# Predicting Personality from Twitter

Evaluation of normalization and summary statistics, and the value of TweetNLP features as predictors

Li Quan Khoo
li.khoo.11@ucl.ac.uk

changtau2005@gmail.com

Emine Yilmaz
E.Yilmaz@cs.ucl.ac.uk
Supervisor

30th April 2015

MEng Computer Science Final Year Project
Department of Computer Science
University College London

This report is submitted as part of the requirement for the MEng Degree in Computer Science at UCL. It is substantially the result of my own work except where explicitly indicated in the text.

| | |
|---|---|
| Title: | Predicting Personality from Twitter: Evaluation of normalization and summary statistics, and the value of TweetNLP features as predictors |
| Student: | Li Quan Khoo (li.khoo.11@ucl.ac.uk / changtau2005@gmail.com) |
| Supervisor: | Emine Yilmaz (E.Yilmaz.cs.ucl.ac.uk) |
| Date: | 30th April 2015 |

# 1 Abstract

We present a 6-month-long multi-objective two-part study on the prediction of Big-Five personality scores from Twitter data, using lexicon-based machine learning methods. We investigate the usefulness of models trained with TweetNLP features, which have not been used in this domain before.

In the first part of the study, we cast personality prediction as a classification problem, and we investigate how prediction performance is affected by different methods of data normalization, such as whether we divide each feature by word count. In the second, main part of our study, we cast it as a regression problem, and we investigate the differences in performance when we use ranks of scores rather than actual scores, and how filtering only for users with over a certain tweet count affects prediction performance.

We report on the different methods used in existing literature, explain background information about the tools we used, and look at the common evaluation metrics used in classification and regression problems and address potential pitfalls when calculating or comparing them. We also suggest a solution on how to reconcile learning parameters for different models optimizing different metrics. Finally, we compare our best results with those in recent publications.

Our main findings are that term frequency `tf`-normalized features perform most consistently, that filtering for users (>200 tweets) improves prediction performance significantly in the regression problem, and that prediction performance using ranked data is comparable to using actual values. We found that models trained with TweetNLP features have comparable or superior performance to those trained with LWIC and MRC features commonly used in literature. Models trained with both have superior performance. Compared against 15 recent models (3 papers, 5 personality scores), our best models are better at prediction than 11 of them.

This report is submitted as part of the requirement for the MEng Degree in Computer Science at UCL. It is substantially the result of my own work except where explicitly indicated in the text.

# 2 Acknowledgments


I would like to thank:

*Emine Yilmaz*
for being such an inspirational supervisor
for your passion and keen analytical grasp of the subject matter
and for being as understanding as ever when the going gets tough.

*Chris Aniftos*
for kindly providing all the data required for this research,
and for the lengthy, weekly meetings and discussions
when we bounced ideas off each other.

*Kevin Bryson*
for being an amazing personal tutor and friend
who is always there
to discuss just about anything weighing on my mind.

Fellow researchers
who have done all this hard work long before I even began.

And all my family, colleagues, acquaintances, and teachers
who have supported and inspired me
all this way.

# Table of contents

# 3 Introduction

We leave traces of our personality everywhere we go. They subtly dictate, to an extent, what we do, what we like, how we behave, and how we are perceived by others. This also means other people can make judgments on our personality based on their perception of ourselves. Empirical evidence suggests that total strangers can make reasonably accurate impressions of someone's personality not just based on physical appearance (Naumann 2009)[40] and facial expressions (Kenny 1992),[41] but also based on cues such as music preferences (Rentfrow 2006)[42] and even choice of footwear (Gillath 2012).[43]

These physical cues can be difficult for people to pick up, and the associations with personality are understandably weak, so reliable predictions are not to be expected, much less by an automated process or computer. However, there are other traces much more useful to us, namely the online footprints many people leave all over their activity over the internet every day - where they shop, what they like, what they are reading about, talking about, blogging about, with whom they are interacting with, and how they are doing all that. Online social networks, often abbreviated to OSNs in the literature, are a very rich source of data for these purposes, as they are essentially representations of the profile owner in online space. Other rich sources of information tend to be creations of the user, such as blogs, emails, and personal essays.

Various aspects of personality can be predicted from just what a user likes on Facebook (Kosinki 2012 ).[3] Kosinki et. al. established a prediction accuracy of between 0.6 to 0.95 for user's behavior outside of personality, for example, drinking, smoking, gender, sexual orientation, religion, etc., from a 100-dimension-PCA-reduced feature set consisting of domains which the user likes. Companies already use candidates' Facebook profiles to get an initial impression of their person before an interview. Perhaps they find value in having an expectation of the user's behavior, or personality, or perhaps they are looking to pick up warning signs? Research in this area is about automating this process, finding out which aspects of personality correlate, and finding applications for them.

Two large-scale studies (n=10k to 50k) using Facebook information (Bachrach 2011, Kosinki 2012)[2,3] attain reasonable predictive accuracy with just what the user likes. However, most other studies, including ours, are limited to the order of a few hundred users, or even fewer, due to lack of a unifying effort such as the MyPersonality Facebook application (Quercia 2011)[10] that made those large-scale studies possible. Having such a large sample allows them to capture the correlations within a massive sparse feature set, which is then dimensionally-reduced. In contrast, smaller-scale research involving OSNs other than Facebook tend to focus on affective language processing - which is a sub-field in computational linguistics, specializing in extracting sentiment, opinion, emotion, or point of view - by analyzing what the user has written rather than his or her interactions with other users or entities on the web.

Affective language processing methods tend to rely on pre-annotated "classes" of words. In sentiment analysis, for example, there are dictionaries of words (or more specifically, word senses) annotated with sentiment scores, which could then be used to derive the sentiment of a written piece of work. Similarly, the LWIC (Pennebaker 2001)[33] and MRC (Coltheart 1981, Wilson 1988)[34,35] group certain words together that represent a simple psychological concept, such as 'work', 'swear words', or 'sexual'. Based on the reasoning that our personality is reflected by how we speak and use certain words, these efforts focus on finding associations between these groups of words and personality scores, and building predictive models out of them. This is what our research is about.

## 3.1 Value of personality prediction

### 3.1.1 Targeted advertising

In terms of marketing, (Kassarjian 1971)[45] explains the segmentation of the market into different segments based on interests, values, personality, and demographic variables. This directly relates to the maximization of consumer surplus by price discrimination, by estimating the true valuation of a product to a customer more accurately at a given time by factoring in their personality. For example, limiting the window of opportunity of a deal has been used as a marketing strategy to increase impulse buys. (Rook, 1985)[46] reported up to 400% increase in potato chip sales through such strategies, and attempted to identify the reasons behind impulsive buying. In a gender-balanced case study (n=212), they found a 0.51 correlation ($p<0.001$) that impulsive shoppers enjoy shopping more than those who do not. Their study suggests that different marketing strategies can have very different impacts depending on the personality of the receiving demographic. More importantly, it means a person's personality affects what he or she likes doing. They also found that "Inner-directed" persons are less likely to be influenced by "other-directed" people and by social pressures. Should a click-through on an ad by a highly influential user pay more, because the user is more likely to mention the product to his or her peers?

Golbeck (2011)[7] suggested that predicted personality could be used to suggest music, based on established correlations in the literature. It is not unreasonable to expect this relation to extend into other forms of products, primarily movies, books, and games, as they are products related to information consumption and entertainment. Current recommender systems make use of collaborative filtering and association rule mining to establish product suggestions, but it is likely that predictive accuracy could be improved by segmenting the sample population by various aspects or personality, at least for certain products like music.

The most convincing argument we have for the use of personality prediction in this domain lies with companies like Facebook and Twitter. They are well-positioned to take advantage of personality prediction in ad-serving decisions, for instance. Facebook already make use of their own tracking cookies and already operates as a third-party advertising network, which they can associate with the user who is browsing one of their affiliated websites (Milic-Frayling 2015).[31] Given that a corporate giant such as Facebook has part of their business model based on dealing in personal information, there are certainly powerful incentives driving this kind of research forward.

#### 3.1.2 Detecting state of mind

The elegance of dictionary-based methods is that as long as we have a dictionary of any form, whether it's one that's readily available or generated from annotated corpuses, is that text can then be used to predict the domain of that dictionary. Sentiment lexicons give us sentiment information, and personality dictionaries, which we use, give us personality scores. There are efforts to predict depression from social media, but the field is is still in its infancy (Choudhury 2013).[26] Choudhury et. al. constructs their own depression lexicon from labeled Twitter data, removing frequent terms via tf-idf normalization, with the objective of a gold-standard resource for further research. Their model predicts onset of a Major Depressive Disorder (MDD) episode ahead of time, with 70% accuracy and 74% precision.

## 3.2 Privacy concerns

While Choudhury's efforts are an example of rigorous research in this area, there is also the incidence of the well-meaning but badly-executed Samaritans Radar application that made the news and was withdrawn due to privacy concerns. In concept, it is extremely simple - it relies on word matches to detect depression, so in principle it is the same as what we are doing. But, due to the fact that they rely on a small number of fixed words that trigger a high positive rate, and the fact that they made such sensitive information public, it raised privacy concerns and was withdrawn. In the end, we need to remember that are dealing with personal information, whether it is about depression, personality, or other aspects of users' private lives.

What might be alarming is that once we have trained models which predict these various personality aspects with reasonable accuracy, we can deduce this information from users who have no idea that they are disclosing such information behind by simply acting as they always have. However, despite these concerns, there is clearly value in personality prediction, in at least these two major areas explored.

Now that we have explained some background and potential uses of this technology, we will examine personality prediction as a machine learning problem.

## 3.3 Challenges

Although collecting the data that generates the features tend to be cheap, for example, through the Twitter / Facebook API, the labels (personality scores) are difficult and laborious to collect. Users have to either take a standardized personality test (called an inventory) themselves, or some studies had the users' acquaintances take a modified 'observer' form of the inventory, and take the average of the user's personality scores. Other than Facebook, a large ($n > 100k$), centralized collection of mappings between users' social network profiles and their personality scores does not exist, so existing literature makes use of datasets which the researchers collected themselves. Such collections have small sample sizes and often have sampling or gender bias (Bachrach, 2011),.[2] Thus, unlike text-retrieval or prediction studies on common datasets such as TREC, this makes replication of results or comparison of effective methods difficult, if not impossible, unless one implements the exact system as documented in the paper we wish to compare to..

There is an important caveat within the personality scores themselves. Mairesse et. al. (2007)[12] and Qiu et. al. (2010)[9] found correlations between personality scores and LWIC and MRC features. However, the correlations and their significance values differ between self-reported and observed personality scores. Most papers use self-reported scores, but some use an average between the two (Back 2009),[1] so this is an important distinction to remember when reading about existing work, as not all scores and results are comparable to one another.

## 3.4 Objectives

We will briefly discuss our main objectives and the motivations leading up to them. They will be covered in more detail when we discuss our method in Section 6.

Among the published papers on predicting personality using Twitter data, we have one of the largest datasets out there (n=372 used in this study, currently over 500 and growing). We hope to replicate results of previous papers and address their shortcomings. Some of them (Golbeck 2011)[7] use evaluation metrics that do not necessarily give an objective view of predictive performance. All of these papers report various levels of correlation between linguistic features and personality scores, but that does not indicate a model's predictive accuracy in any reasonably accurate way, especially since Pearson correlation significance p-values are generated using a probabilistic model, for small n, having a significant correlation in the sample population does not necessarily mean this holds for the true underlying population; Scikit's documentation recommends at least 500 data points, for example. We want to investigate how our model's performance is evaluated by different statistics, and find if there are any ones which may be misleading in the context of this problem.

We also need to address the issue of feature normalization. Raw feature scores are not ideal for training, as we are naturally biasing towards learning features of users with more tweets as they would have more words in total matching the set of words corresponding to that feature. There are many well-known numerical statistics used in information retrieval and search systems that attempts to address this bias. Term frequency normalization (`tf`)'s advantage is in its simplicity as it is the simple division by number of words for that user, and it is used in existing literature (Golbeck 2011),[7] but there are other ways to mitigate this bias, such as truncating the corpus at 2 standard deviations (Qiu 2010).[9] Bachrach et. al. (Bachrach, 2011)[2] chose to convert the feature scores into quantiles instead. We investigate `tf`, `pt` (dividing by number of tweets), and `bm25` (BM25+) (Lv 2011).[19]

Finally, we wish to find if there are other dictionaries which contain word clusters which are useful personality predictors. We will discuss this in further detail in Section 3.5.4, but TweetNLP word clusters (Gimpel 2012, Owoputi 2013)[14,15] are automatically generated by Brown clustering on a massive sampling of actual tweets, instead of hand-written like LWIC and MRC. If we find useful associations, then this may be a promising way to make use of implicit language features to construct word clusters in place of dictionaries. The other way to construct dictionaries is to obtain labeled corpuses as one would do in sentiment analysis, but as we have explained, personality labels are slow and laborious to collect. Since we are adding over 1,000 new features in the form of TweetNLP labels, clusters, and some of our own computed features, dimensionality reduction becomes necessary, but it is not yet known what is a good number of dimensions to reduce to, using our selected features. We want to investigate this.

## 3.5 Background

In the following subsections, we introduce individual concepts and tools related to this project.

### 3.5.1 Big Five Personality Model

The five personality factors Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism, were conceived by Tupes and Christal as fundamental traits that emerged from analyses of personality tests. Despite being subjected to different tests, languages, and models of analysis, the big five model does not change, leading to its acceptance as the current definitive model of personality (Mccrae, 1992; Digman, 1990).[37,38]

As it allows us to quantify a user's personality, it is of particular relevance to our work, and all the other papers related to personality prediction from online social networks also make of big-five personality scores for the same reasons explained above.

The following 5 paragraphs consists of excerpts from (Bachrach, 2011),[2] describing each factor in their own words, which we think is a very good explanation.

– begin excerpt –

*Conscientiousness* measures preference for an organized versus spontaneous approach in life. People high on Conscientiousness are more likely to be well-organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. Low Conscientiousness individuals are generally more easy-going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.

*Openness* to experience measures a person's imagination, curiosity, seeking of new experiences and interest in culture, ideas, and aesthetics. It is related to emotional sensitivity, tolerance, and political liberalism. People high on Openness tend to have high appreciation for art, adventure, and new or unusual ideas. Those with low Openness tend to be more conventional, less creative, more authoritarian. They tend to avoid changes and are usually more conservative and close-minded.

*Extraversion* measures a person's tendency to seek stimulation in the external world, company of others, and express positive emotions. Extraverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative, and do not mind being the center of attention, and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.

*Agreeableness* measures the extent to which a person is focused on maintaining positive social relations. High Agreeableness people tend to be friendly and compassionate, rather than cold or suspicious. They are more likely to behave in a cooperative way, trust other people, and adapt to their needs.

*Neuroticism*, often referred to as emotional instability, is a tendency to experience mood swings and negative emotions such as guilt, anger, anxiety, and depression. Highly Neurotic people are more likely to experience stress and nervousness, while those with lower Neuroticism tend to be calmer and self-confident.

– end of excerpt –

The distribution of these scores is not even across our sample population. The 10-bin histograms of individual Big-Five scores are plotted in Section 7.1, to investigate the independence of score distribution and the number of tweets, illustrate this distribution across different slices of the sample population.

#### 3.5.2 LWIC

Homepage: http://www.liwc.net/

LWIC (Pennebaker 2001)[33] stands for Linguistic Inquiry and Word Count. It is a text analysis software program designed by Pennebaker, Booth, and Francis et. al. Simply put, LWIC maps words or word patterns to 64 predefined groups, like certain pronouns, swear words, or words related to certain ideas like religious, workplace-related, or sexual-related words, to name a few.

These mappings are boolean rather than weighted, so a word or word pattern is either in a group or not. A word can be in more than one group. A word pattern is a combination of alphabetic characters and a wildcard, to potentially match different word-endings. Other than the wildcard, matches for the alphabetic part have to be exact (after lowercasing).

LWIC is of interest to us as existing literature (Golbeck 2011, Qiu 2010),[7,9] and others has found significant correlations between the number of LWIC words used in a user's social network profile and their personality scores. We use all LWIC word classes as features.

#### 3.5.3 MRC

Homepage: http://www.psych.rl.ac.uk/

MRC (Coltheart 1981, Wilson 1988)[34,35] is a collection of many different groupings of words, compiled from many different publications in the past. It consists of about 150,000 words in total. In the dictionary, each word is mapped to a vector of numbers or letters. We extract the following classes of words from MRC:

| Column | Code | Property |
|---|---|---|
| AOA | - | Age of acquisition |
| FAM | - | Familiarity |
| CONC | - | Concreteness |
| IMAG | - | Imagery |
| MEANP | - | Mean Pavio Meaningfulness |
| MEANC | - | Mean Colerado Meaningfulness |
| STATUS | F | Alien |
| | A | Archaic |
| | C | Capital |
| | Q | Colloquial |
| | W | Nonce word |
| | E | Nonsense |
| | O | Obsolete |
| | P | Poetical |
| | H | Rhetorical |
| | R | Rare |
| | $ | Specialized |
| | S | Standard |
| IRREG | N | No plural |
| | P | Plural acting singular |

#### 3.5.4 TweetNLP

Homepage: http://www.ark.cs.cmu.edu/TweetNLP/

We use two components of Carnegie Mellon University's TweetNLP library (Gimpel 2012, Owoputi 2013).[14,15] The first part is the POS (part of speech) tagger. Given a sequence of whitespace-delimited character sequences, the POS tagger tags each sequence with one of the following POS labels (Gimpel, 2012).[14]

| Token | Description |
|---|---|
| N | common noun |
| O | pronoun; not possessive |
| S | nominal + possessive |
| ^ | proper noun |
| Z | proper noun + possessive |
| L | nominal + verbal e.g. *i'm, let's* |
| M | proper noun + verbal |
| V | verb including copula, auxiliaries |
| A | adjective |
| R | adverb |
| ! | interjection |
| D | determiner |
| P | pre- or postposition, or subordinating conjunction |
| & | coordinating conjunction |
| T | verb particle |
| X | existential *there* |
| Y | X + verbal |
| # | hashtag |
| @ | at-mention |
| | discourse marker |
| U | URL or email address |
| E | emoticon |
| $ | numeral |
| , | punctuation |
| G | other abbreviations, foreign words, possessive endings, symbols, garbage |

The second part of TweetNLP used are the word clusterings produced by Brown clustering. We are using the `50mpaths` cluster, which is the largest out of all available ones. These are 1,000 clusters of words produced from a random sample of 100,000 tweets per day between September 10, 2008 to August 14, 2012, after filtering out non-English tweets. Based on the papers by the creators of TweetNLP (Gimpel 2012, Owoputi 2013),[14,15] these are clusterings of functionally-similar words independent of correctness of spelling. Simply put, these are groupings of words based on their relative positions as found in English sentences i.e. based on the function that word serves.

We consider each POS tag from the POS tagger as well as each of those clusters produced by Brown clustering as a class of words, and we derive features based on the number of words a user uses in each of these word classes. By making use of these clusters, we hope to account for the fact that Twitter text is noisy, and curated dictionaries such as LWIC which require exact word match would miss misspelled words.

There is no existing literature on predicting personality using TweetNLP features derived from authored text that we are aware of, so this is completely new ground. LWIC and MRC dictionaries have features constructed out of words serving a common purpose, such as pronouns or adverbs. And since TweetNLP words are functionally similar based on how they are grouped, our hope is that because TweetNLP cluster construction does not require human intervention or curation (like LWIC or MRC) or collection of personality scores, if we find significant correlations between such clusters, even larger clusters could be constructed to improve classification performance.

We will briefly explain the concept behind Brown clustering (Brown 1992),[13] since this is how the TweetNLP word clusters are constructed. Our explanation is based on Michael Collin's (2015).[28] The observation is that similar words are used in similar contexts. For example, adverbs tend to precede adjectives, verbs, or other adverbs. Considering two different words, both of them have a probability distribution of words occurring before and after them. We can use these probabilities to define a notion of similarity between these two words are considering.

Formally, the Brown clustering algorithm relies on a partitioning function $C$, that partitions our vocabulary of all words $V = \{w_1, w_2, ...w_T\}$, into $n$ clusters.

$$C : V \longrightarrow 1, 2, ..., k$$

Of all the different partitions, we seek to maximize the 'quality' of the partitioning, which we call $Quality(C)$. This quality is computed as the log likelihood of the corpus. We have the function $e(w_i, c)$, which defines the probability of cluster $c$ emitting the word $w_i$, and the function $q(w_i|w_{i-1})$, which describes the probability of the current word $w_i$ following the previous word $w_{i-1}$.

The log likelihood of the corpus is defined in terms of $C$, $q$, and $e$ as:

$$p(w_1, w_2, ..., w_T) = \prod_{i=1}^{n} e(w_i|C(w_i))q(C(w_i)|C(w_{i-1})$$

Intuitively, that expresses, for all possible groupings of words in our vocabulary $V$, what is the likelihood of the corpus we are looking at. More conveniently,

$$\begin{aligned} Quality(C) &= log\, p(w_1, w_2, ..., w_T) \\ &= \sum_{i=1}^{n} log\, e(w_i|C(w_i))q(C(w_i)|C(w_{i-1}) \end{aligned} \tag{1}$$

The above equation happens to simplify into the equation for mutual information. $c$ and $c'$ are any pair of clusters being considered, and $G$ is a constant.

$$Quality(C) = \sum_{c=1}^{k} \sum_{c'=1}^{k} p(c, c') log \frac{p(c, c')}{p(c)p(c')} + G \quad (2)$$

The modern implementation of Brown's clustering starts with a seed set of $m$ most frequent words in the corpus. At each step, from $i = (m+1)...|V|$, it adds one word $w$ from the rest of the vocabulary into the set of words being considered. At first, each word belongs to its own cluster, and the algorithm performs a merge by calculating $Quality(C)$ over the $m+1$ words. Finally, when all words in the vocabulary have been added, it merges the clusters in $m-1$ steps, to create the full binary tree.

In Brown clustering, words are clustered together because they are most interchangeable within the sentences in the corpus. We think such clusters might be associated with certain personality aspects, by capturing the fact that users with certain personality types may tend to use certain sentence constructions, non-standard words or expressions, which are language aspects not captured by dictionary-based features.

#### 3.5.5 Other features

So far, we've covered specialized toolkits to extract specific data from Twitter and how we use them to generate our training features. Our remaining features are based more loosely on various papers and findings.

The first we consider is word lengthening. Prosodic indicators (e.g. high pitch, prolonged duration, intensity etc.) have long been known as ways for a speaker to emphasize a word (Bolinger 1965).[36] However, in written text, most of these features are lost, although sometimes they are reintroduced in the form of typographic styling. In cases where this is not always possible (like with Twitter), users resort to word lengthening and capitalization, or repeated punctuation marks.

Brody et. al. (2011)[16] reported strong relations between word-lengthening and sentiment. Based on their dataset, this phenomenon occurs roughly once in every 6 tweets - very common indeed. Although we are not measuring sentiment, it would be reasonable to assume that there would be some relation between users' personality and how they emphasize what they want to express, and by how much.

We identify several features, including:

- runs of the same letter (>2) in a word
- runs of '!' or '?' consecutively
- runs of capital letters in a word

For each of these, we log how many runs were there in total, how long each run was, and the mean and mode length of the runs.

# 4 Prior work

In this section, we discuss literature surrounding the subject matter, focusing on the different methods used.

## 4.1 OSNs reflect personality

By attempting to predict a user's personality from his or her Twitter profile, we are implicitly making the assumption that these variables are correlated in some way.

Back et. al. (2009)[1] conducted an experiment (n=236) that not only found significant correlations between a user's personality scores, but the correlations to their actual personality scores are much stronger than those of their idealized selves. They conducted their studies across three OSNs (Facebook, StudiVZ, SchuelerVZ) used true and idealized scores from the TIPI and BFI-10 personality inventories, aggregated from the user and 4 of their acquaintances.

## 4.2 Facebook

Facebook is a popular OSN to work with. While the Facebook-specific features available are quite different from those in Twitter, this gives us an idea of what kind of accuracy we can expect from OSN data.

Kosinki et. al. (2012)[3] conducted a study on users from the myPersonality Facebook application (n=58,466) used Facebook likes to predict users' personality scores and several other dichotomous variables (binary labels), such as whether they are single, whether they use drugs, their sexual orientation etc., and managed to yield Pearson correlation coefficients of between 0.3 to 0.43 (p<0.001) between actual and predicted personality scores. They compiled the various things their set of users liked on Facebook and reduced the dimensionality to 100 dimensions via SVD (singular value decomposition), and then ran logistic or linear regression with10-fold cross validation on the data.

Bachrach et. al. (2011)[2] worked on a wider set of Facebook features, including the number of Facebook friends, groups, likes, photos and statuses posted, and the number of times a user was tagged in a photograph. Using a larger set of users (n=180,000) from the myPersonality Facebook application, they showed that the user's big five personality score, ranked within that set of users, corresponds to the user's rank in terms of a Facebook feature. When they attempted to predict personality via multivariate linear regression (10-fold cross-validation), and they managed $R^2$ scores of between 0.01 (agreeableness) and 0.33 (extraversion). This was by far the largest study in terms of number of users involved, and one of their criticisms of other studies was that the sample size is small, or biased in some way.

Gosling et. al.'s work (2011)[4] is perhaps an example of the kind of work Bachrach et. al.'s criticisms are addressed at. Their sample (n=159) was drawn from psychology students in Washington University, but they use the same Facebook features, plus a handful of others. There are marked differences in their findings in comparison to Bachrach et. al.'s, as the large majority of significant correlations ($p<0.05$) are to extraversion only. The correlation metric is not clear, but is assumed to be Pearson correlation.

## 4.3 Microblogs

Microblogs are OSNs with very low text limit in each post. Examples include Twitter and Sina Weibo. The nature of the work on microblogs is very different from that on Facebook; many of the papers we will discuss use linguistic features from what the user typed, rather than features specific to microblogs, as seen in Facebook's case. The nature of this work, then, is much closer to similar work on blogs and essays and textual / sentiment analysis than than the work on Facebook features.

The methods for collecting the data usually involve recruiting the users in some way, collecting their Big Five personality information using one of the various inventories such as BFI-10 and then requesting their microblog account username (or equivalent), and then extracting information from their blogs or Tweets posted before they took the personality test.

Qiu et. al. (2010)[9] collected 28,978 tweets from users (n=142) between May 25th and June 25th in 2011 (mean word count=2362.72, sd=2535.97) and computed LWIC dictionary scores for each user. They found correlations between LWIC features and personality scores (there are too many to list here). The authors also summarized previous work done for predicting personality using LWIC features extracted from all kinds of other corpuses including blogs, personal essays, and emails in a table, and the significance of these correlations only sparsely overlap between the different corpora. This may be attributed to the different forms of writing involved. When comparing their results to our work (and most other work), we are interested in what they call cue-validity correlations, which are correlations of LWIC dictionary scores to personality scores obtained by the user's self-report (rather than his or her observers'). This is a psychology paper, and the authors did not attempt to predict personality.

Golbeck et. al. (2011)[7] is a relatively widely-cited paper for personality prediction in microblogs. Their sample consists of users' (n=279) most recent 2,000 tweets (mean word count=1914, mean tweet count=142.2, min and max no. of tweets=4 and 350 respectively). Twitter specific features that they used include no. of followers, no. of users the user is following, and the number of @mentions, replies, hashtags, links, and words used per tweet. Via Gaussian process regression and ZeroR with 10-fold cross-validation, they managed to predict personality scores within 11% and 18% of their actual values.

Gou et. al.'s study (2013)[6] involved 256 Twitter users from IBM (at least >200 tweets). Their work attempts to predict more than the Big-Five personality features, including the users' basic values (such as openness to change), and needs (such as curiosity), but for the Big-Five personality scores, they only use LWIC features. They found that for 80% of their sample population, predicted personality scores correlates with the actual personality score. More interestingly, they reported that by sampling just 200 tweets, they managed to produce within 10% rank of the results obtained using the full set of tweets.

Quercia (2011)[10] are the team who worked on myPersonality. Using Twitter-specific features like number of followers and follower count, as well as Facebook data for users who have presence on both OSNs, they found significant correlations between log of number of followers and number of other people being followed. Using their dataset (n=335), they performed 10-fold-validated decision-tree regression, achieving RMSEs of between 0.69 and 0.88, for a normalization range of [1, 5] (effective range of 4). Their work does not involve lexicon-based analysis of the users' text, so it provides a different point of comparison for our work.

There is a very recent paper by Liu et. al. (2015)[11] that makes use of an active learning method. This is interesting, because we have explained that personality labels are expensive to collect, so their system suggests to them the users which are predicted to make the most improvement, and they recruit these users for their study. Due to the date of publication of this paper, we have already all of the data used in this paper, but it is something worth doing for future endeavors. In a sample set of 100, they reported Pearson correlation coefficients of between 0.10 (for extraversion) and 0.21 (for conscientiousness).

### 4.4 Other

Oberlander et. al.'s work (2006),[8] conducted a study on the blog corpuses of 71 users (relatively biased gender ratio, 47 females, 24 males), using bi- and trigrams rather than LWIC features for prediction. The nature of their evaluation is slightly different than what we have seen discussed so far. For their binary classifier (for each personality feature), they split the users evenly into 2, 3, or 5 bins, but for the odd numbered splits, they discard the middle bin(s), leaving the 2 at either extreme. They discard the middle bins to evaluate their classifiers' performance on extremes of data.

## 5 Data

> *Project disclaimer:*
> I was not personally involved in data collection. Chris, with whom I have very limited collboration under Emine's supervision, gave me the tweets he collected in the form of a MongoDB Collection.

The tweets were collected through Chris' website (http://www.urself.org/) and through Amazon's Mechanical Turk service, recruited through advertisements on Twitter and via a Turk task, from 21st May 2014 to 14th August 2014. Participants would be given their personality scores at the end, and asked to verify their Twitter account via the Twitter API. For the participants recruited through Mechanical Turk, they received monetary compensation as well.

The inventory used to collect the Big-Five Personality scores is IPIP (http://ipip.ori.org/), hence based on our explanations in Section 3.3, they are mostly incompatible with observer-reported scores when comparing to existing literature. Unfortunately, we do not have demographic information about the sample population, but when using our results as comparison, ANOVA methods can be used to test for significant differences.

When we began our investigation, our dataset contains 372 users' personality scores and their tweets. After manually filtering out 16 users with mostly (>80%) automated tweets, we are left with 356 users and their tweets. In the process, we made sure that all the tweets we have area in English, because LWIC and MRC are in English, and TweetNLP word clusters are constructed based on Tweets that passed the `langid.py` language checker as English.

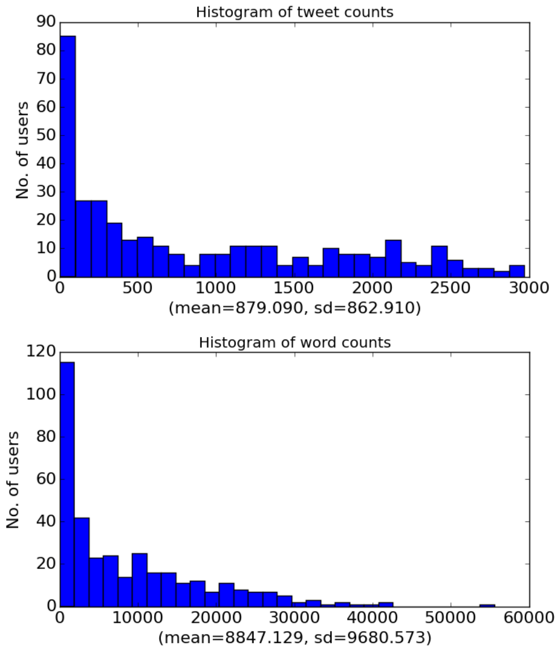



Figure 1: Distribution of word and tweet counts across all users, after filtering out automated tweets, but before filtering for users who have >200 tweets.

In section 7.1, we plot histograms of users' personality scores. The plots show that the user's personality score is independent of the number of tweets submitted. Knowing this fact, for our classification and one of our regression tasks, we filtered only for users who posted more than 200 tweets, to filter out inactive users and lessen noise. We end up with n=243 users when doing so. The threshold of 200 was picked based on Gou et. al.'s result (2013),[6] which showed that using 200 tweets, ranked personality scores can be predicted to within 10% accuracy of the score predicted using every single tweet. For the other two regression tasks where we do not perform the filtering, we use all n=356 data points.

Because all of our features are computed from the users' tweet corpuses, all users have the complete feature set, i.e. we do not have any one data point with a feature whose value is unknown or undefined.

# 6 Method

All of the features we have are simple counts. Most of them involve counting words or punctuation marks, a few others come from Twitter data such as the number of tweets. For the word counts, we simply concatenate all of the user's tweets together, and pass this long spacebar-delimited string into a function. For generating TweetNLP's labels (such as hashtags), the library requires exactly this. For the other features, we explode the string on the spacebar character, and for each token, we count the relevant punctuation marks, then strip them, lowercased the word, and attempt to match it against individual sets of words in LWIC, MRC, and TweetNLP 50mpaths clusters. Only exact matches count. This gives us the raw counts of the features we are investigating.

The LWIC, MRC, and TweetNLP dictionaries are distributed in the form of files. To implement them, internally, we hash each word defined in a cluster in a dictionary to that cluster to match them quickly. For the LWIC dictionary, however, as we explained in Section 3.5.2, a large proportion of their definitions are in the form `word*`, to match different word-endings. For those, we define a regular expression for each such definition and match each word against them, at significant performance cost.

Simply using raw counts is biasing towards users with the longest corpus. These would be users who are most active on Twitter, or have been active for the longest. Hence, we want investigate the effect on classification or regression accuracy using common normalization methods (for document length) in text retrieval:

- `tf`: term frequency scores (by dividing by total number of words)
- `pt`: per tweet scores (by dividing by total number of tweets)
- `bm25`: BM25+ scores

We attempted to transform features into BM25 ranks by recommendation of the project supervisor, as this metric is very commonly used in search retrieval settings. We have no idea how these features are going to perform in this context. We used a modified form of BM25 called BM25+ (Lv 2011),[19] and it requires calculating the inverse document frequency, or IDF. There are various formulas for IDF with different techniques for smoothing.

We used simple inverse frequency IDF. Where $n$ is the number of users we have and $X$ is their feature vectors, the denominator part is the number of users whose this feature is not equal to zero:

$$IDF = \log \frac{n+1}{\{x \in X : x_t \neq 0\}} \tag{3}$$

Where $x$ is the raw counts of the feature for a user, $|D|$ is the length of the user's tweets corpus in words, and $avgdl$ is the mean number of words used by each user, the BM25+ formula is given by Equation4. $k_1$ and $b$ are tuning parameters, which we set to 1.6 and 0.75 respectively in absence of optimization (Mannning 2009).[29] $\delta$ is the "+" in BM25+, and serves to not over-penalize long documents (Lv 2011).[19] We set it to the value of 1, based on Lv's paper.

$$BM25+ = IDF \cdot \frac{x \cdot (k_1 + 1)}{x + k_1 \cdot (1 - b + b \cdot \frac{|D|}{avgdl})} + \delta \tag{4}$$

## 6.1 Classification

We start off our investigation by framing personality prediction as a classification problem, to familiarize ourselves with the data and answer a few questions about data representation. The objectives of this task is to:

- Confirm that there is at least some form of correlation between personality scores and our features, and if there are, how strongly they correlate.
- Find out if there is any correlation between personality scores and TweetNLP features, and the comparative performance between using it and LWIC + MRC
- Investigate which kind of normalization gives the best performance (`tf`, `pt`, or `bm25`).
- Find out if any of our features defined in Section 3.5.5 are useful predictors.

In order to define a classification problem, first we need to segregate the labels into different classes. Initially, we segregated them into bins of the same size i.e. 33.33 each, as personality scores have a range of [0, 100]. However, the attributes agreeableness and intellect have skewed distributions, and one of the bins would not have enough data points for all 10 folds. Because a few particular bins are so sparsely populated, we have the problem of undefined precision or recall. Simply skipping the folds would also unduly affect evaluation calculations (Forman 2010),[24] making the model seem better than it actually is.

To sidestep the entire issue, we transformed the personality scores into three quantiles instead, by first ranking them and then specifying boundaries for those ranks. This is in line with what Bachrach et. al.

(2011)[2] did in their Facebook study. We do not convert feature scores into ranks like they did as there is no need to do so; rank conversion necessarily involves loss of information, namely background distribution of the scores. Theoretically, we can recover the original personality score within a certain range, provided that we record the mean and standard deviation or min and max of personality scores of each quantile before conversion.

We constructed the training and validation sets with stratified K-folds with 10 folds. We use principal component analysis (PCA) and reduce data dimensionality via singular value decomposition (SVD). PCA can only reduce the feature space to a maximum number of dimensions equal to the number of training items (as there cannot be more variance in the data than there are points of data), hence we are limited to n=243 dimensions at most. Starting from this maximum value, we reduce it one by one until we have just 1 left. After each dimensionality reduction, we then train our 3-way classifier by L1-regularized logistic regression and calculate the 3 metrics we use for evaluating classification: AUC (area under receiver operating characteristic curve), AP (average precision = area under precision-recall curve), and F1 score on the validation set. Each metric's value is first computed over a fold, and then the mean of these values are taken to give us our final value.

*Caveat:*
The other way to average metrics is to collect all scores from all folds together and then plot one single ROC or PR curve to calculate AUC and AP, but there are implications in doing so (Forman 2010).[24] It will downgrade classifiers that perform well if they have poor calibration across folds. Calibration here means the drift between: the plot of the classifier's true positive rate (TPR) against classifier thresholds, and the plot of false positive rate (FPR) against the thresholds.

L1 regularization already selects for features implicitly, but initially, when we performed regression on the full feature set without dimensionality reduction, we got less favorable results. From the Pearson correlation table in the appendix, for any given feature, we see that most features are not useful to predict a given personality score, and a simple method like L1 regularization alone is not enough to filter these out.

To investigate how much a feature set (e.g. LWIC or TweetNLP) contributes, we re-run evaluation using only a subset of features, by zeroing out those that we do not want. The results are tabulated in Figures 6, 7, 8, and 9, in Section 7.

#### 6.1.1 Review

We will briefly digress to review some machine learning terminology, along with properties of our evaluation metrics and some important caveats working with them. After that, then we will return to discuss why we chose to perform L1-regularized logistic regression.

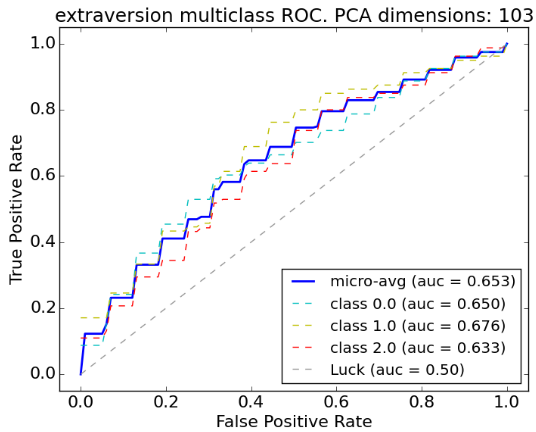



Figure 2: Example ROC plot with corresponding AUC from our classification results. There are actually 4 ROC curves in this figure, 3 for the binary classifier of each class, and the micro-average.

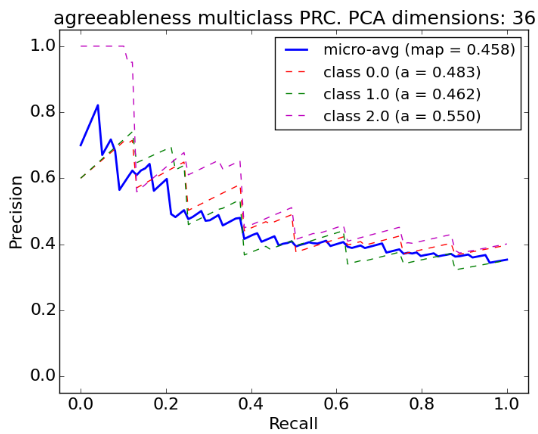



Figure 3: Example PR plot with corresponding AP from our classification results. There are actually 4 PR plots in this figure, 3 for the binary classifier of each class, and the micro-average.

AUC is the area under the receiver operating characteristic curve, which is a plot of true positive rate against false positive rate when the classifier's decision threshold is varied. ROC plots often have a reference line with AUC = 0.5 across, which represents the ROC curve of a random guesser for two equally-sized classes of labels. In our case, however, we have three equally-sized quantiles, so the baseline AUC of a random guesser is $\frac{1}{3}$. The range of AUC is [0, 1].

A precision-recall or PR plot is produced by first ranking all items by their 'score' in terms of the loss function, in ascending order. Precision $p$ and recall $r$ can be computed at each of these points and a PR curve plotted. The average precision (AP) of the model corresponds to the average value of precision as a function of recall $p(r)$, which happens to correspond to the area under the PR plot. The AP of a random guesser for 3 equally-sized classes is $\frac{4}{9}$. The range of AP is [0, 1].

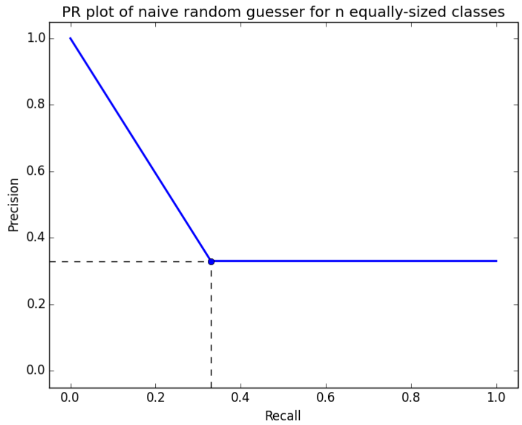



Figure 4: PR plot for naive random guesser with n equally-sized classes. Turning point is $(\frac{1}{n}, \frac{1}{n})$

*Caveat:*
There is an unachievable region in P-R space (Boyd 2012).[23] The size of this region depends on the skew of the data (between positive and negative classes), and Boyd suggested that when calculating relative improvement in terms of AP, this region should be subtracted from both models first.

The F score (or measure) is a measure of a classifier's accuracy. $F_1$ is a weighting where precision and recall contributes equally to the score. The precision and recall of a random guesser for 3 equally-sized classes are both $\frac{1}{3}$, hence the $F_1$ score is also $\frac{1}{3}$. The range of $F_1$ is [0, 1].

$$F_\beta = \frac{(1+\beta^2)}{\beta^2} \cdot \frac{p \cdot r}{p + r} \tag{5}$$

Where $y$ and $x$ are the labels and training vectors, $\theta$ are the model parameters and $b$ is the bias term, $m$ is the number of training examples, and $n$ the 'length' of $x$, the formulation for the optimization problem presented by a regularized linear model is:

$$\arg\min_{\theta,b} L(\theta, x) = \sum_{i=1}^{m} L(\theta, x) + R(\theta) \tag{6}$$

The loss function $L(\theta, x)$ for logistic regression is:

$$L(\theta, x) = \log \frac{1}{1 + e^{-y(\theta^T x + b)}} \tag{7}$$

where $R(\theta)$ is the regularization parameter. Where $\lambda$ is the scaling factor,

The $L^1$ norm (of the model's parameters $\theta$) used to penalize coefficients $\theta$ in so-called lasso methods are:

$$R(\theta) = \lambda \sum_{j=1}^{n} |\theta_j| \tag{8}$$

The $L^2$ norm in ridge-regression methods is:

$$R(\theta) = \lambda \sum_{j=1}^{n} \theta_j^2 \tag{9}$$

Finally, both $L^1$ and $L^2$ norms are used in elastic net regularization:

$$R(\theta) = \lambda_2 \sum_{j=1}^{n} \theta_j^2 + \lambda_1 \sum_{j=1}^{n} |\theta_j| \tag{10}$$

### 6.1.2 Why L1-regularized logistic regression

It is worth discussing why we used L1 regularized logistic regression. The main difference between linear and logistic regression is that logistic regression accounts for biases in class distributions. We were examining the problem in terms of bins initially, where each bin can have very different numbers of users, until we ran into the problem of undefined AP we mentioned earlier. In that context with unbalanced classes, using logistic regression makes sense.

As for the choice between L1 and L2, we use results from Ng (2004)[17] and Vapnik (1984).[18] According to Ng, it is well known that, for unregularized discriminative models such as linear regression, sample complexity (number of training examples needed to learn well) increases approximately linearly with the Vapnik-Chervonenkis (VC) dimension of the model. The VC dimension of the model is its complexity or expressive power, defined in terms of the maximum cardinality of points it can shatter. According to Vapnik, the VC dimension for most models grows about linearly with the number of parameters, which grows at least linearly with the number of input features. This means that the number of training examples for most models needs to grow at least linearly with the number of input features for most models to learn well, otherwise mechanisms such as regularization is needed to encourage the parameters to remain small to prevent overfitting.

According to Ng, L1-regularized *logistic* regression has a sample complexity that grows logarithmically with the number of irrelevant features, which matches the best known bounds. For L2-regularized logistic regression, Ng proved that it is a rotationally invariant algorithm, and for any rotationally-invariant algorithm, (including SVMs and neural networks) may not be suitable for the cases where there are only a few relevant features, or where number of training examples is significantly smaller than the input dimension, which is our case.

## 6.2 Regression

In order to make our results comparable to existing literature (which have results for regression), we perform ridge (L2-regularized least-squares) regression on the data, evaluating with mean square error (MSE), mean absolute error, (MAE), coefficient of determination (R2) scores, and the Pearson (PEAR) and Spearman rank (SPEAR) correlation coefficients. The difference between Pearson and Spearman rank coefficients is that Spearman does not assume both variables to have a linear correlation. Look at Bachrach et. al.'s results, some ranked Facebook features do not correlate linearly with ranked personality scores, hence if we have the same situation, Spearman rank correlation will capture this better than Pearson correlation.

The reason we chose ridge regression is to see if there are any differences between L1 and L2 regularization. The advantage of L1 regularization, other than what was already mentioned in Ng's paper (2004),[17] is that it selects for features and creates a sparse model, but since we are already reducing data dimensionality by PCA (which improved classification performance even when using L1), we use L2-regularization to construct a non-sparse model using the features after dimensionality reduction.

We first ran ridge regression on the set of all users with more than 200 tweets (n=243), the same set as the one we performed the classification task with. We then did the same for the set of all users, but without filtering for tweets, to observe the difference in performance. One of two things could happen - one is that the extra training points are beneficial, the other is that these points contain so few features that they contribute only to noise.

Then, as a final exercise, we perform ridge regression a third time, on ranked personality scores and ranked `tf`-normalized features. This is exactly what Bachrach et. al. (2011)[2] for their Facebook features, and we want to see how this changes predictive accuracy, if any.

MSE and MAE are sensitive to the range of our labels, so we first normalize the raw personality scores to the range [0, 1], to make our results comparable. We only perform regression on `tf` features this time, as they have the most consistent performance in our classification task.

Again, we will review the metrics and method we are using, and then discuss our results in Section 7.

### 6.2.1 Review

Where $y$ is the true value, $\hat{y}$ is the predicted value, $\theta$ is the linear model's parameters, $x$ are the feature vectors, and $b$ is the bias term,

$$\hat{y} = \theta^T x + b \tag{11}$$

MSE and MAE are two of the most straightforward metrics. These two metrics are not ideal, as they do not capture the variance present in the data. We would expect labels with a larger spread to have higher MSE and MAE, simply because their variance is higher (compare with the equation given for $R^2$). this makes comparing models difficult with just MSE and MAE, especially when the models are trained and evaluated on different datasets where the variance is not always known. Secondly, these metrics are sensitive to the spread of the labels, so it is meaningless to compare with MSE and MAE scores without knowing the labels' ranges beforehand. Both MSE and MAE have the range of [0,$\infty$].

$$MSE = \frac{\sum (y - \hat{y})^2}{n} \tag{12}$$

$$MSE = \frac{\sum |y - \hat{y}|}{n} \tag{13}$$

The definition of root mean squared error (RMSE), is ambiguous, in that sometimes the square root extends to the denominator (Holmes 2000),[27] and sometimes it does not (Liu 2015).[11] It depends on whether the expression is about the mean of the square root of the squared distances, or simply the square root of the MSE. Hence where the formula is not declared explicitly, even with normalized data, the values could be off by a multiple of $\sqrt{n}$. Derivation of RMSE (whichever form) from MSE is trivial, so we stick to MSE so it's unambiguous.

The coefficient of determination, R2, is calculated as 1, minus the ratio, of the square of difference between true and predicted values, and the square of difference between true values and their mean. Intuitively speaking, R2 is the fraction of the variance in y that is explained by the variance in x. R2 has the range of [$-\infty$,1].

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \tag{14}$$

Correlation coefficients have the range [-1,1]. The Pearson correlation coefficient, $r$, is the centered cosine similarity between two sets of points $x$ and $y$. For our case, it's between $y$ and $\hat{y}$.

$$PEAR = r = \frac{\sum(y-\bar{y})(\hat{y}-\bar{\hat{y}})}{\sqrt{\sum(y-\bar{y})^2}\sqrt{(\hat{y}-\bar{\hat{y}})^2}} \tag{15}$$

The Spearman rank correlation coefficient, $\rho$, is equal to the Pearson correlation coefficient between ranked variables. After converting $y$ and $\hat{y}$ to ranked scores $y_r$ and $\hat{y}_r$,

$$SPEAR = \rho = 1 - \frac{6\sum(y_r-\hat{y}_r)^2}{n(n^2-1)} \tag{16}$$

*Caveat:*
The mean of correlation coefficients (like PEAR and SPEAR) across k-folds do not equal their value for the entire sample population. To estimate sample population correlation with subsample correlations, one must first perform Fisher's Z-transform (which is equivalent to arctanh) on the correlations, before summing them, and then reversing the transform (tanh) to get a better estimate (Corey 1998).[25] Even then, this is an estimate only, and for our population, we found this value to be a non-negligible overestimate. This result generalizes to the fact that the computed sample correlation is only a biased estimate of the true population correlation, since the sample population is a subset.

To avoid the entire issue with estimating sample population correlations, we did the exact opposite of what we did in the classification problem. Instead of computing values across individual folds, which, for calculating AUC and PR is necessary, (Forman 2010)[24] we only concatenate the predicted ys after training on each fold, and then compute all 5 of our regression metrics across all ys at once. We seek to maximize the correlation coefficient in our case.

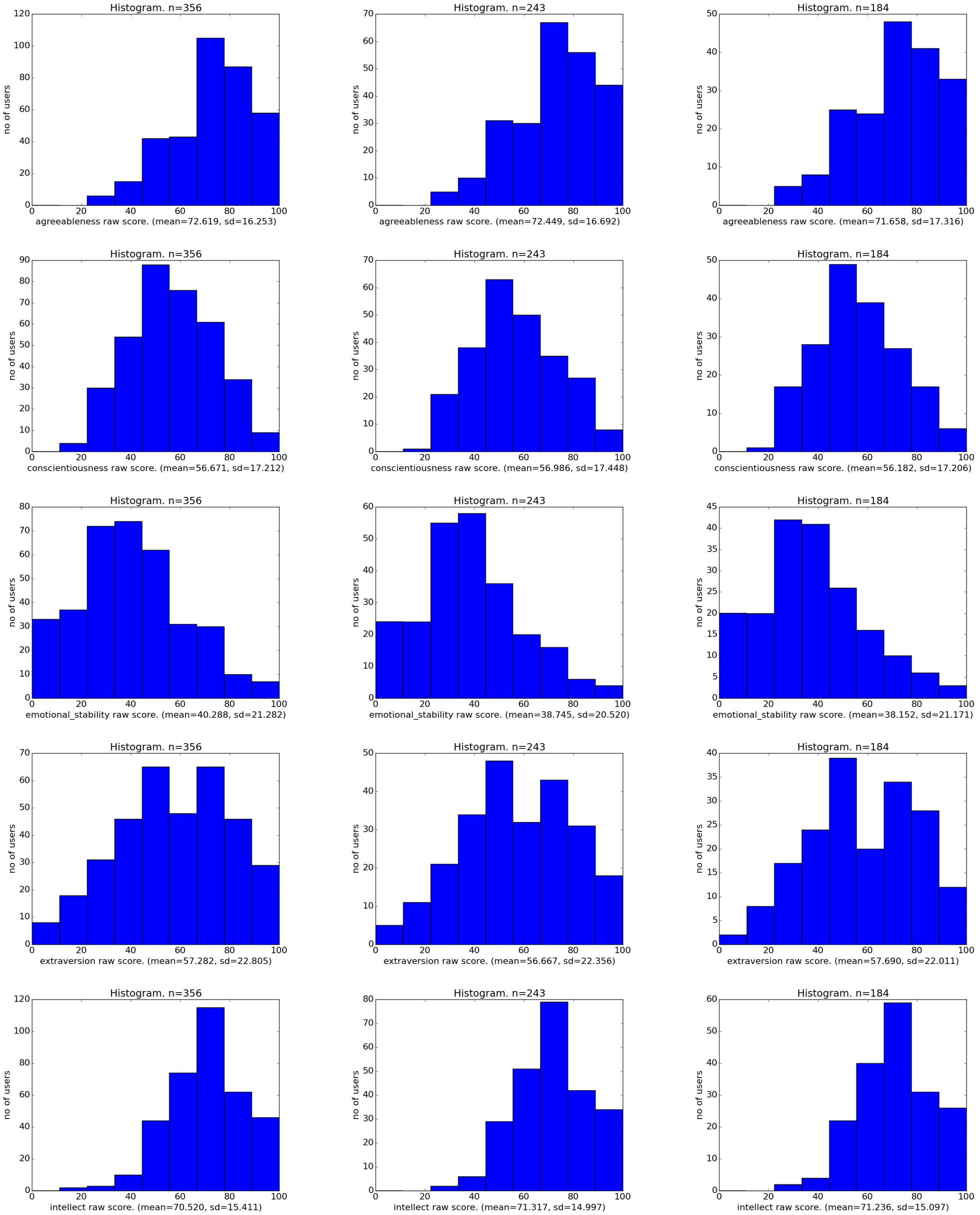
# 7 Results

(Results continued overleaf)

## 7.1 Personality score distribution

We plot histograms of users' personality scores, filtering for users with over 200 or 500 tweets, or with no filtering. The plots show that each personality score is independent of the number of tweets posted by the user.

All users | users with >200 tweets | users with >500 tweets



## 7.2 Classification by L1-regularized logistic regression into tritiles with actual values

The tables below show the results from classification. Users' personality scores have been transformed by first ranking them, and then segregating the rank scores into tritiles (3 quantiles).

These tables are dense as this is a multi-dimensional problem, and they need some explanation. Each row in the table corresponds to one metric, either AUC, AP, or F1. This is the metric that the models are maximizing for. Each column also corresponds to one metric, and this is how each one of those models evaluate, using all 3 metrics. Hence, the diagonal across the little 3x3 matrices always have the highest value in a column. This format allows us to compare the performance of models when optimizing for and evaluating with different metrics.

The rows in each table are grouped in threes - each such group is a set of results obtained using features normalized either with `tf`, `pt`, or `bm25`. Each table (overleaf) is the result of from training models on a subset of the entire feature set, i.e. whether we only use LWIC and MRC features, TweetNLP features, or all features at once.

Using Figure 5 to explain, each number is the micro-average of classification performance (evaluated using that metric) of the 3 binary classifiers trained by transforming the label into one-hot encodings. This is the thick blue line. We have 3 classifiers because we have 3 quantiles, and these correspond to the dashed lines in the chart. In actuality, we have 10 cross-validation folds, so each individual line is already the mean across all 10 validation folds, and the micro-average is the micro-averages of the means across 10 folds.

The mean column in the table then, corresponds to the mean of these micro-averages (for all 5 personality scores), when using the one set of training parameters (in this case, just the number of target PCA dimensions) that maximizes $\sum$score for all 5 personality attributes. This set of features corresponds to the maximum of the mean line in Figures 10, 11, and 12. This is the average performance if we are restricted to the same training parameters to train all 5 classifiers.

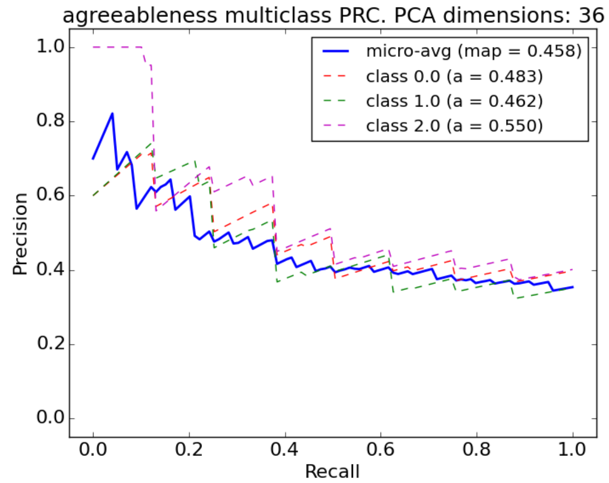



Figure 5: Explanatory PRC plot corresponding to the number at position (2,1) in the Figure 6. This plot is generated using a subset of features selected to maximize AUC of agreeableness.

| normal-<br>-ization | maxim-<br>-izing | agreeableness | | | conscientiousness | | | emotional stability | | | extraversion | | | intellect | | | mean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 |
| tf | AUC | 0.6081 | 0.4579 | 0.4167 | 0.6812 | 0.5039 | 0.6250 | 0.6530 | 0.4542 | 0.3750 | 0.6529 | 0.4811 | 0.3333 | 0.6080 | 0.4386 | 0.4167 | 0.6118 | 0.4462 | 0.3917 |
| | AP | 0.6081 | 0.4579 | 0.4167 | 0.6792 | 0.5153 | 0.6250 | 0.6430 | 0.4883 | 0.3750 | 0.6529 | 0.4811 | 0.3333 | 0.6080 | 0.4386 | 0.4167 | 0.6117 | 0.4474 | 0.4250 |
| | F1 | 0.5309 | 0.3619 | 0.5000 | 0.6792 | 0.5153 | 0.6250 | 0.6077 | 0.4442 | 0.5417 | 0.6077 | 0.4368 | 0.4583 | 0.5283 | 0.3608 | 0.5833 | 0.6117 | 0.4474 | 0.4250 |
| pt | AUC | 0.5972 | 0.4383 | 0.3333 | 0.6644 | 0.4875 | 0.5417 | 0.6346 | 0.4843 | 0.3750 | 0.6144 | 0.4163 | 0.3750 | 0.5960 | 0.4076 | 0.5000 | 0.6061 | 0.4386 | 0.4167 |
| | AP | 0.5962 | 0.4388 | 0.3333 | 0.6603 | 0.4897 | 0.4583 | 0.6346 | 0.4847 | 0.3750 | 0.5928 | 0.4345 | 0.3750 | 0.5931 | 0.4130 | 0.5000 | 0.6016 | 0.4399 | 0.3833 |
| | F1 | 0.5364 | 0.3603 | 0.5417 | 0.5909 | 0.4381 | 0.6250 | 0.5813 | 0.4222 | 0.5417 | 0.6010 | 0.3978 | 0.4583 | 0.5659 | 0.3876 | 0.5833 | 0.5707 | 0.4058 | 0.4500 |
| bm25 | AUC | 0.6162 | 0.4515 | 0.4167 | 0.6652 | 0.4795 | 0.4583 | 0.6265 | 0.4440 | 0.4167 | 0.5736 | 0.3919 | 0.2500 | 0.5813 | 0.3742 | 0.3333 | 0.5857 | 0.4182 | 0.4167 |
| | AP | 0.6162 | 0.4515 | 0.4167 | 0.6546 | 0.4984 | 0.5417 | 0.6216 | 0.4488 | 0.4583 | 0.5574 | 0.3996 | 0.2500 | 0.5547 | 0.3904 | 0.4583 | 0.5809 | 0.4232 | 0.4083 |
| | F1 | 0.5770 | 0.4188 | 0.5000 | 0.5928 | 0.4317 | 0.7083 | 0.5877 | 0.3991 | 0.5000 | 0.4870 | 0.3354 | 0.3333 | 0.5467 | 0.3809 | 0.5417 | 0.5406 | 0.3786 | 0.4417 |

Figure 6: Classification using LWIC + MRC + TweetNLP + misc features.

### 7.2.1 Table of results

Miscellaneous features are the simple counts of the following features, normalized by either `tf`, `pt`, or `bm25`: favourites, followers, friends, words they used in total, tweets they used in total, and all the other features we talked about in Section 3.5.5.

A naive predictor (i.e. random guesser) for 3 equally-sized classes have a baseline of AUC=$\frac{1}{3}$=0.333, F1=$\frac{1}{3}$=0.333, and AP=$\frac{4}{9}$=0.444.

| normal- -ization | maxim- -izing | agreeableness | | | conscientiousness | | | emotional stability | | | extraversion | | | intellect | | | mean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 |
| tf | AUC | 0.6081 | 0.4579 | 0.4167 | 0.6812 | 0.5039 | 0.6250 | 0.6530 | 0.4542 | 0.3750 | 0.6529 | 0.4811 | 0.3333 | 0.6080 | 0.4386 | 0.4167 | 0.6118 | 0.4462 | 0.3917 |
| | AP | 0.6081 | 0.4579 | 0.4167 | 0.6792 | 0.5153 | 0.6250 | 0.6430 | 0.4883 | 0.3750 | 0.6529 | 0.4811 | 0.3333 | 0.6080 | 0.4386 | 0.4167 | 0.6117 | 0.4474 | 0.4250 |
| | F1 | 0.5309 | 0.3619 | 0.5000 | 0.6792 | 0.5153 | 0.6250 | 0.6077 | 0.4442 | 0.5417 | 0.6077 | 0.4368 | 0.4583 | 0.5283 | 0.3608 | 0.5833 | 0.6117 | 0.4474 | 0.4250 |
| pt | AUC | 0.5972 | 0.4383 | 0.3333 | 0.6644 | 0.4875 | 0.5417 | 0.6346 | 0.4843 | 0.3750 | 0.6144 | 0.4163 | 0.3750 | 0.5960 | 0.4076 | 0.5000 | 0.6061 | 0.4386 | 0.4167 |
| | AP | 0.5962 | 0.4388 | 0.3333 | 0.6603 | 0.4897 | 0.4583 | 0.6346 | 0.4847 | 0.3750 | 0.5928 | 0.4345 | 0.3750 | 0.5931 | 0.4130 | 0.5000 | 0.6016 | 0.4399 | 0.3833 |
| | F1 | 0.5364 | 0.3603 | 0.5417 | 0.5909 | 0.4381 | 0.6250 | 0.5813 | 0.4222 | 0.5417 | 0.6010 | 0.3978 | 0.4583 | 0.5659 | 0.3876 | 0.5833 | 0.5707 | 0.4058 | 0.4500 |
| bm25 | AUC | 0.6162 | 0.4515 | 0.4167 | 0.6652 | 0.4795 | 0.4583 | 0.6265 | 0.4440 | 0.4167 | 0.5736 | 0.3919 | 0.2500 | 0.5813 | 0.3742 | 0.3333 | 0.5857 | 0.4182 | 0.4167 |
| | AP | 0.6162 | 0.4515 | 0.4167 | 0.6546 | 0.4984 | 0.5417 | 0.6216 | 0.4488 | 0.4583 | 0.5574 | 0.3996 | 0.2500 | 0.5547 | 0.3904 | 0.4583 | 0.5809 | 0.4232 | 0.4083 |
| | F1 | 0.5770 | 0.4188 | 0.5000 | 0.5928 | 0.4317 | 0.7083 | 0.5877 | 0.3991 | 0.5000 | 0.4870 | 0.3354 | 0.3333 | 0.5467 | 0.3809 | 0.5417 | 0.5406 | 0.3786 | 0.4417 |

Figure 7: Classification using LWIC + MRC + TweetNLP + misc features. (duplicated from Figure 6 for convenience)

| normal- -ization | maxim- -izing | agreeableness | | | conscientiousness | | | emotional stability | | | extraversion | | | intellect | | | mean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 |
| tf | AUC | 0.5835 | 0.4194 | 0.2917 | 0.6892 | 0.5217 | 0.5833 | 0.6487 | 0.4805 | 0.3750 | 0.5712 | 0.4201 | 0.2500 | 0.5847 | 0.4109 | 0.4167 | 0.6110 | 0.4504 | 0.3833 |
| | AP | 0.5822 | 0.4236 | 0.3333 | 0.6892 | 0.5217 | 0.5833 | 0.6423 | 0.4863 | 0.4167 | 0.5708 | 0.4245 | 0.2500 | 0.5737 | 0.4275 | 0.4583 | 0.6110 | 0.4504 | 0.3833 |
| | F1 | 0.5779 | 0.4100 | 0.5000 | 0.6658 | 0.4941 | 0.6667 | 0.6191 | 0.4391 | 0.4583 | 0.4979 | 0.3468 | 0.4583 | 0.5536 | 0.4048 | 0.4583 | 0.5577 | 0.4016 | 0.4333 |
| pt | AUC | 0.6086 | 0.4451 | 0.2083 | 0.6575 | 0.4828 | 0.5417 | 0.6381 | 0.4924 | 0.3333 | 0.5395 | 0.3967 | 0.4167 | 0.5488 | 0.4121 | 0.4167 | 0.5856 | 0.4290 | 0.3583 |
| | AP | 0.6023 | 0.4493 | 0.2083 | 0.6427 | 0.4867 | 0.5417 | 0.6256 | 0.5028 | 0.3333 | 0.5395 | 0.3967 | 0.4167 | 0.5488 | 0.4121 | 0.4167 | 0.5769 | 0.4319 | 0.3667 |
| | F1 | 0.5752 | 0.4241 | 0.5417 | 0.6270 | 0.4523 | 0.6667 | 0.6161 | 0.4460 | 0.4167 | 0.5395 | 0.3967 | 0.4167 | 0.5417 | 0.4019 | 0.5833 | 0.5682 | 0.4189 | 0.4583 |
| bm25 | AUC | 0.5787 | 0.4173 | 0.3750 | 0.6210 | 0.4395 | 0.5000 | 0.6529 | 0.4895 | 0.4167 | 0.4857 | 0.3284 | 0.2083 | 0.5470 | 0.3989 | 0.5000 | 0.5724 | 0.4123 | 0.3500 |
| | AP | 0.5664 | 0.4281 | 0.3750 | 0.6083 | 0.4433 | 0.5000 | 0.6529 | 0.4895 | 0.4167 | 0.4721 | 0.3307 | 0.1667 | 0.5424 | 0.4046 | 0.3750 | 0.5710 | 0.4126 | 0.3333 |
| | F1 | 0.5755 | 0.4217 | 0.4167 | 0.6140 | 0.4258 | 0.6667 | 0.6169 | 0.4427 | 0.4583 | 0.4251 | 0.3037 | 0.2500 | 0.5279 | 0.3948 | 0.6250 | 0.5488 | 0.3985 | 0.4083 |

Figure 8: Classification using LWIC + MRC + misc features only.

| normal- -ization | maxim- -izing | agreeableness | | | conscientiousness | | | emotional stability | | | extraversion | | | intellect | | | mean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 | AUC | AP | F1 |
| tf | AUC | 0.5834 | 0.4240 | 0.3333 | 0.6785 | 0.5023 | 0.6250 | 0.6346 | 0.4327 | 0.3333 | 0.5992 | 0.4342 | 0.4167 | 0.5994 | 0.4288 | 0.5000 | 0.6034 | 0.4282 | 0.3917 |
| | AP | 0.5760 | 0.4291 | 0.3750 | 0.6641 | 0.5093 | 0.3333 | 0.6244 | 0.4630 | 0.4167 | 0.5805 | 0.4406 | 0.2917 | 0.5994 | 0.4288 | 0.5000 | 0.6021 | 0.4369 | 0.4333 |
| | F1 | 0.5124 | 0.3496 | 0.4583 | 0.6785 | 0.5023 | 0.6250 | 0.6070 | 0.4486 | 0.5417 | 0.5578 | 0.3848 | 0.5417 | 0.5373 | 0.3761 | 0.5833 | 0.5657 | 0.4039 | 0.4417 |
| pt | AUC | 0.5963 | 0.4195 | 0.2917 | 0.6654 | 0.4932 | 0.5000 | 0.6282 | 0.4339 | 0.5417 | 0.6219 | 0.4359 | 0.4167 | 0.5857 | 0.4122 | 0.4167 | 0.6075 | 0.4400 | 0.4333 |
| | AP | 0.5944 | 0.4314 | 0.3750 | 0.6642 | 0.4953 | 0.5000 | 0.6215 | 0.4795 | 0.4167 | 0.6163 | 0.4447 | 0.3333 | 0.5857 | 0.4122 | 0.4167 | 0.6075 | 0.4400 | 0.4333 |
| | F1 | 0.5075 | 0.3750 | 0.5417 | 0.6562 | 0.4845 | 0.6250 | 0.5955 | 0.4345 | 0.5417 | 0.5136 | 0.3705 | 0.4167 | 0.5458 | 0.3666 | 0.5833 | 0.5591 | 0.3950 | 0.4500 |
| bm25 | AUC | 0.6152 | 0.4489 | 0.3333 | 0.6606 | 0.4714 | 0.4583 | 0.6265 | 0.4368 | 0.3333 | 0.5558 | 0.3837 | 0.2500 | 0.6298 | 0.4289 | 0.3750 | 0.5887 | 0.4225 | 0.4167 |
| | AP | 0.6130 | 0.4563 | 0.3750 | 0.6464 | 0.4893 | 0.6250 | 0.6265 | 0.4523 | 0.3750 | 0.5410 | 0.3910 | 0.2500 | 0.6238 | 0.4317 | 0.3333 | 0.5879 | 0.4247 | 0.4333 |
| | F1 | 0.5956 | 0.4371 | 0.5000 | 0.5842 | 0.4217 | 0.7083 | 0.5848 | 0.3949 | 0.5000 | 0.4889 | 0.3367 | 0.2917 | 0.4755 | 0.3410 | 0.5417 | 0.5445 | 0.3800 | 0.4417 |

Figure 9: Classification using TweetNLP + misc features only.

### 7.2.2 Plots

Please see overleaf for descriptions and explanations of these plots.

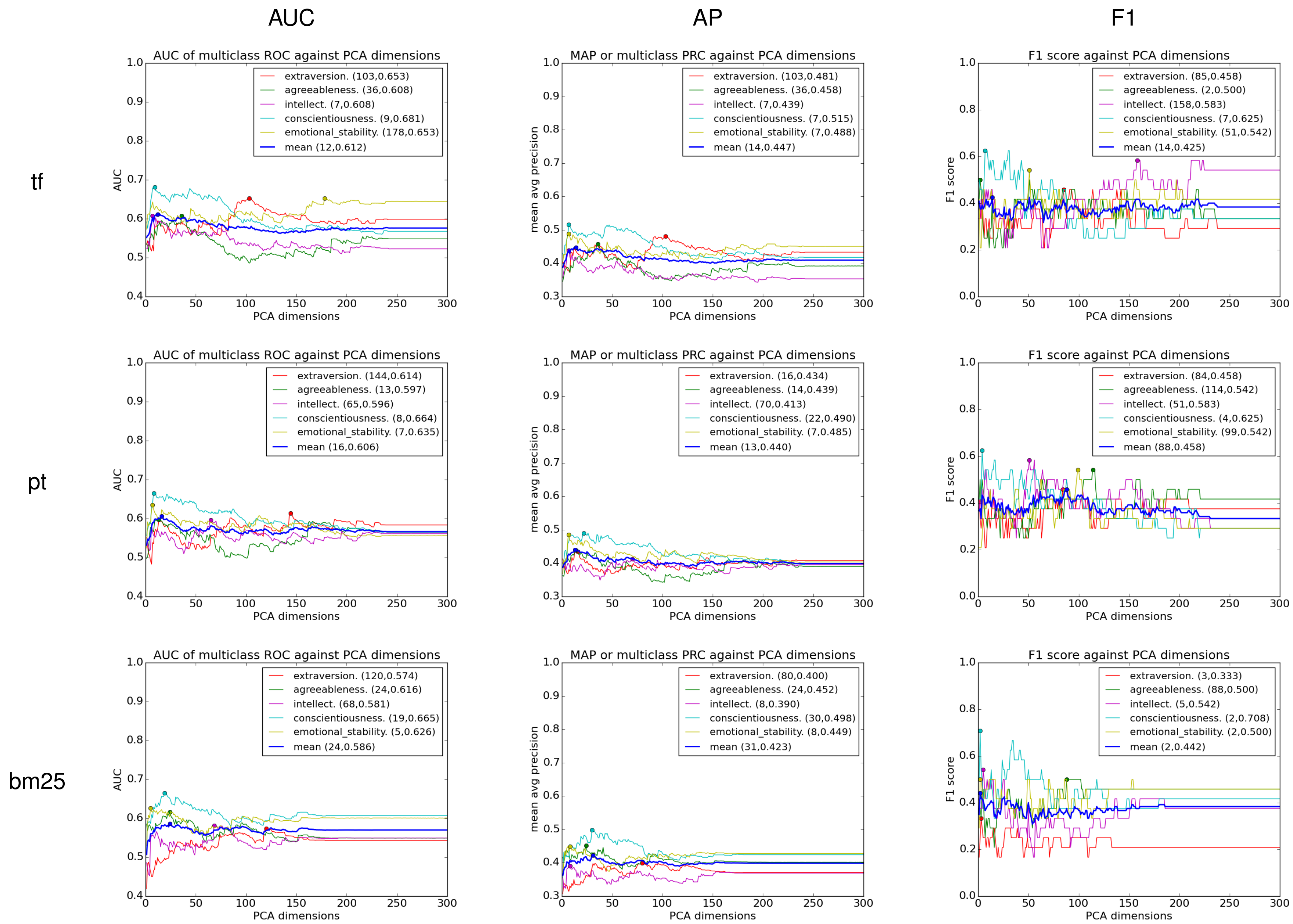



Figure 10: AUC, AP, and F1 scores against PCA dimensions, using LWIC + MRC + TweetNLP + misc features.

Figures 10, 11, and 12 show how the AUC, AP, and F1 scores change with number of target PCA dimensions, for each personality score. The points highlighted are the maximal values for each respective line, so they show the number of target PCA dimensions which produce the values given in Figures 7, 8, and 9. The flat tail end of each graph is either caused by the number of target PCA dimensions exceeding the number of training points (n), in which case it is reduced to n, or the features beyond a point have been zeroed out to remove them (Figures 11 and 12).

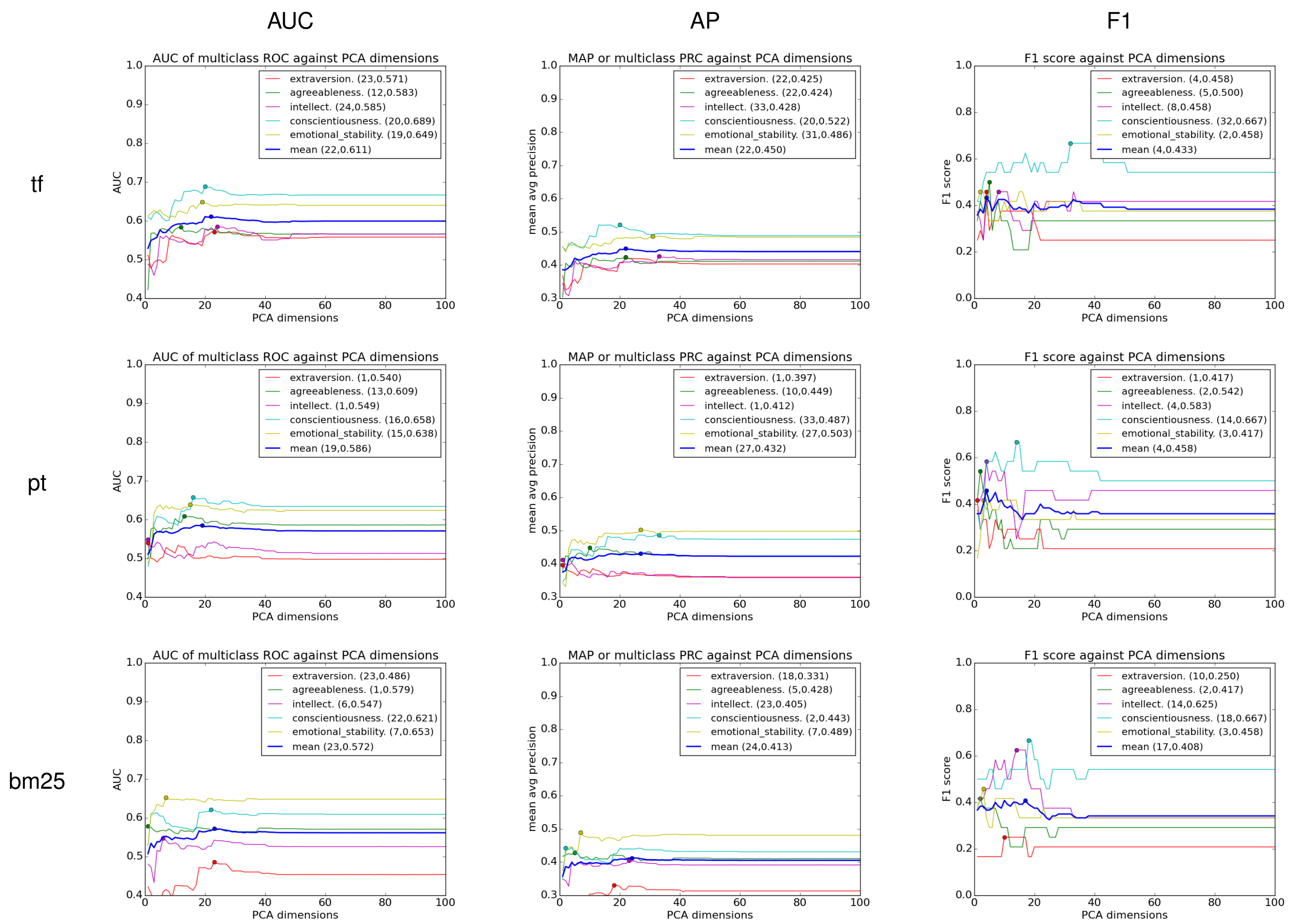



Figure 11: AUC, AP, and F1 scores against PCA dimensions, Classification using LWIC + MRC + misc features only.

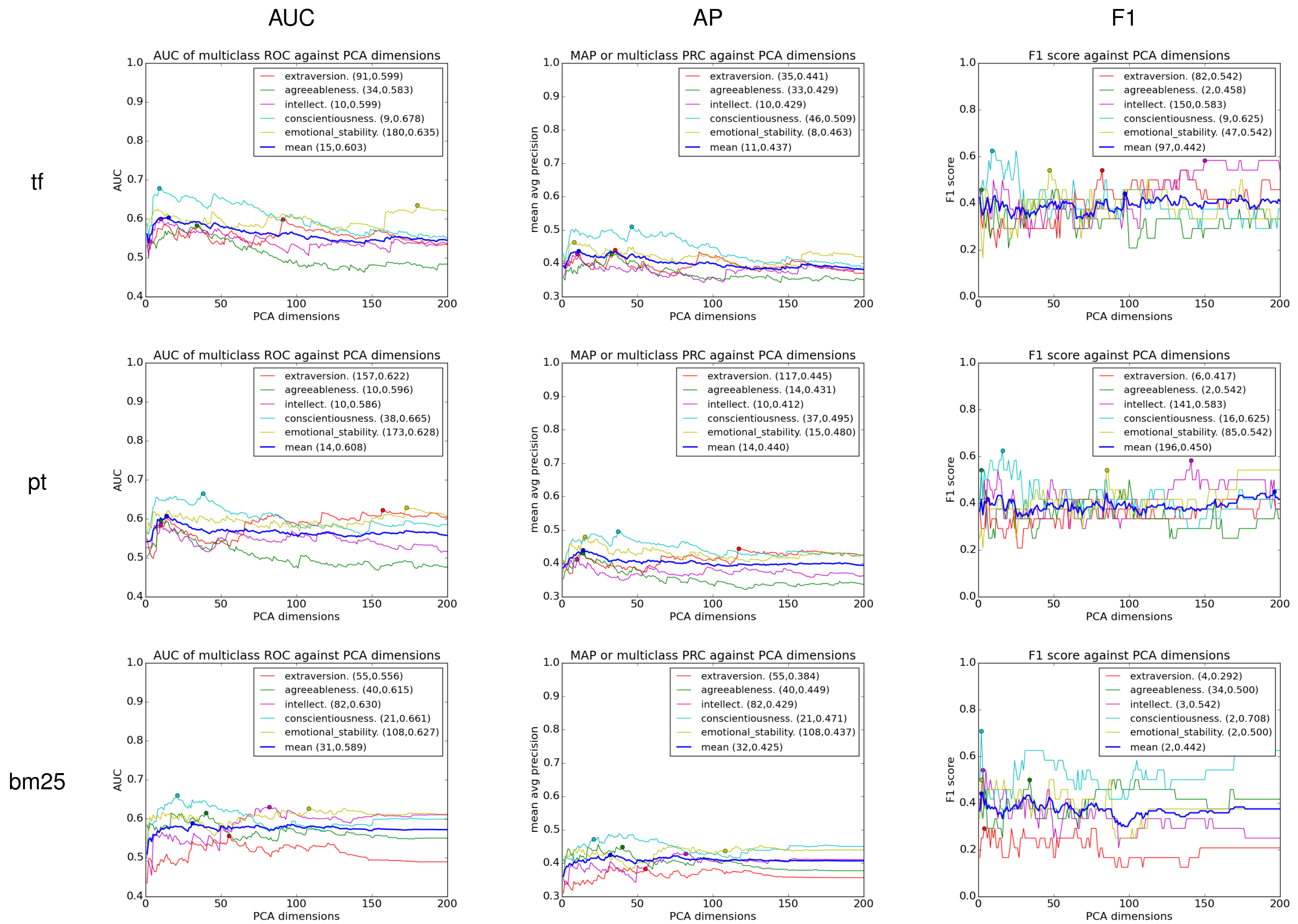



Figure 12: AUC, AP, and F1 scores against PCA dimensions, Classification using TweetNLP + misc features only.

Here, we plot the original ROC and PR curves for each personality score, since AUC, AP, and F1 scores used in Figures 10, 11, and 12. The ROC curve is generated using the features that maximizes AUC, for ROC it's using those that maximizes AP. For brevity, we only plot for `tf`-transformed features from the full feature set. Hence these correspond to the values of the diagonals in the first row of Figure 10, except for F1 scores and the mean. Class 0 is the lowest quantile, and class 2 is the highest quantile. The dashed lines are the performance of binary classifiers trained on data transformed into one-hot encodings for that class.

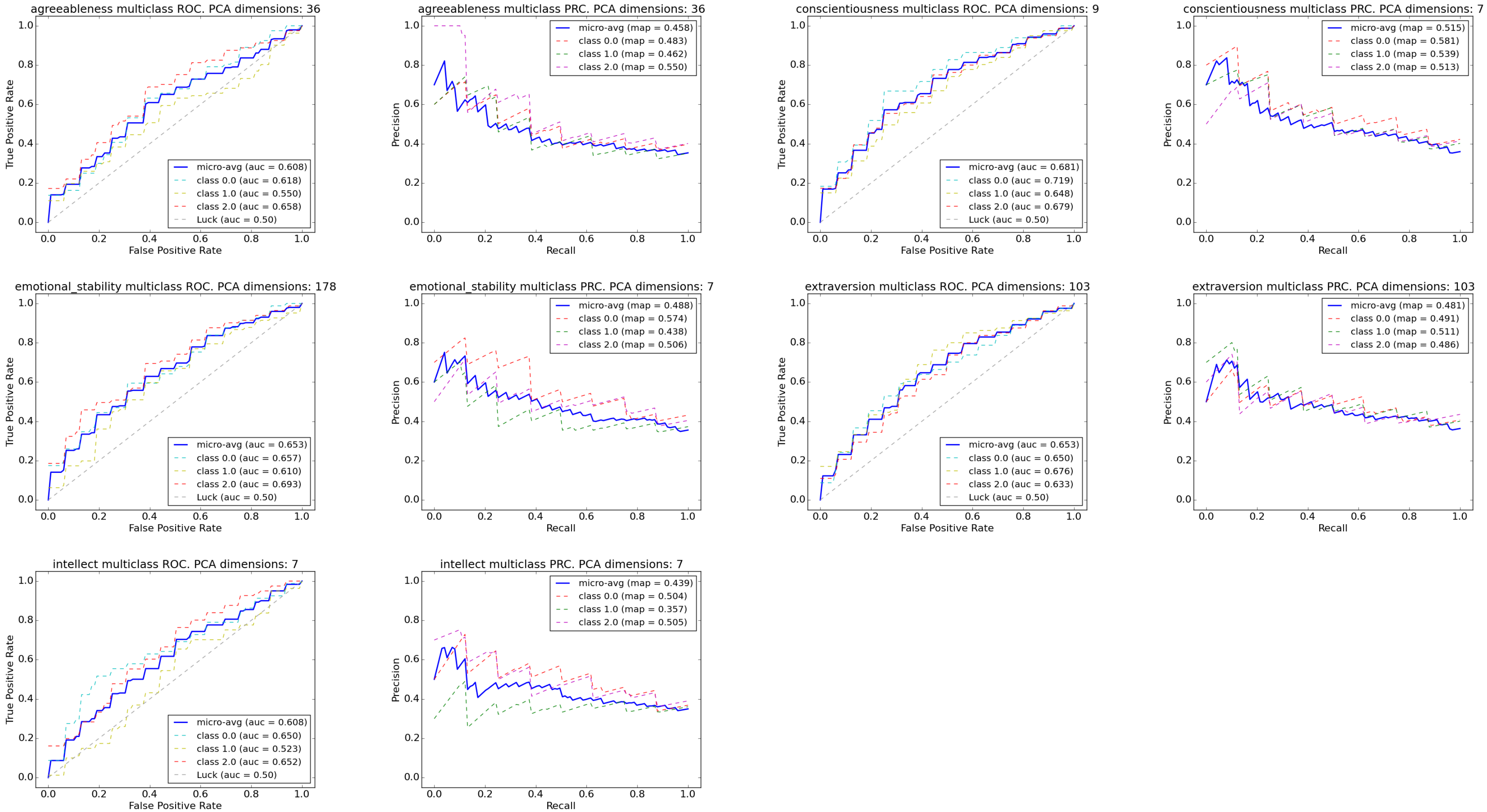
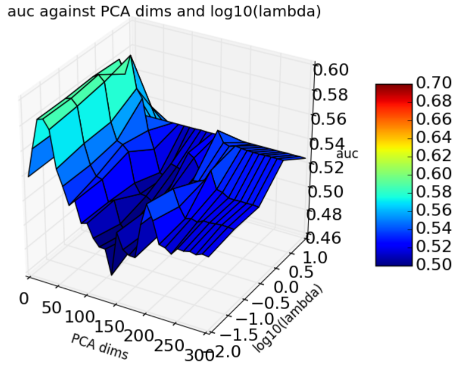
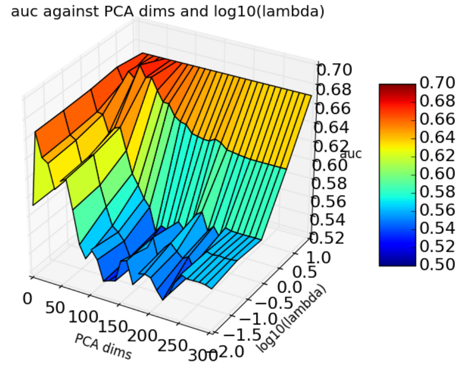



Figure 13: ROC and AUC curves for models trained using all features.

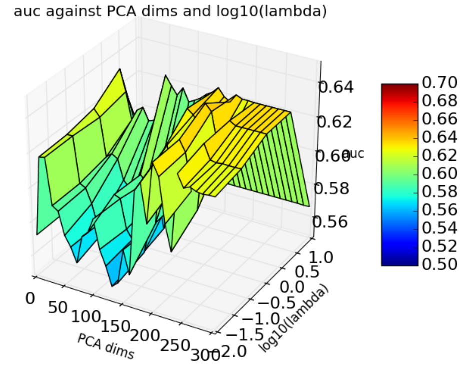
These plots are lower-sampling-resolution versions of the plots in Figure 10, but performed across a range of values for $\lambda$, the L1 regularization strength for logistic regression. The z-axis (vertical axis) is the AUC score. The features used here are the `tf` ones. AP and F1 plots were omitted for brevity. There plots are to look for an optimal (or close to optimal) value for $\lambda$. Notice that the maximal point is always at $\lambda = 1$.



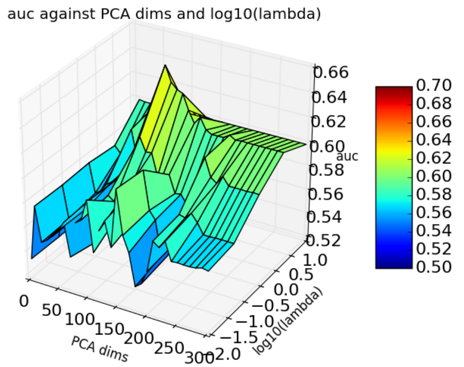

Agreeableness



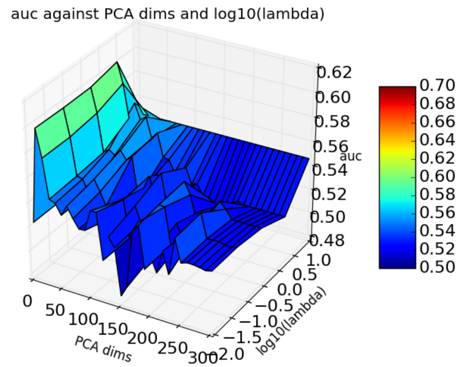

Conscientiousness




Emotional stability




Extraversion




Intellect

Figure 14: AUC score against number of dimensions after PCA and $log_{10}(\lambda)$, using all features

## 7.3 Discussion (classification)

In this section, unless indicated otherwise, assume that we we are discussing the results obtained using the full set of features (LWIC + MRC + TweetNLP + misc).

From Figures 7, 8, and 9, we hoped to discern a pattern about which statistic is better for evaluating the classification problem (classification across tritiles of personality scores), but as we can see from the tables, there is no clear pattern observed.

However, by taking the mean of the values across each 3x3 matrix, we can judge the relative accuracy of classification using all 3. For example, using `tf` features, classification for conscientiousness is better than for intellect, as all 9 numbers are higher. That is saying no matter which metric we choose to maximize, using $\lambda = 1$ as the regularization parameter, we can expect classification of conscientiousness to be always better than for extraversion.

By comparing the results across different data normalization methods, then, (`tf`, `pt`, and `bm25`), we can conclude that `pt` features (raw counts divided by the user's number of tweets) perform worse in classification across tritiles of personality scores than `tf` features for almost all cases, across all subsets of data (LWIC + MRC, TweetNLP, or all). `bm25` features have comparable performance to `tf` features. However, it performs consistently worse for extraversion.

`tf` features perform best most consistently amongst the 3 tested normalization methods, across different feature subsets when classifying based on quantile labels for personality scores.

Comparing Figures 8 and 9, using `tf` features, we make the observation that classification performance using only TweetNLP features is comparable to that when using LWIC + MRC alone.

However, comparing Figures 7 and 8, we observe that adding TweetNLP features to the LWIC + MRC feature set improves classification performance for intellect, and emotional stability when optimizing for and evaluating with F1 score. For extraversion and conscientiousness, the result is mixed, as the metrics do not agree.

From Figure 10, we can see why extraversion benefits so much from TweetNLP features; in the upper left figure, there is a marked peak at about 100 features, whereas the other features have more or less stable peaks at various intervals, or one at close to left edge of the chart. This suggests that extraversion is a concept that requires many more features to capture than the others.

Figure 14 shows that $\lambda = 1$ is optimal for AUC, on `tf` features, when classifying using all features. Based on the results in Figures 10, 11, and 12, we see that the peaks and troughs of the plots agree (across a row of plots), so it suggests that $\lambda = 1$ is also optimal (or close to optimal) for when we are maximizing AP or F1 instead of AUC.

What the plots do not tell us is whether $\lambda = 1$ is optimal outside of `tf` features. We would argue that there would be little reason to not be running on the full set of features at any given time (reducing dimensionality by PCA if necessary), and that we have established the fact that `tf` features are the ones that perform most consistently. Of course, we only ran it for half powers of 10 for $\lambda$, so there is still room for fine tuning around 1.0, although this value may well deviate if features are added or removed. In a similar line of thought, the `bm25` features were generated with default parameter values $b = 0.75$, $k_1 = 1.6$ and $\delta = 1$ for the function, and may benefit from fine-tuning these values. In our regression task, we sample 20 values of $\lambda$, to push this set of features closer to its performance limit.

## 7.4 Ridge regression with actual values (>200 tweets)

### 7.4.1 Table of results

MSE and MAE are calculated over labels normalized to the range [0, 1]. `tf`-normalized features used. Correlation metrics of naive predictors are undefined as variance of predictions is zero. See Section 7.2 for descriptions on how to read the tables.

| | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| NAIVE | N/A | N/A | N/A | 0.053 | 0.188 | N/A | N/A | N/A | 0.048 | 0.179 | N/A | N/A | N/A | 0.047 | 0.171 | N/A | N/A | N/A | 0.053 | 0.191 | N/A | N/A | N/A | 0.037 | 0.156 |

Figure 15: Baseline performance of naive mean-predictor.

| optim-<br>-izing | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.265 | 0.255 | 0.053 | 0.050 | 0.183 | 0.385 | 0.355 | 0.148 | 0.041 | 0.162 | 0.468 | 0.465 | 0.216 | 0.037 | 0.153 | 0.273 | 0.261 | 0.072 | 0.049 | 0.180 | 0.399 | 0.381 | 0.154 | 0.032 | 0.145 |
| SPEAR | 0.242 | 0.263 | -0.650 | 0.087 | 0.231 | 0.384 | 0.357 | 0.146 | 0.041 | 0.163 | 0.466 | 0.470 | 0.217 | 0.037 | 0.153 | 0.269 | 0.267 | 0.039 | 0.051 | 0.181 | 0.398 | 0.383 | 0.153 | 0.032 | 0.144 |
| R2 | 0.257 | 0.248 | 0.066 | 0.050 | 0.182 | 0.385 | 0.355 | 0.148 | 0.041 | 0.162 | 0.466 | 0.470 | 0.217 | 0.037 | 0.153 | 0.270 | 0.256 | 0.073 | 0.049 | 0.181 | 0.398 | 0.379 | 0.158 | 0.032 | 0.144 |
| MSE | 0.257 | 0.248 | 0.066 | 0.050 | 0.182 | 0.385 | 0.355 | 0.148 | 0.041 | 0.162 | 0.466 | 0.470 | 0.217 | 0.037 | 0.153 | 0.270 | 0.256 | 0.073 | 0.049 | 0.181 | 0.398 | 0.379 | 0.158 | 0.032 | 0.144 |
| MAE | 0.257 | 0.248 | 0.066 | 0.050 | 0.182 | 0.370 | 0.338 | 0.126 | 0.042 | 0.161 | 0.466 | 0.464 | 0.214 | 0.037 | 0.152 | 0.273 | 0.261 | 0.072 | 0.049 | 0.180 | 0.398 | 0.383 | 0.153 | 0.032 | 0.144 |

Figure 16: Regression using `tf`-normalized LWIC + MRC + TweetNLP + misc features. (users with >200 tweets)

| optim-<br>-izing | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.215 | 0.239 | 0.012 | 0.052 | 0.183 | 0.367 | 0.357 | 0.086 | 0.043 | 0.168 | 0.445 | 0.466 | 0.194 | 0.038 | 0.156 | 0.199 | 0.157 | -0.006 | 0.053 | 0.189 | 0.342 | 0.340 | 0.059 | 0.035 | 0.151 |
| SPEAR | 0.212 | 0.245 | 0.007 | 0.053 | 0.184 | 0.346 | 0.362 | -0.038 | 0.049 | 0.176 | 0.445 | 0.466 | 0.194 | 0.038 | 0.156 | 0.196 | 0.173 | 0.025 | 0.051 | 0.185 | 0.341 | 0.346 | 0.090 | 0.034 | 0.149 |
| R2 | 0.206 | 0.219 | 0.039 | 0.051 | 0.183 | 0.349 | 0.347 | 0.114 | 0.042 | 0.167 | 0.444 | 0.464 | 0.196 | 0.038 | 0.156 | 0.184 | 0.166 | 0.029 | 0.051 | 0.186 | 0.334 | 0.342 | 0.107 | 0.033 | 0.148 |
| MSE | 0.206 | 0.219 | 0.039 | 0.051 | 0.183 | 0.349 | 0.347 | 0.114 | 0.042 | 0.167 | 0.444 | 0.464 | 0.196 | 0.038 | 0.156 | 0.184 | 0.166 | 0.029 | 0.051 | 0.186 | 0.334 | 0.342 | 0.107 | 0.033 | 0.148 |
| MAE | 0.211 | 0.229 | 0.033 | 0.051 | 0.182 | 0.359 | 0.353 | 0.112 | 0.042 | 0.166 | 0.439 | 0.446 | 0.193 | 0.038 | 0.156 | 0.196 | 0.172 | 0.025 | 0.051 | 0.185 | 0.322 | 0.332 | 0.103 | 0.034 | 0.148 |

Figure 17: Regression using `tf`-normalized LWIC + MRC + misc features only. (users with >200 tweets)

| optim-<br>-izing | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.241 | 0.226 | 0.052 | 0.050 | 0.183 | 0.387 | 0.343 | 0.150 | 0.040 | 0.162 | 0.453 | 0.452 | 0.195 | 0.038 | 0.155 | 0.263 | 0.246 | 0.067 | 0.049 | 0.181 | 0.415 | 0.393 | 0.170 | 0.031 | 0.143 |
| SPEAR | 0.202 | 0.248 | 0.023 | 0.052 | 0.185 | 0.384 | 0.357 | 0.141 | 0.041 | 0.165 | 0.453 | 0.455 | 0.204 | 0.037 | 0.154 | 0.252 | 0.280 | -0.003 | 0.053 | 0.185 | 0.404 | 0.395 | 0.160 | 0.031 | 0.144 |
| R2 | 0.237 | 0.220 | 0.056 | 0.050 | 0.183 | 0.387 | 0.343 | 0.150 | 0.040 | 0.162 | 0.453 | 0.455 | 0.204 | 0.037 | 0.154 | 0.259 | 0.241 | 0.067 | 0.049 | 0.182 | 0.415 | 0.393 | 0.170 | 0.031 | 0.143 |
| MSE | 0.237 | 0.220 | 0.056 | 0.050 | 0.183 | 0.387 | 0.343 | 0.150 | 0.040 | 0.162 | 0.453 | 0.455 | 0.204 | 0.037 | 0.154 | 0.259 | 0.241 | 0.067 | 0.049 | 0.182 | 0.415 | 0.393 | 0.170 | 0.031 | 0.143 |
| MAE | 0.235 | 0.216 | 0.055 | 0.050 | 0.182 | 0.383 | 0.341 | 0.141 | 0.041 | 0.162 | 0.453 | 0.455 | 0.204 | 0.037 | 0.154 | 0.263 | 0.246 | 0.067 | 0.049 | 0.181 | 0.415 | 0.393 | 0.170 | 0.031 | 0.143 |

Figure 18: Regression using `tf`-normalized TweetNLP + misc features only. (users with >200 tweets)

### 7.4.2 Plots

These are plots of predicted (y-axis) against actual personality score (x-axis), using the best training parameters (minimizes MSE, MAE; maximizes PEAR, SPEAR, R2). The black diagonal line is how the dots should line up with an ideal predictor, and the black horizontal line around where most of the lines are clustered is the mean of that personality score in the actual data. Correlation-based metrics are in shades of red, whereas distance-based metrics are in shades of blue. The colored lines are the lines of least square error fitted on the chart, and the $m$ value is the gradient of this line. We can see that evaluating with distance-based metrics often gives us a model that predicts the mean score, whereas models optimized using correlation-based metrics have greater uncertainty in their predictions.

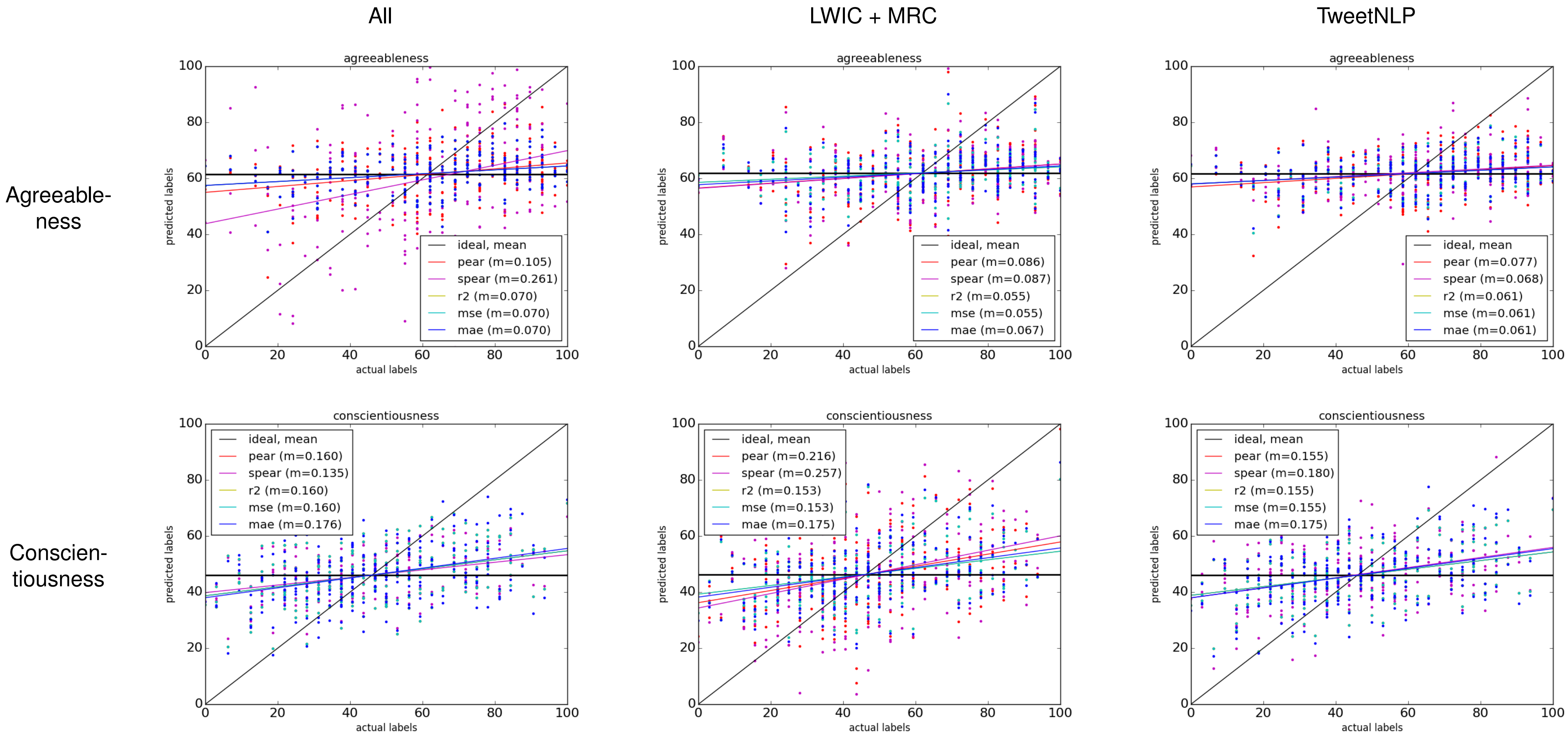



Figure 19: Plot of predicted (y-axis) against actual personality score (x-axis)

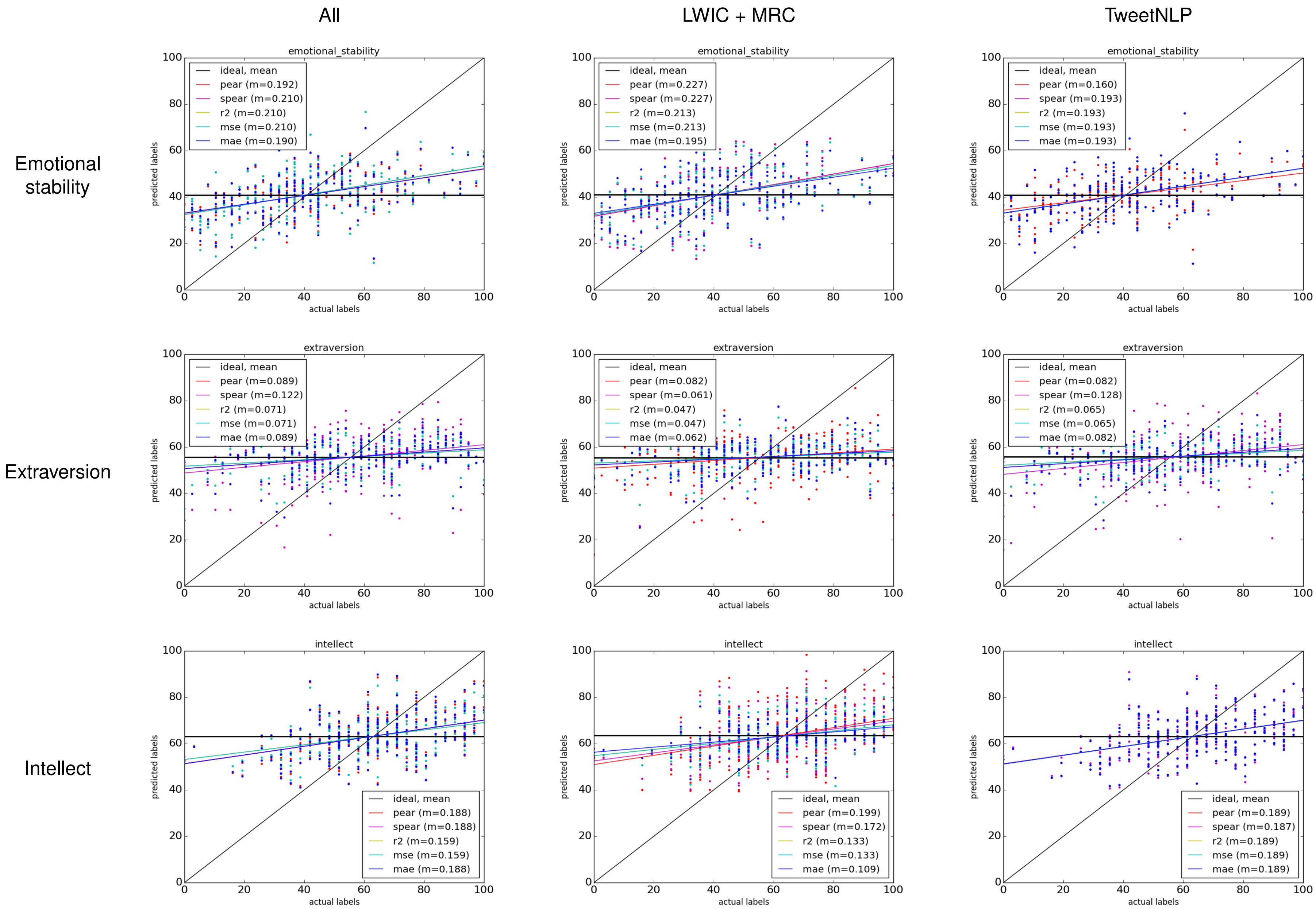



Figure 20: (continued from Figure 30) Plot of predicted (y-axis) against actual personality score (x-axis)

We plot the various models' performances against PCA dimensions and $\lambda$, optimizing for each metric individually. For brevity, we only included plots for PEAR, R2, and MSE, as PEAR/SPEAR, and MSE/MAE plots are very similar. Notice that optimal $\lambda$ and $d$ tend to be quite close regardless of metric. The blue lines intersecting it extend the entire range of the 3 axes. These are models trained with all actual personality scores and features (as opposed to rank). The 'flat' end of the graphs at the highest PCA dimensions correspond to the 'flat' ends in Figure 11, when we set target PCA dimensions exceeding number of training points (n) we have - in which case it reduces to n dimensions.

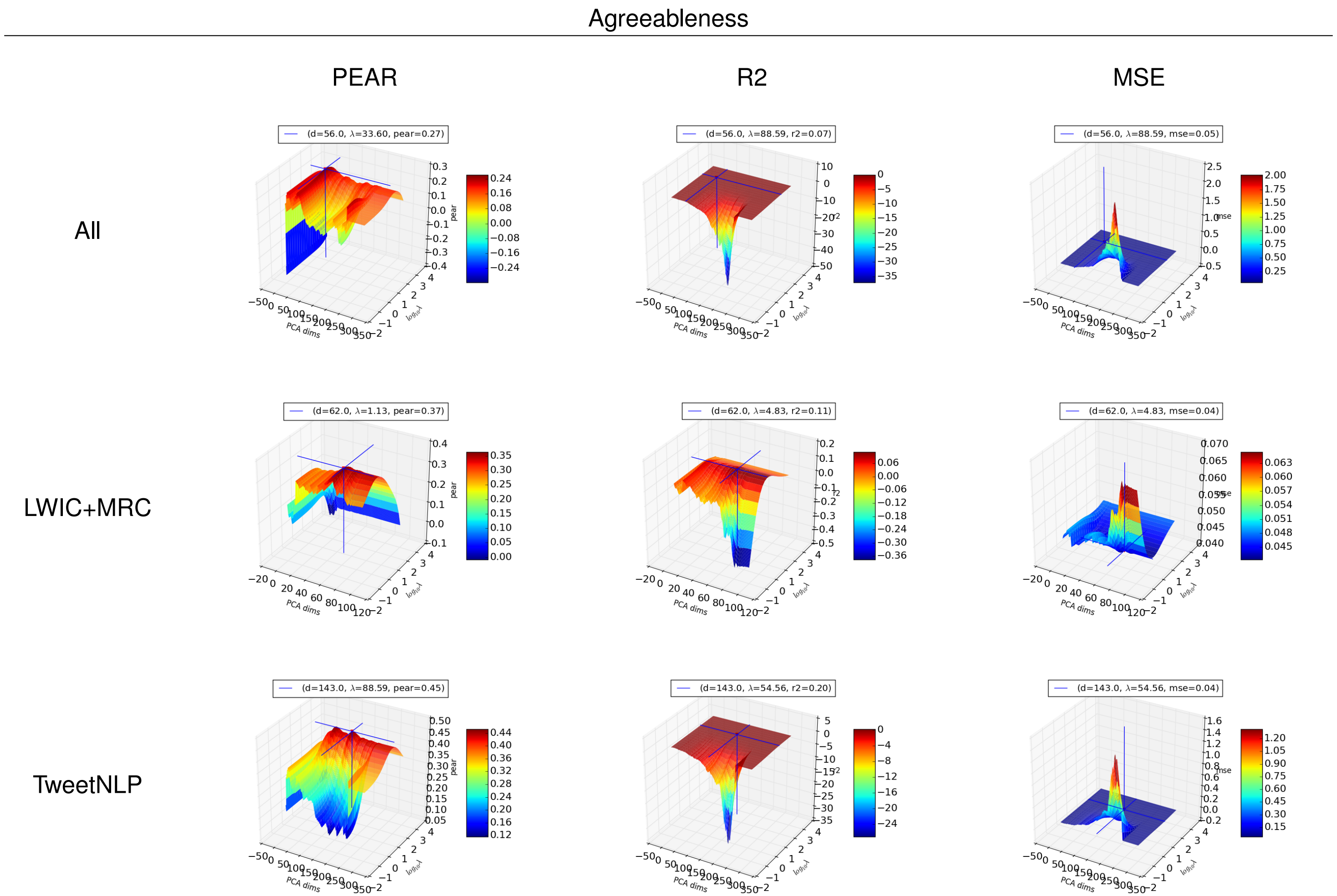



Figure 21: Plot of optimal point for agreeableness

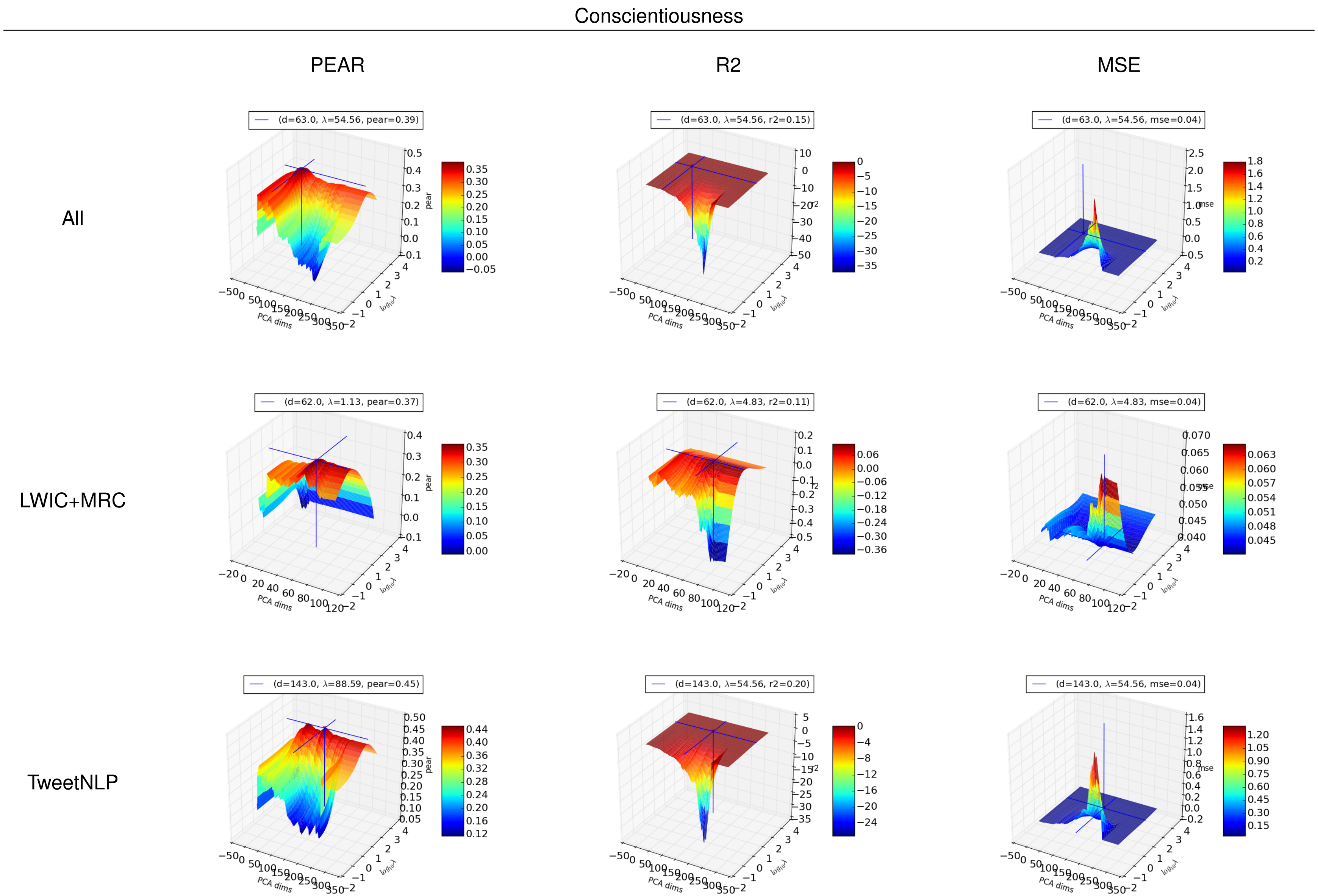



Figure 22: Plot of optimal point for conscientiousness

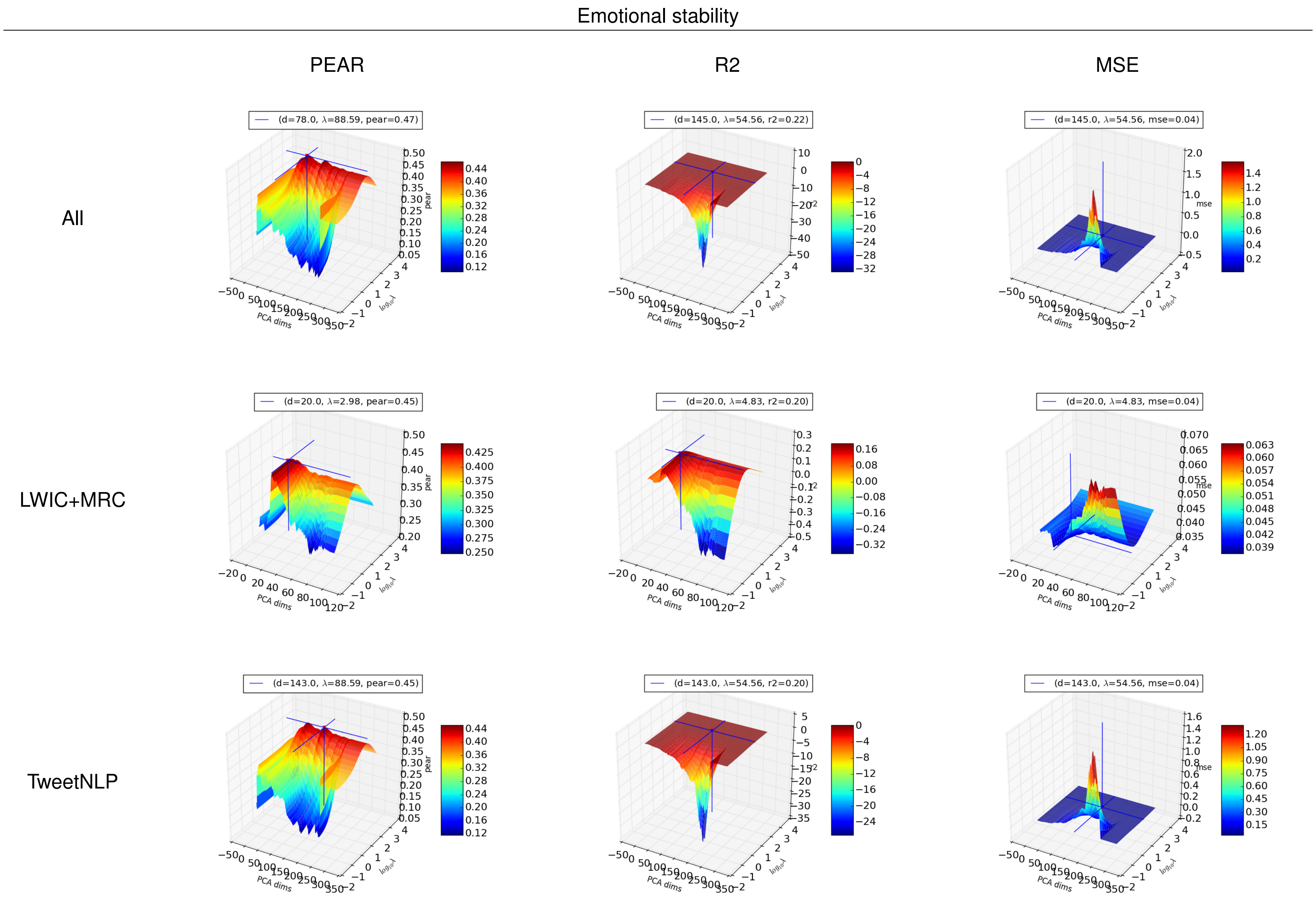



Figure 23: Plot of optimal point for emotional stability

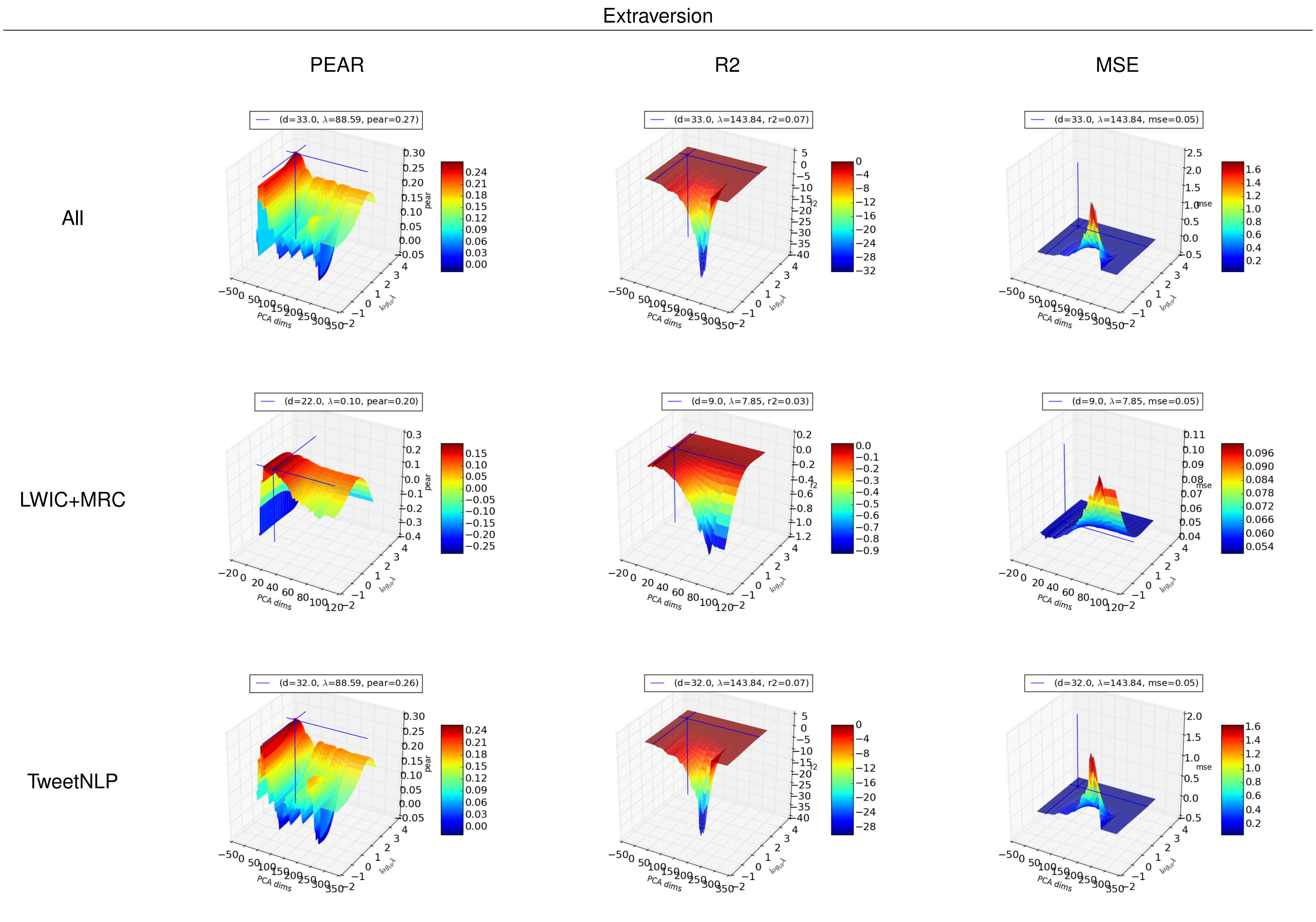



Figure 24: Plot of optimal point for extraversion

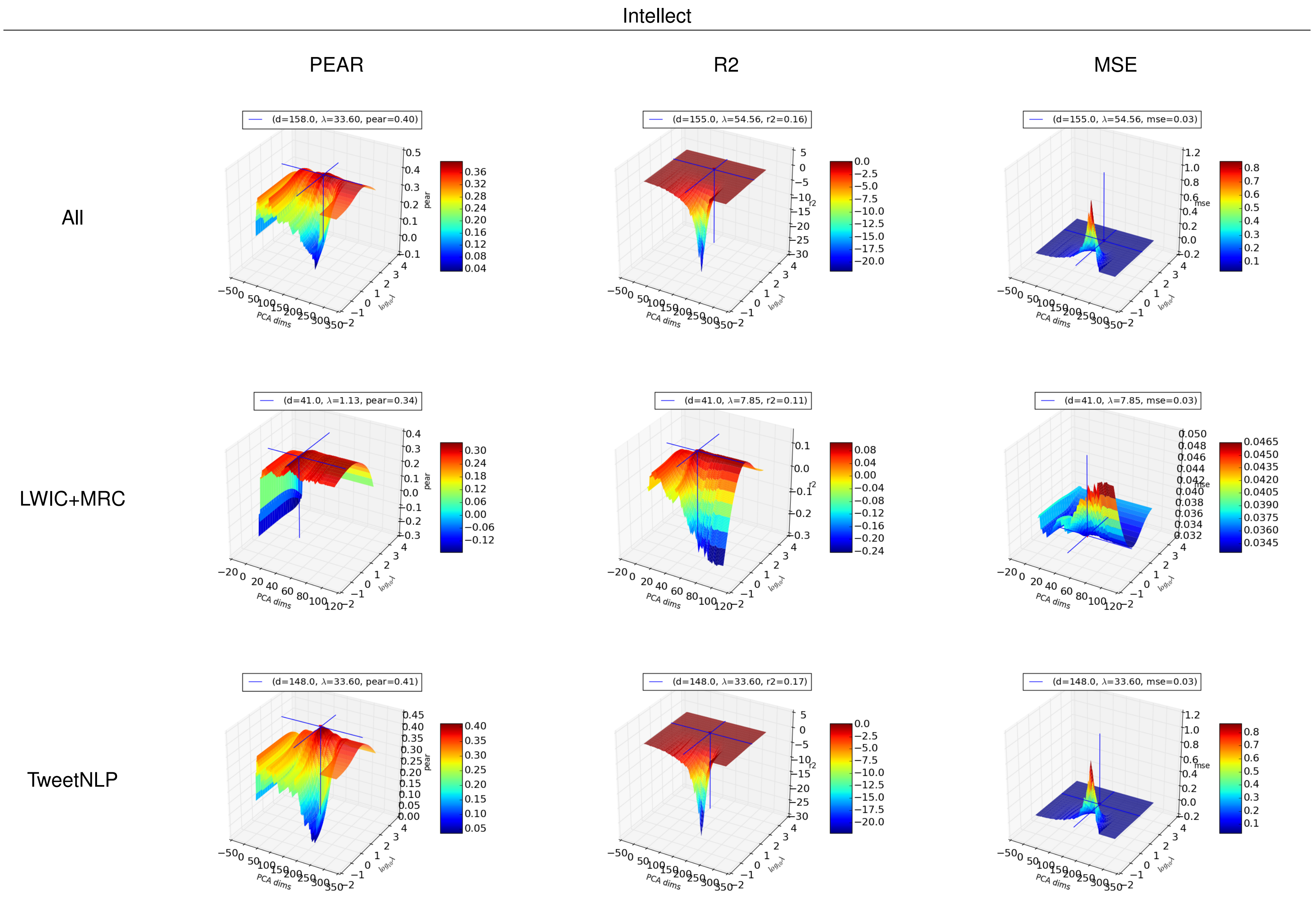



Figure 25: Plot of optimal point for intellect

## 7.5 Ridge regression with actual values

### 7.5.1 Table of results

MSE and MAE are calculated over labels normalized to the range [0, 1]. `tf`-normalized features used. Correlation metrics of naive predictors are undefined as variance of predictions is zero. See Section 7.2 for descriptions on how to read the tables.

| | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| NAIVE | N/A | N/A | N/A | 0.026 | 0.131 | N/A | N/A | N/A | 0.030 | 0.141 | N/A | N/A | N/A | 0.045 | 0.171 | N/A | N/A | N/A | 0.052 | 0.191 | N/A | N/A | N/A | 0.024 | 0.122 |

Figure 26: Baseline performance of naive mean-predictor.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.221 | 0.243 | 0.041 | 0.025 | 0.129 | 0.297 | 0.297 | 0.046 | 0.028 | 0.136 | 0.320 | 0.289 | 0.101 | 0.041 | 0.163 | 0.144 | 0.154 | -0.286 | 0.067 | 0.201 | 0.282 | 0.295 | 0.079 | 0.022 | 0.117 |
| SPEAR | 0.221 | 0.243 | 0.041 | 0.025 | 0.129 | 0.297 | 0.297 | 0.046 | 0.028 | 0.136 | 0.287 | 0.302 | 0.030 | 0.044 | 0.168 | 0.138 | 0.159 | -1.644 | 0.138 | 0.279 | 0.277 | 0.305 | 0.066 | 0.022 | 0.117 |
| R2 | 0.219 | 0.243 | 0.047 | 0.025 | 0.129 | 0.292 | 0.287 | 0.083 | 0.027 | 0.135 | 0.319 | 0.291 | 0.101 | 0.041 | 0.164 | 0.119 | 0.121 | 0.012 | 0.051 | 0.187 | 0.282 | 0.295 | 0.079 | 0.022 | 0.117 |
| MSE | 0.219 | 0.243 | 0.047 | 0.025 | 0.129 | 0.292 | 0.287 | 0.083 | 0.027 | 0.135 | 0.319 | 0.291 | 0.101 | 0.041 | 0.164 | 0.119 | 0.121 | 0.012 | 0.051 | 0.187 | 0.282 | 0.295 | 0.079 | 0.022 | 0.117 |
| MAE | 0.216 | 0.219 | 0.043 | 0.025 | 0.129 | 0.296 | 0.293 | 0.074 | 0.027 | 0.135 | 0.320 | 0.289 | 0.101 | 0.041 | 0.163 | 0.127 | 0.130 | 0.009 | 0.052 | 0.187 | 0.277 | 0.305 | 0.066 | 0.022 | 0.117 |

Figure 27: Regression using `tf`-normalized LWIC + MRC + TweetNLP + misc features.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.182 | 0.164 | 0.019 | 0.026 | 0.130 | 0.131 | 0.198 | -0.045 | 0.031 | 0.145 | 0.206 | 0.231 | 0.042 | 0.043 | 0.168 | 0.184 | 0.167 | 0.030 | 0.050 | 0.187 | 0.177 | 0.244 | -0.007 | 0.024 | 0.122 |
| SPEAR | 0.132 | 0.196 | -0.032 | 0.027 | 0.133 | 0.076 | 0.237 | -0.531 | 0.045 | 0.156 | 0.123 | 0.266 | -0.801 | 0.082 | 0.192 | 0.109 | 0.172 | -0.070 | 0.056 | 0.192 | 0.161 | 0.253 | -0.305 | 0.031 | 0.134 |
| R2 | 0.178 | 0.156 | 0.031 | 0.026 | 0.129 | 0.112 | 0.142 | 0.009 | 0.029 | 0.143 | 0.206 | 0.231 | 0.042 | 0.043 | 0.168 | 0.178 | 0.163 | 0.031 | 0.050 | 0.187 | 0.164 | 0.241 | 0.027 | 0.023 | 0.121 |
| MSE | 0.178 | 0.156 | 0.031 | 0.026 | 0.129 | 0.112 | 0.142 | 0.009 | 0.029 | 0.143 | 0.206 | 0.231 | 0.042 | 0.043 | 0.168 | 0.178 | 0.163 | 0.031 | 0.050 | 0.187 | 0.164 | 0.241 | 0.027 | 0.023 | 0.121 |
| MAE | 0.178 | 0.156 | 0.031 | 0.026 | 0.129 | 0.016 | -0.010 | -0.005 | 0.030 | 0.142 | 0.204 | 0.226 | 0.040 | 0.043 | 0.168 | 0.184 | 0.167 | 0.030 | 0.050 | 0.187 | 0.169 | 0.243 | 0.026 | 0.023 | 0.120 |

Figure 28: Regression using `tf`-normalized LWIC + MRC + misc features only.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.207 | 0.217 | 0.014 | 0.026 | 0.131 | 0.296 | 0.291 | 0.076 | 0.027 | 0.135 | 0.336 | 0.298 | 0.112 | 0.040 | 0.162 | 0.159 | 0.176 | -0.811 | 0.094 | 0.234 | 0.285 | 0.287 | 0.081 | 0.022 | 0.117 |
| SPEAR | 0.207 | 0.217 | 0.014 | 0.026 | 0.131 | 0.291 | 0.293 | 0.072 | 0.027 | 0.135 | 0.315 | 0.299 | 0.090 | 0.041 | 0.166 | 0.154 | 0.189 | -2.348 | 0.174 | 0.309 | 0.254 | 0.299 | -0.033 | 0.025 | 0.120 |
| R2 | 0.202 | 0.211 | 0.040 | 0.025 | 0.129 | 0.293 | 0.285 | 0.085 | 0.027 | 0.135 | 0.336 | 0.298 | 0.112 | 0.040 | 0.162 | 0.104 | 0.122 | 0.009 | 0.052 | 0.189 | 0.285 | 0.287 | 0.081 | 0.022 | 0.117 |
| MSE | 0.202 | 0.211 | 0.040 | 0.025 | 0.129 | 0.293 | 0.285 | 0.085 | 0.027 | 0.135 | 0.336 | 0.298 | 0.112 | 0.040 | 0.162 | 0.104 | 0.122 | 0.009 | 0.052 | 0.189 | 0.285 | 0.287 | 0.081 | 0.022 | 0.117 |
| MAE | 0.202 | 0.211 | 0.040 | 0.025 | 0.129 | 0.286 | 0.274 | 0.081 | 0.027 | 0.135 | 0.336 | 0.298 | 0.112 | 0.040 | 0.162 | 0.109 | 0.119 | -0.010 | 0.053 | 0.187 | 0.277 | 0.295 | 0.068 | 0.022 | 0.116 |

Figure 29: Regression using `tf`-normalized TweetNLP + misc features only.

### 7.5.2 Plots

These are plots of predicted (y-axis) against true personality score (x-axis), using the best training parameters, using training data and labels, but without filtering for minimum tweet count. The ranks have been normalized to have the same range as raw personality score, [0, 100], for comparison against Figure 30. Correlation-based metrics are in shades of red, whereas distance-based metrics are in shades of blue. See Figure 19 for an extended description of plots.

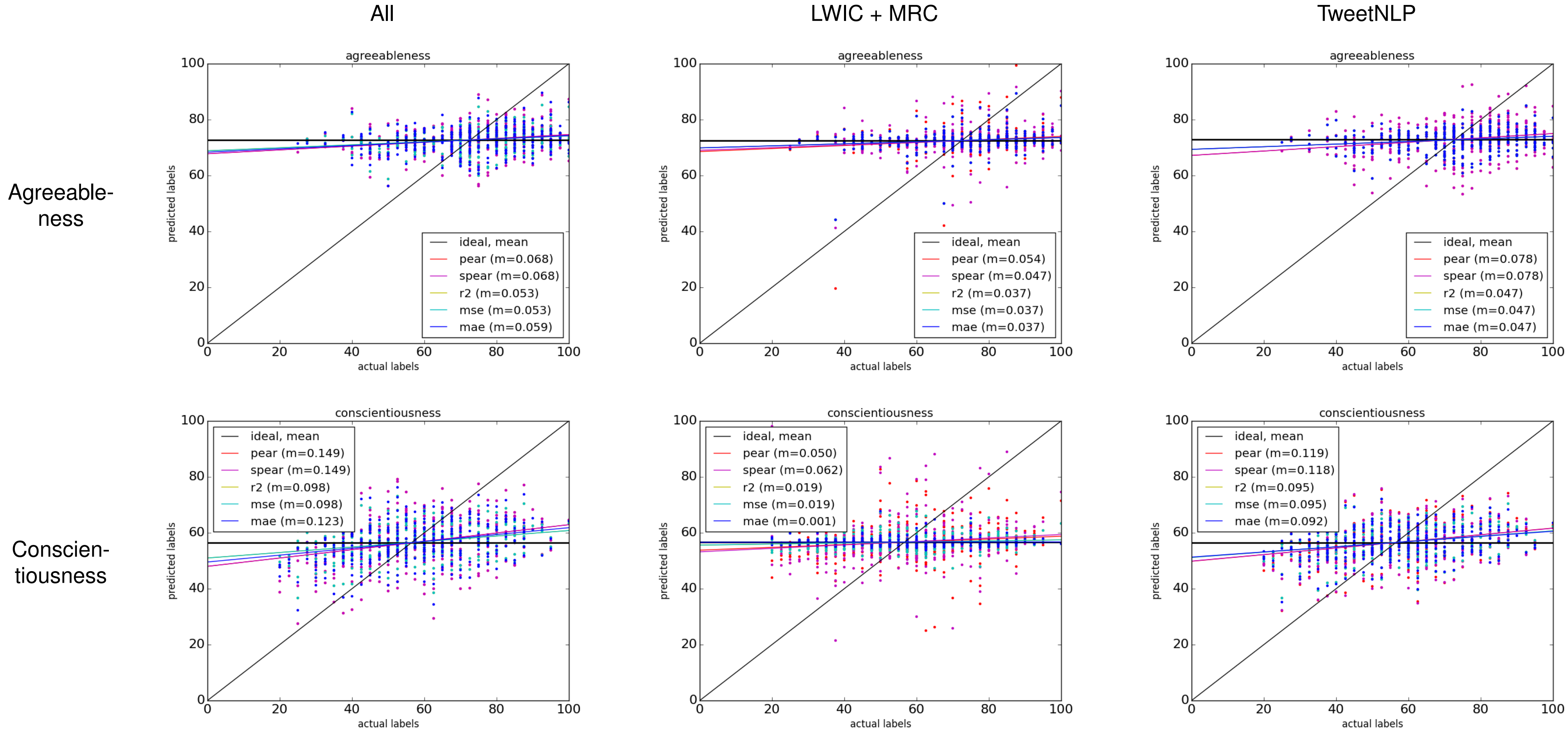



Figure 30: Plot of predicted (y-axis) against actual personality score (x-axis)

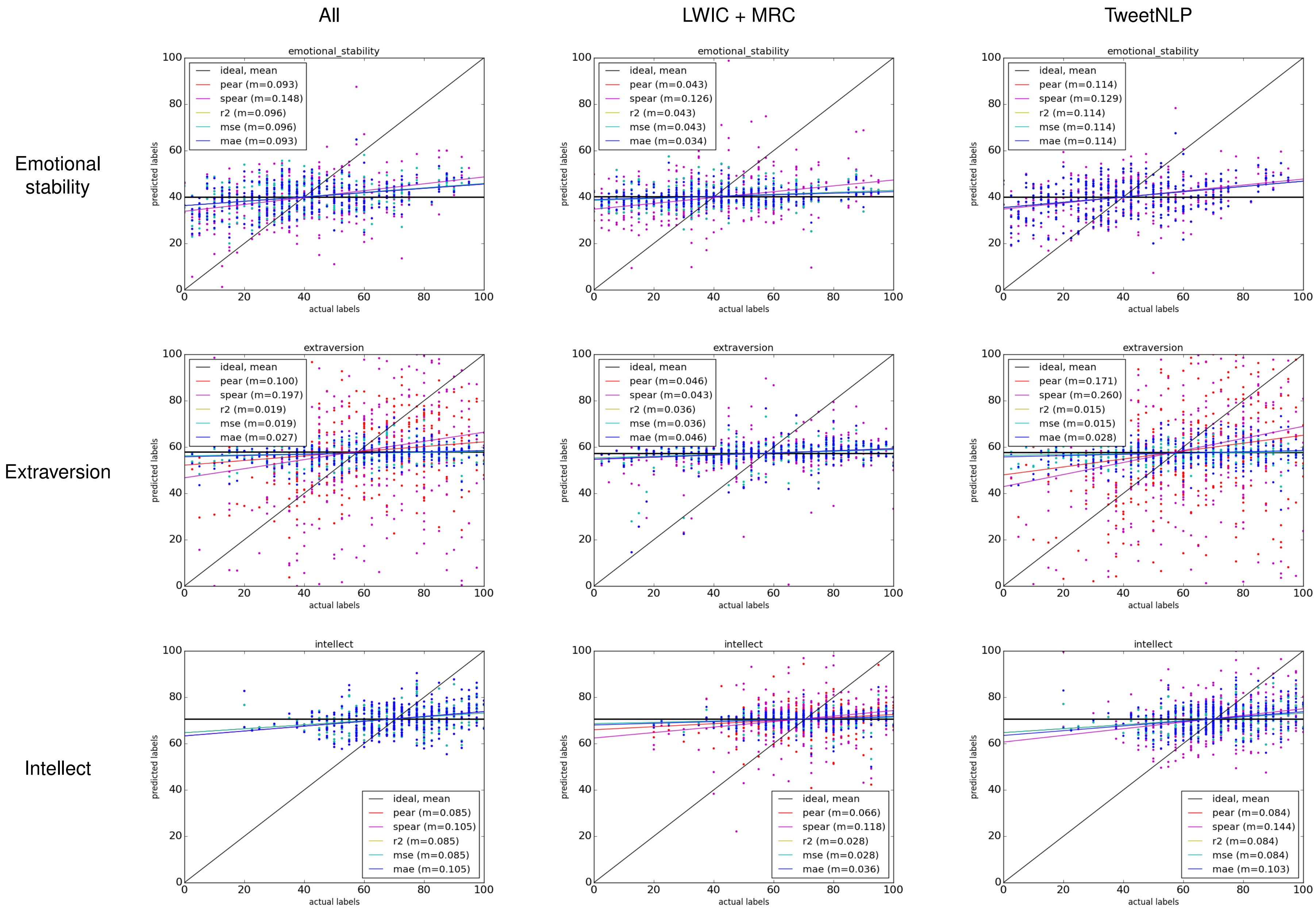



Figure 31: (continued from Figure 30) Plot of predicted (y-axis) against actual personality score (x-axis)

See Figure 21 for descriptions of these plots.

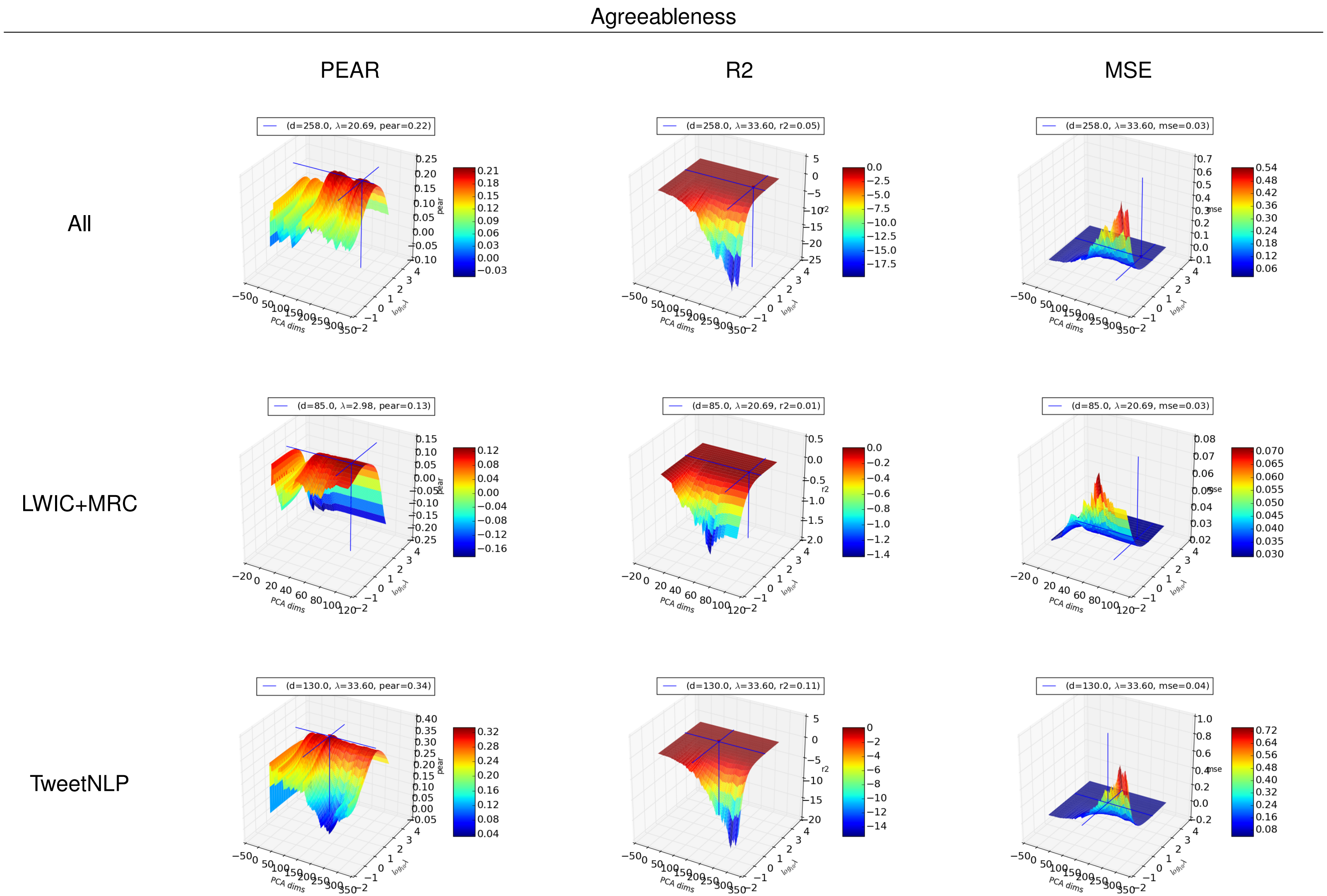



Figure 32: Plot of optimal point for agreeableness

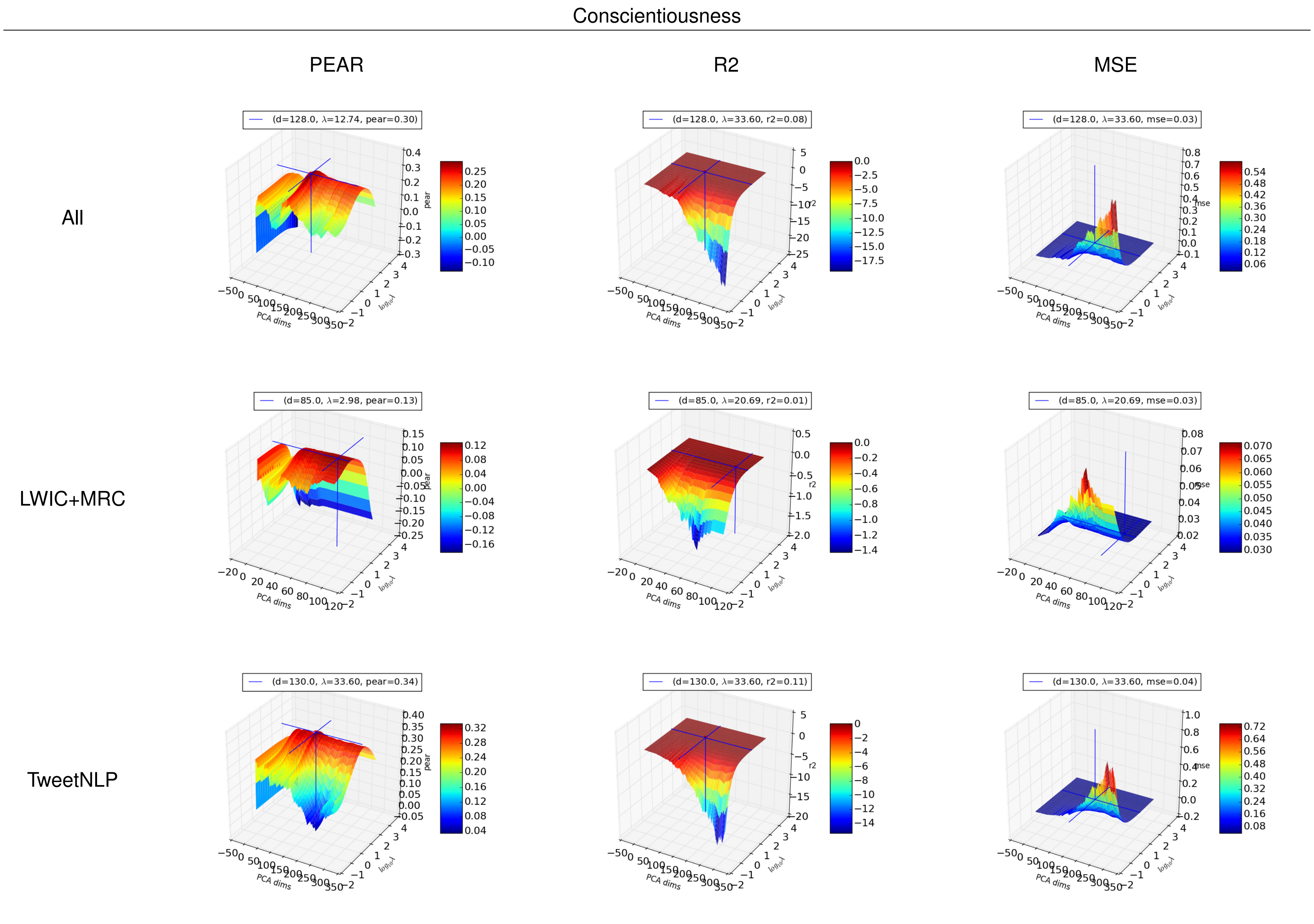



Figure 33: Plot of optimal point for conscientiousness

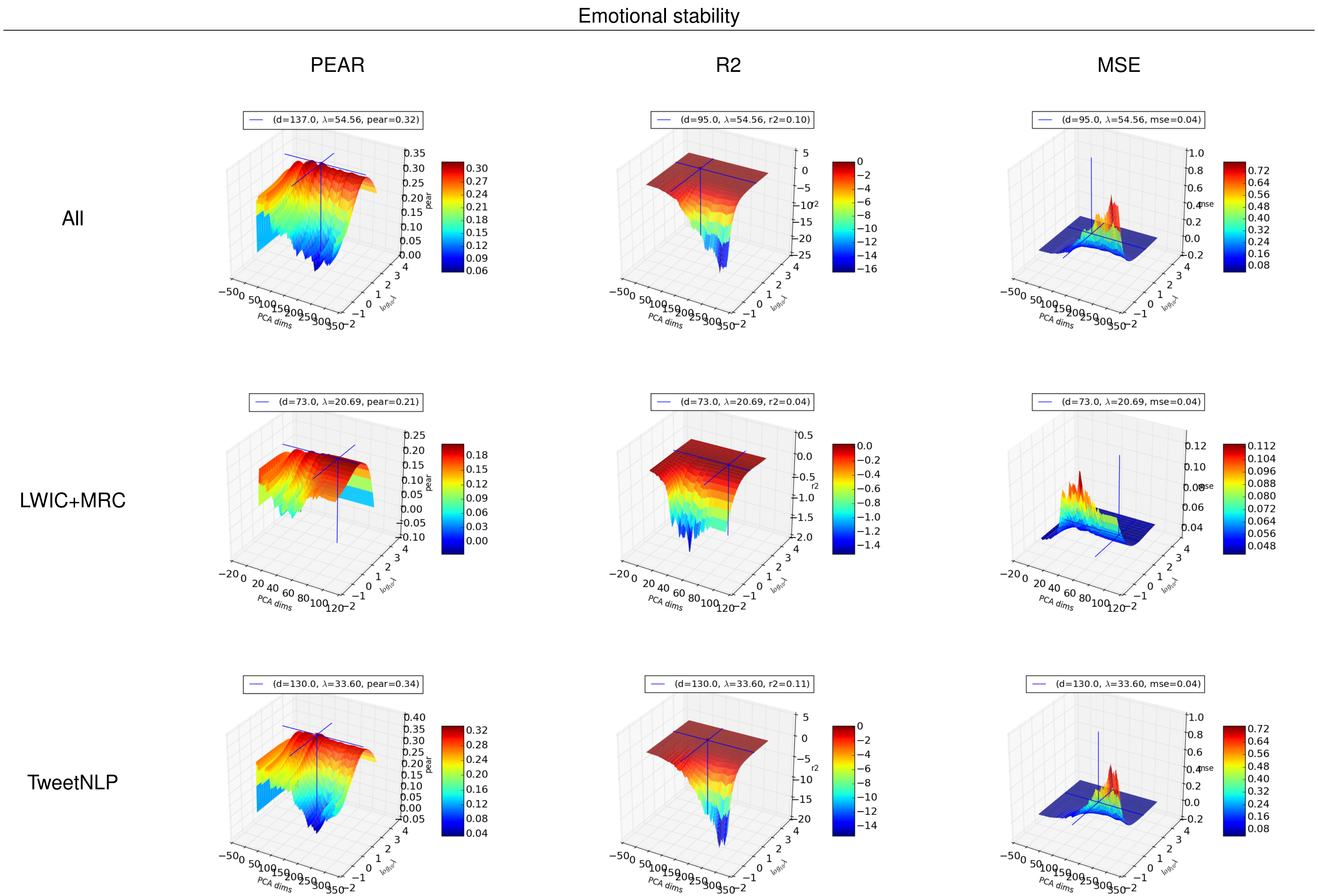



Figure 34: Plot of optimal point for emotional stability

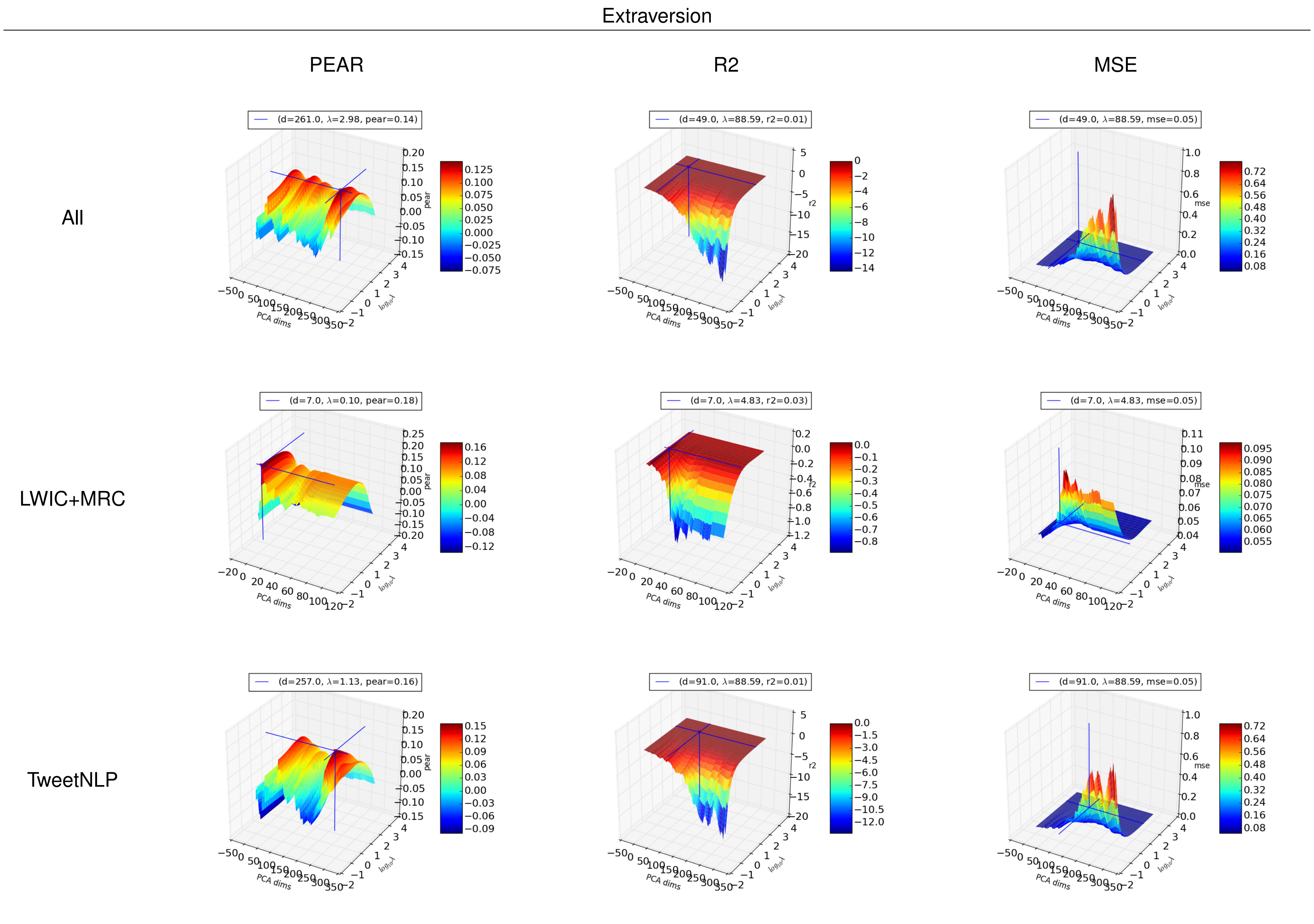



Figure 35: Plot of optimal point for extraversion

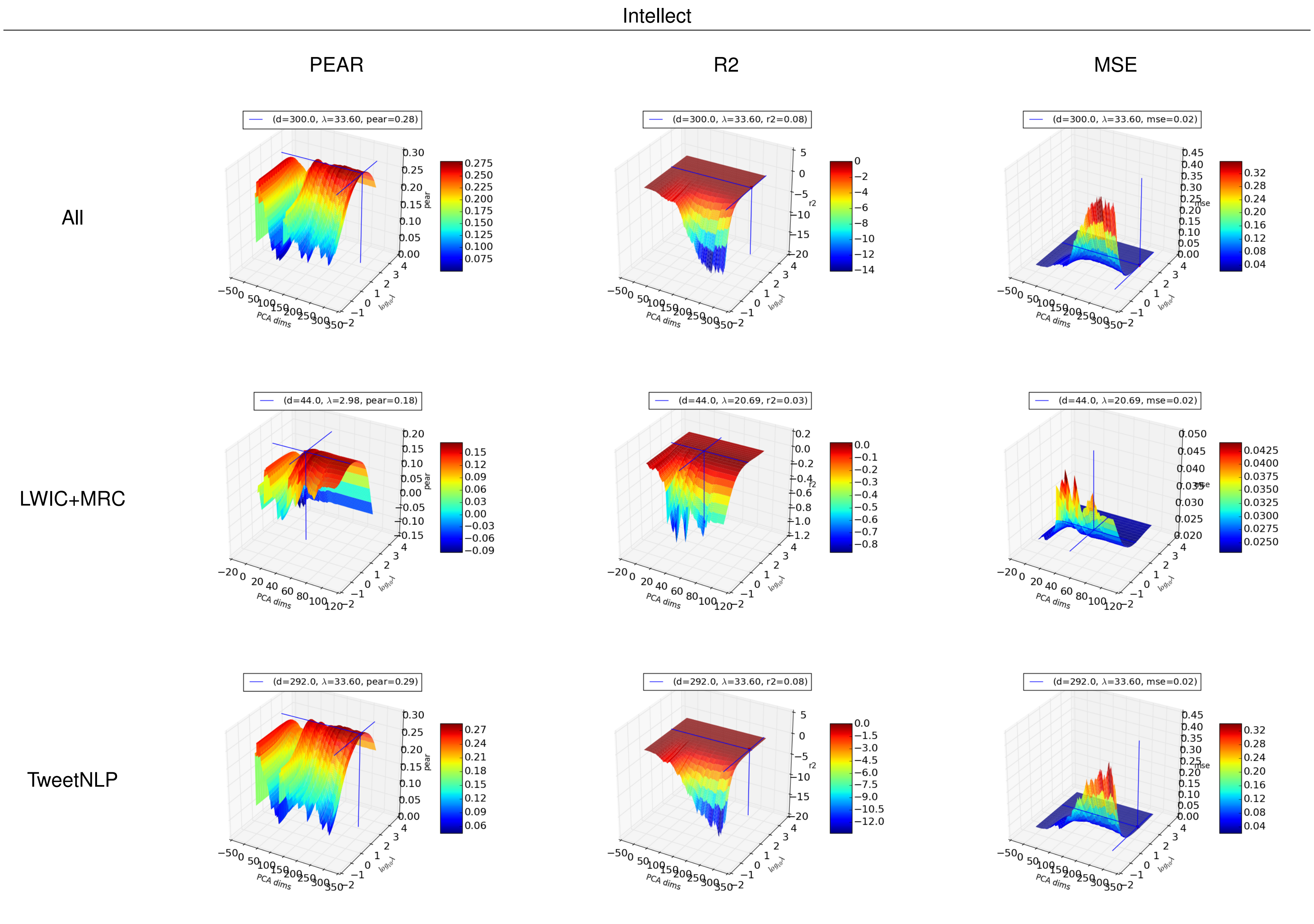



Figure 36: Plot of optimal point for intellect

## 7.6 Ridge regression with ranked values

### 7.6.1 Table of results

MSE and MAE are calculated over labels normalized to the range [0, 1]. tf-normalized features used. Correlation metrics of naive predictors are undefined as variance of predictions is zero. Since ranks are being used, all personality scores have the same mean. See Section 7.2 for descriptions on how to read the tables.

| | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| NAIVE | N/A | N/A | N/A | 0.084 | 0.251 | N/A | N/A | N/A | 0.084 | 0.251 | N/A | N/A | N/A | 0.084 | 0.251 | N/A | N/A | N/A | 0.084 | 0.251 | N/A | N/A | N/A | 0.084 | 0.251 |

Figure 37: Baseline performance of naive mean-predictor.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.264 | 0.257 | 0.042 | 0.080 | 0.241 | 0.297 | 0.291 | 0.087 | 0.077 | 0.239 | 0.330 | 0.315 | 0.109 | 0.075 | 0.233 | 0.101 | 0.093 | -0.232 | 0.103 | 0.262 | 0.283 | 0.296 | 0.080 | 0.077 | 0.236 |
| SPEAR | 0.257 | 0.261 | 0.049 | 0.080 | 0.240 | 0.259 | 0.305 | -0.057 | 0.089 | 0.245 | 0.330 | 0.315 | 0.109 | 0.075 | 0.233 | 0.076 | 0.117 | -0.358 | 0.114 | 0.270 | 0.278 | 0.298 | 0.043 | 0.080 | 0.235 |
| R2 | 0.256 | 0.255 | 0.066 | 0.078 | 0.241 | 0.297 | 0.291 | 0.087 | 0.077 | 0.239 | 0.330 | 0.315 | 0.109 | 0.075 | 0.233 | 0.067 | 0.067 | 0.002 | 0.084 | 0.248 | 0.283 | 0.296 | 0.080 | 0.077 | 0.236 |
| MSE | 0.256 | 0.255 | 0.066 | 0.078 | 0.241 | 0.297 | 0.291 | 0.087 | 0.077 | 0.239 | 0.330 | 0.315 | 0.109 | 0.075 | 0.233 | 0.067 | 0.067 | 0.002 | 0.084 | 0.248 | 0.283 | 0.296 | 0.080 | 0.077 | 0.236 |
| MAE | 0.263 | 0.259 | 0.064 | 0.078 | 0.240 | 0.284 | 0.302 | 0.062 | 0.079 | 0.236 | 0.329 | 0.311 | 0.103 | 0.075 | 0.231 | 0.088 | 0.097 | -0.007 | 0.084 | 0.247 | 0.276 | 0.279 | 0.062 | 0.079 | 0.234 |

Figure 38: Regression using tf-normalized LWIC + MRC + TweetNLP + misc features.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.168 | 0.143 | 0.003 | 0.084 | 0.248 | 0.132 | 0.226 | -0.033 | 0.087 | 0.251 | 0.209 | 0.267 | 0.023 | 0.082 | 0.243 | 0.141 | 0.142 | 0.004 | 0.083 | 0.248 | 0.162 | 0.193 | -0.055 | 0.088 | 0.250 |
| SPEAR | 0.121 | 0.193 | -0.372 | 0.115 | 0.267 | 0.082 | 0.248 | -0.374 | 0.115 | 0.266 | 0.170 | 0.298 | -0.933 | 0.162 | 0.287 | 0.141 | 0.142 | 0.004 | 0.083 | 0.248 | 0.132 | 0.230 | -0.277 | 0.107 | 0.257 |
| R2 | 0.158 | 0.133 | 0.023 | 0.082 | 0.246 | 0.103 | 0.154 | 0.008 | 0.083 | 0.251 | 0.203 | 0.234 | 0.041 | 0.080 | 0.245 | 0.123 | 0.119 | 0.014 | 0.083 | 0.248 | 0.137 | 0.203 | 0.018 | 0.082 | 0.247 |
| MSE | 0.158 | 0.133 | 0.023 | 0.082 | 0.246 | 0.103 | 0.154 | 0.008 | 0.083 | 0.251 | 0.203 | 0.234 | 0.041 | 0.080 | 0.245 | 0.123 | 0.119 | 0.014 | 0.083 | 0.248 | 0.137 | 0.203 | 0.018 | 0.082 | 0.247 |
| MAE | 0.158 | 0.133 | 0.023 | 0.082 | 0.246 | 0.126 | 0.190 | 0.004 | 0.084 | 0.251 | 0.209 | 0.267 | 0.023 | 0.082 | 0.243 | 0.130 | 0.128 | 0.014 | 0.083 | 0.248 | 0.156 | 0.216 | 0.007 | 0.083 | 0.245 |

Figure 39: Regression using tf-normalized LWIC + MRC + misc features only.

| optim- | agreeableness | | | | | conscientiousness | | | | | emotional stability | | | | | extraversion | | | | | intellect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -izing | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE | PEAR | SPEAR | R2 | MSE | MAE |
| PEAR | 0.248 | 0.238 | 0.044 | 0.080 | 0.241 | 0.305 | 0.307 | 0.090 | 0.076 | 0.239 | 0.344 | 0.322 | 0.117 | 0.074 | 0.231 | 0.127 | 0.140 | -0.493 | 0.125 | 0.282 | 0.285 | 0.289 | 0.081 | 0.077 | 0.236 |
| SPEAR | 0.243 | 0.242 | 0.003 | 0.084 | 0.244 | 0.303 | 0.313 | 0.091 | 0.076 | 0.237 | 0.343 | 0.325 | 0.116 | 0.074 | 0.231 | 0.115 | 0.145 | -1.553 | 0.214 | 0.353 | 0.280 | 0.294 | 0.048 | 0.080 | 0.234 |
| R2 | 0.242 | 0.235 | 0.055 | 0.079 | 0.242 | 0.303 | 0.304 | 0.091 | 0.076 | 0.237 | 0.344 | 0.322 | 0.117 | 0.074 | 0.231 | 0.005 | 0.006 | -0.004 | 0.084 | 0.250 | 0.285 | 0.289 | 0.081 | 0.077 | 0.236 |
| MSE | 0.242 | 0.235 | 0.055 | 0.079 | 0.242 | 0.303 | 0.304 | 0.091 | 0.076 | 0.237 | 0.344 | 0.322 | 0.117 | 0.074 | 0.231 | 0.005 | 0.006 | -0.004 | 0.084 | 0.250 | 0.285 | 0.289 | 0.081 | 0.077 | 0.236 |
| MAE | 0.244 | 0.237 | 0.044 | 0.080 | 0.241 | 0.295 | 0.309 | 0.076 | 0.077 | 0.236 | 0.344 | 0.322 | 0.117 | 0.074 | 0.231 | 0.080 | 0.082 | -0.031 | 0.086 | 0.249 | 0.279 | 0.283 | 0.069 | 0.078 | 0.234 |

Figure 40: Regression using tf-normalized TweetNLP + misc features only.

### 7.6.2 Plots

These are plots of predicted (y-axis) against true personality score (x-axis), using the best training parameters, using *ranked* training data and labels. The ranks have been normalized to [0, 1], but the graphs plot 100x the values for comparison against Figure 30. Correlation-based metrics are in shades of red, whereas distance-based metrics are in shades of blue. See Figure 19 for an extended description of plots.

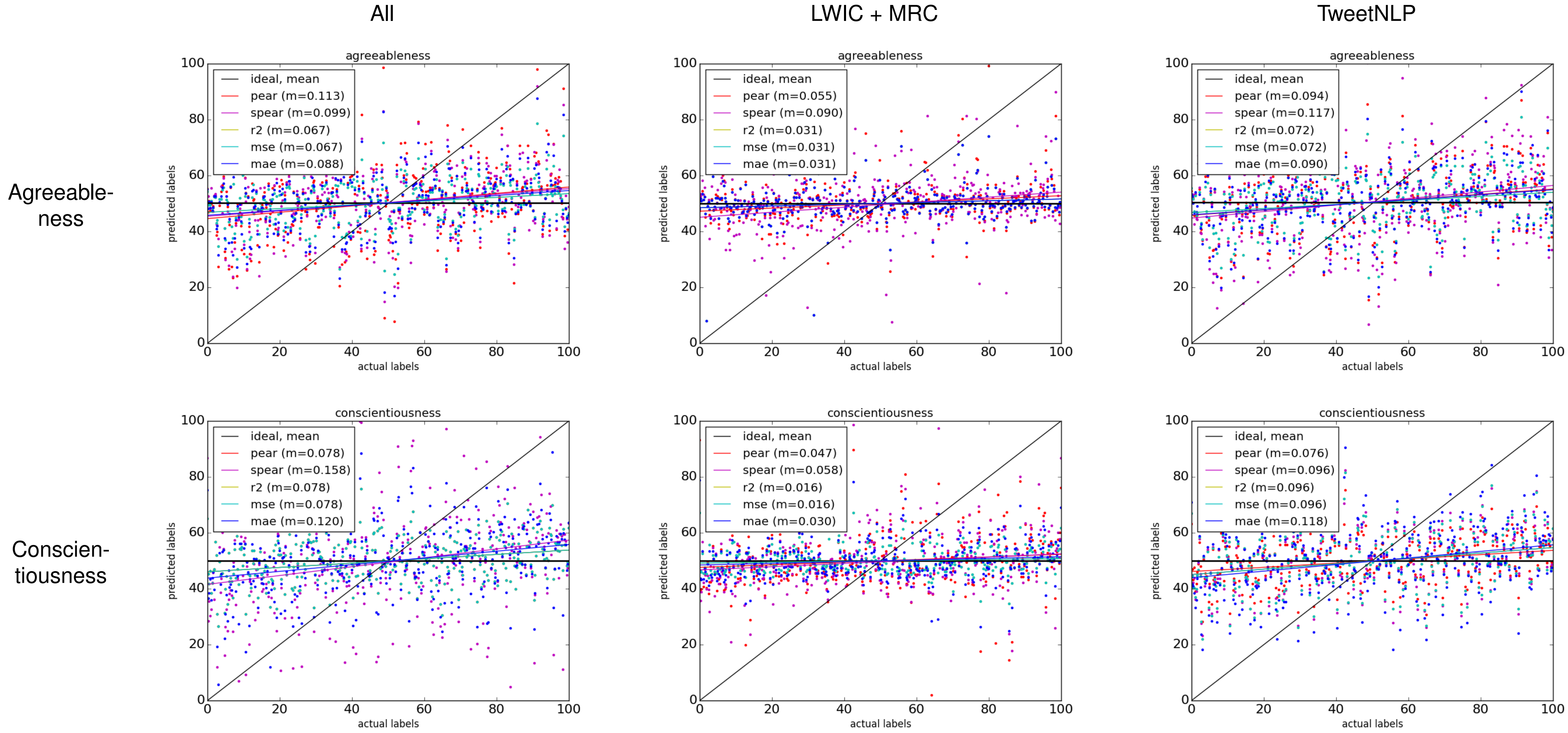



Figure 41: Plot of predicted (y-axis) against actual personality ranked (x-axis)

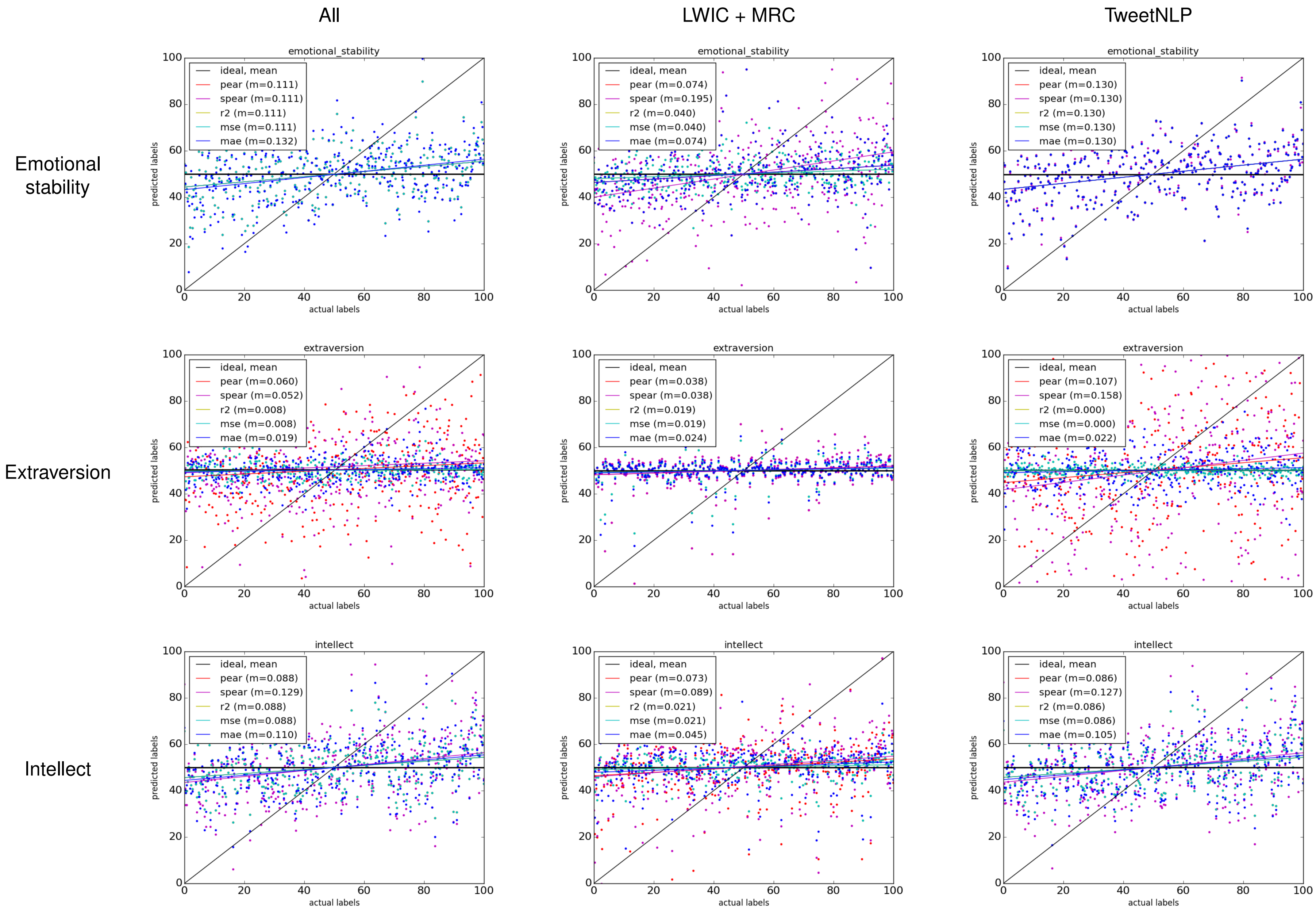



Figure 42: (continued from Figure 41) Plot of predicted (y-axis) against actual personality ranked (x-axis)

See Figure 21 for descriptions of these plots.

Agreeableness

| | PEAR | R2 | MSE |
|---|---|---|---|
| All | | | |
| LWIC+MRC | | | |
| TweetNLP | | | |

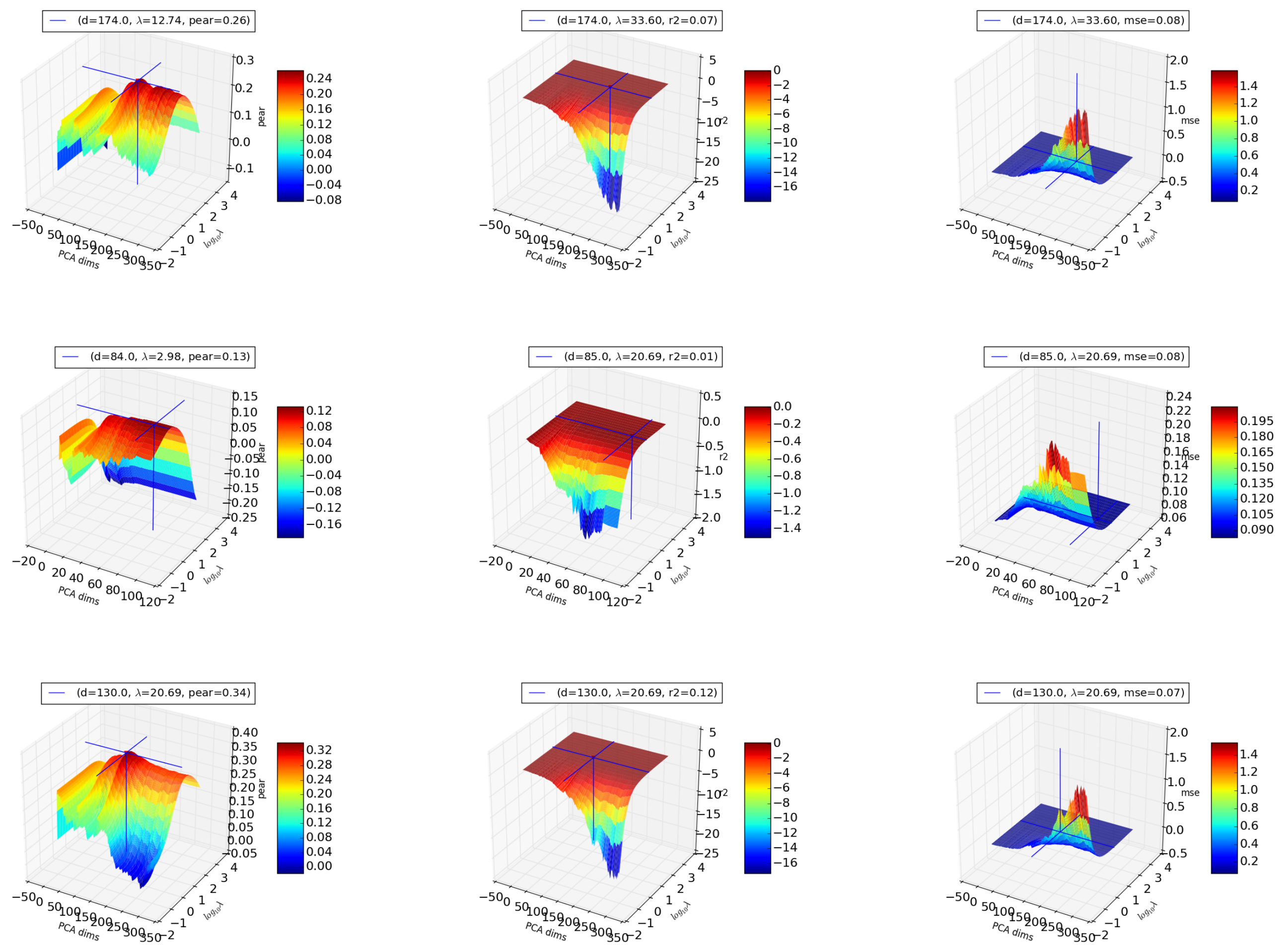



Figure 43: Plot of optimal point for agreeableness

Conscientiousness

| | PEAR | R2 | MSE |
|---|---|---|---|
| All | | | |
| LWIC+MRC | | | |
| TweetNLP | | | |

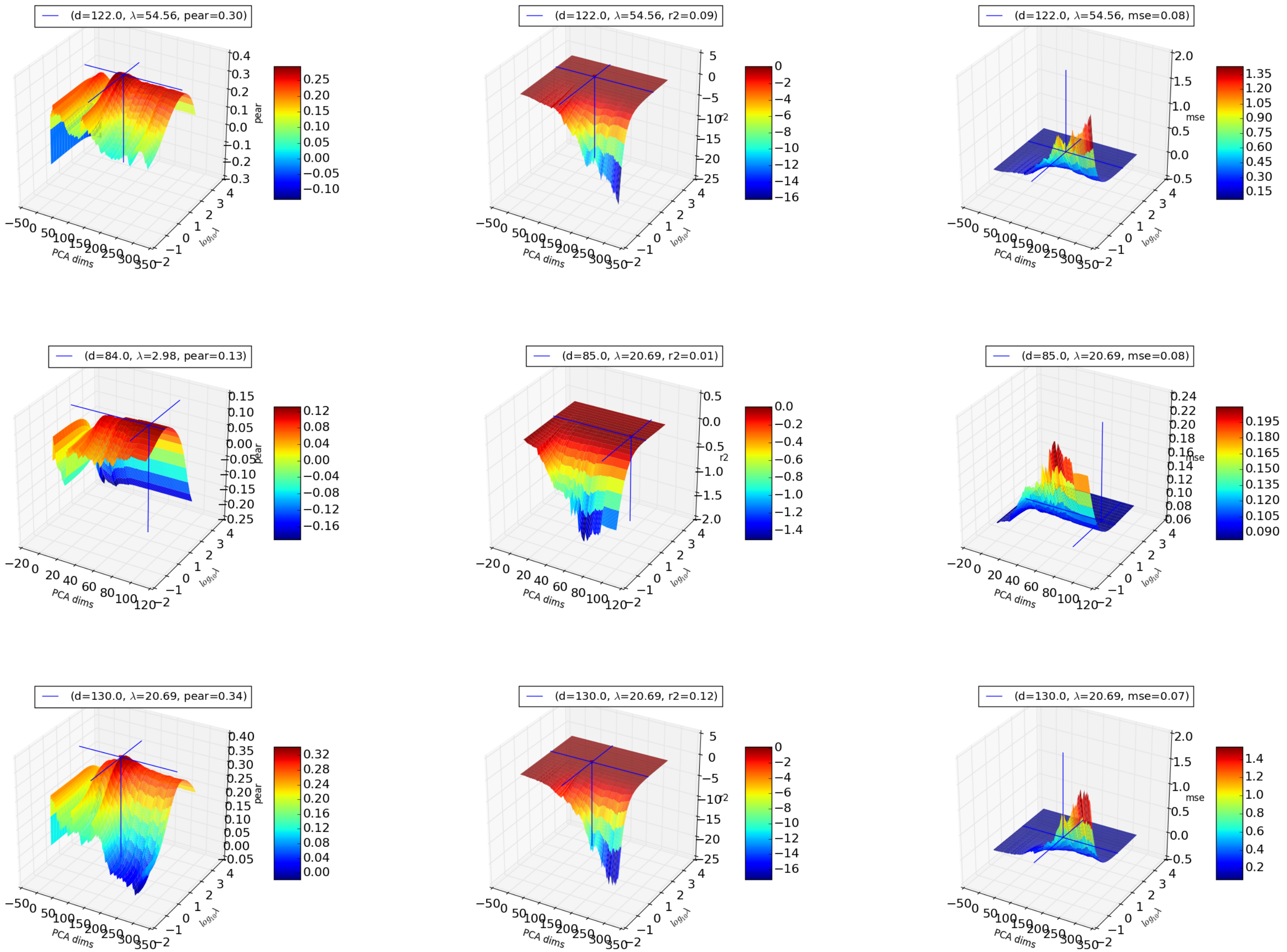



Figure 44: Plot of optimal point for conscientiousness

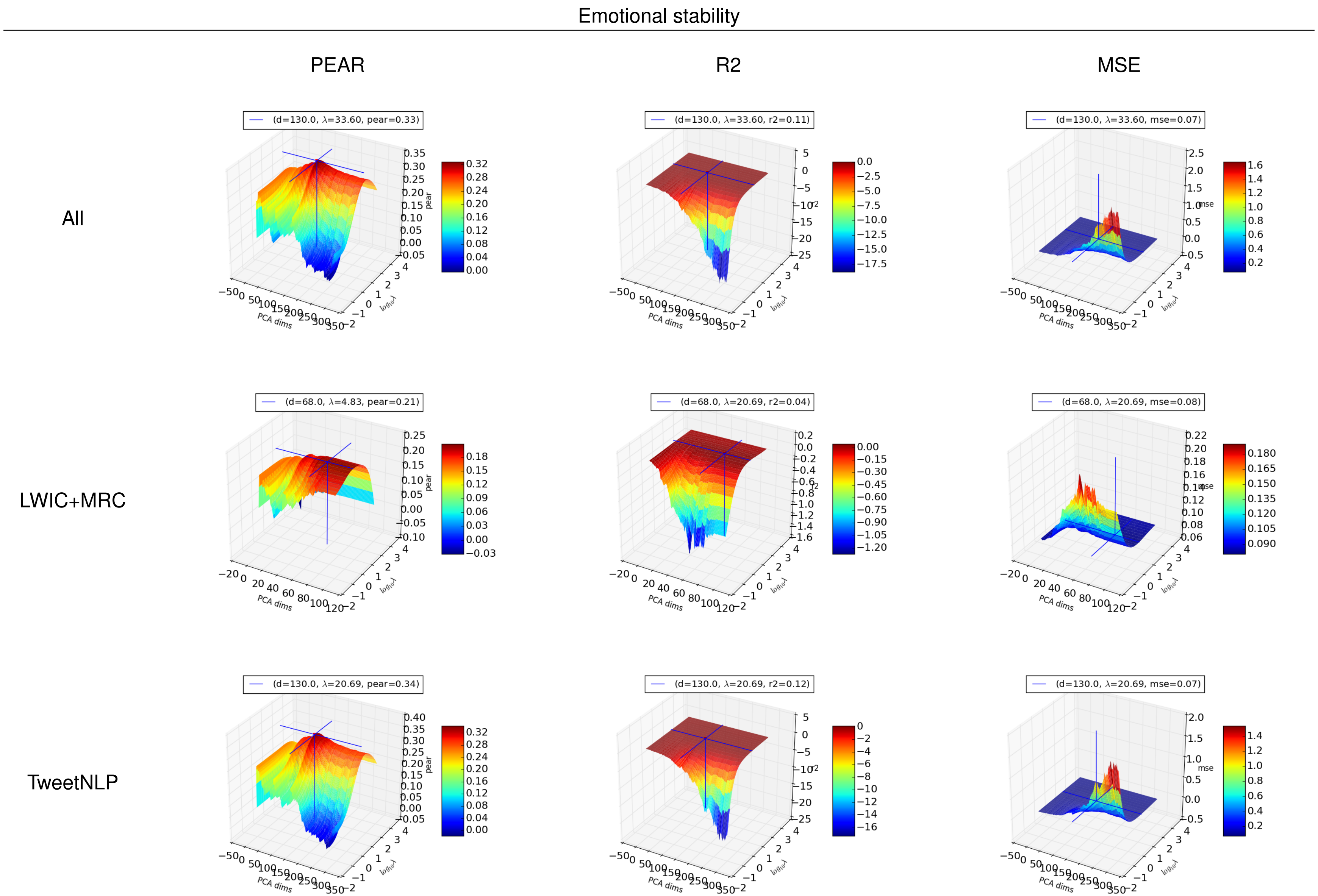



Figure 45: Plot of optimal point for emotional stability

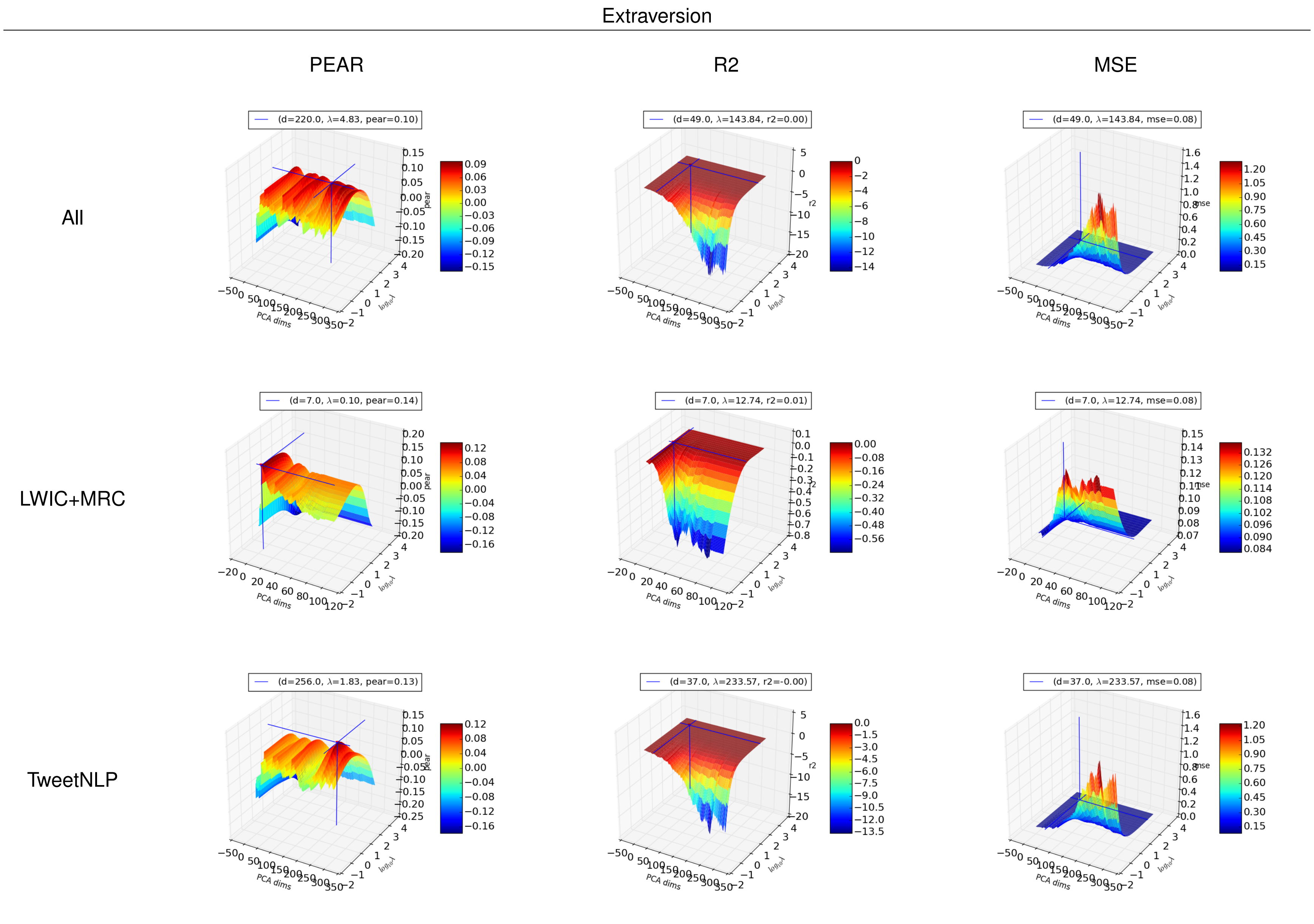



Figure 46: Plot of optimal point for extraversion

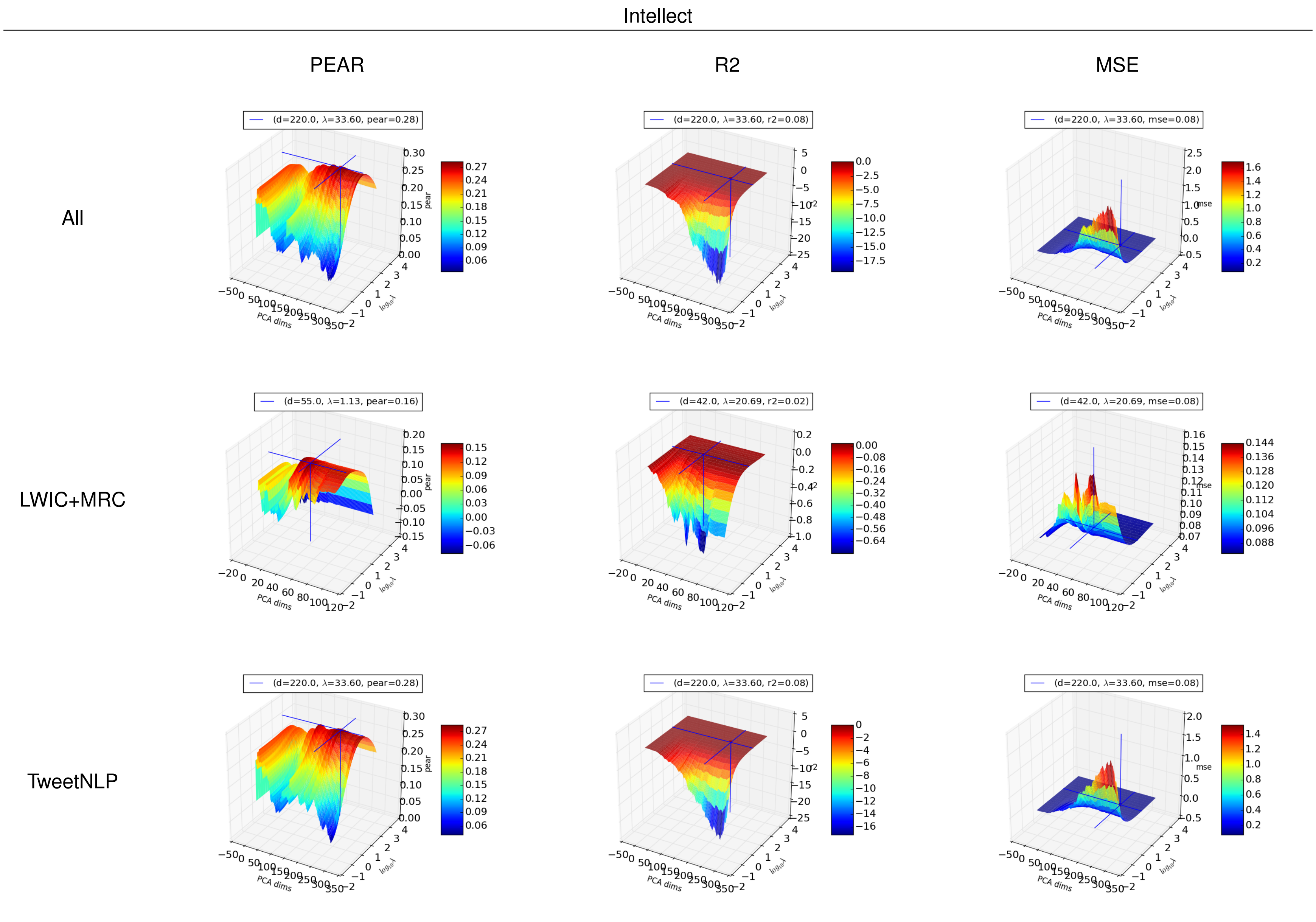



Figure 47: Plot of optimal point for intellect

## 7.7 Discussion (regression)

The main metric we are going to use to evaluate the performance of our models is R2, which measures the proportion of variance in the predicted values caused by variance in the original values. This is because when evaluated with the other metrics (PEAR, SPEAR, MSE, MAE), R2-optimized models almost always have comparable performance when evaluated with other metrics, in comparison to models optimized using those other metrics, whereas unfortunately the converse is not true..

### 7.7.1 Clean data

First, let us look at the results obtained from models trained with a relatively rich dataset, where we filtered out users with fewer than 200 tweets (n=243). The original sample population size is n=356.

If we observe Figures 16 and 27, we see that regression performance improved when we filtered users, despite removing just under a third of our available data. Models maximizing R2 score double in performance, except in extraversion, where the increase is 5x.

Looking at Figures 16, 17, and 18, when used on their own, TweetNLP features outperform LWIC + MRC features in all cases, from 44% better in agreeableness, to 131% in extraversion. Models trained using TweetNLP features have superior R2, PEAR, and SPEAR scores, and they are comparable using MAE and MSE. When used together, it improves the R2 score compared to just using LWIC + MRC features by between 10% (emotional stability) to 150% (extraversion). For intellect, performance increases by 50%, but this is a drop in performance compared to models trained just using TweetNLP features.

### 7.7.2 Noisy data

Now we will compare models' performance on noisy data, when we do not filter users. If we compare Figures 28, 29 and Figures 39, 40 we see that the models' performance are nearly identical whether we use actual scores or ranked scores, regardless of whichever subset of features we choose to train with.

Figures 27 and 29, show that models trained using all features have comparable performance to models trained only using LWIC + MRC features, in all areas except extraversion. Evaluating and optimizing using R2, models trained using all (noisy) data show improvements between a 52% (actual) to 100% (ranked) improvement for agreeableness, and between 8.2x (actual) to 10x (ranked) for conscientiousness. Predictive accuracy for extraversion in noisy data is very weak despite our best efforts. However, this result matches with Golbeck (2011)[7] and Liu et. al.'s findings (2015),[11] as it tends to perform worse than the other features, if not the worst. They only used LWIC + MRC features, but we've shown that adding TweetNLP features either hurts prediction for extraversion, or does not contribute any sort of improvement.

Perhaps extraversion is a higher level concept that requires a multi-faceted approach to capture (such as using total time spent on Twitter per how long the user has been registered etc., something we did not capture), rather than being something that can be predicted using text features. Or, perhaps the correlations are too weak to be learned from a few hundred users, especially when a significant number of them contribute to noise. TweetNLP word clusters are small and consist of many misspelled words, so when used in isolation, each cluster is a highly specific signal. Hence when used without general LWIC + MRC features, the classifier does not perform as well, especially if the user's corpus is small.

### 7.7.3 Overall discussion

These 3d plots show the 'shape' of the problem. Before we created these detailed plots, we sketched out the topography by roughly sampling a wide surface. We found that regardless of the number of PCA target dimensions we set (between 0 and n), the optimal $\lambda$ for every metric lies within a narrow band, for all 5 personality scores. this band is quite neatly constrained and does not shift as we change the target PCA dimensions.

These plots also serve to give insight into how we might be able to reconcile the optimal parameters found by maximizing or minimizing the various metrics we investigated. We will use the results for conscientiousness in Figure 39 as an example. We see that maximizing SPEAR gives us a model that evaluates unfavorably using the other metrics. Now, we omitted the SPEAR 3d plots as they are almost always the same shape as PEAR's, but we will include them both here to illustrate our point.

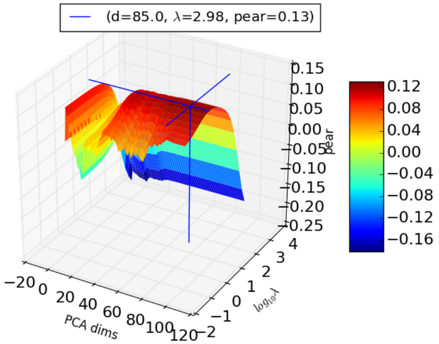



Figure 48: PEAR-maximizing conscientiousness model, trained on noisy data with actual values)

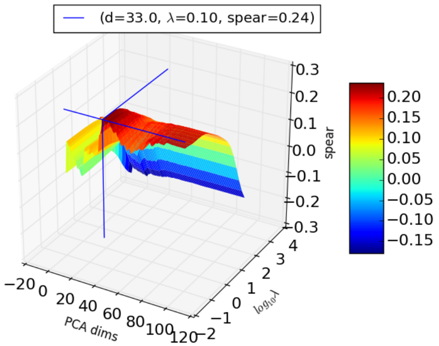



Figure 49: SPEAR-maximizing conscientiousness model, trained on noisy data with actual values)

Looking at Figures 48 and 49, we see that there are two or three ridges, and the maximum point happens to be at different locations. In this case, let's say we need to reconcile a single $\lambda$ and PCA dimension to train our predictor with. Taking the mean or the average would be the incorrect thing to do, because of the deep troughs between these suggested points of interest. This needs further work, but for now, an idea would be to have each model propose a 'vote' with their coordinates $\lambda, d$, we compute the mean, and take the point closest to the mean rather than the mean itself.

In other words, something like, where set $P$ is the set of all points proposed and point $p_i = (\lambda_i, d_i) \in P$,

$$\underset{(\lambda_i, d_i) \in P}{\arg\min} \sqrt{(\lambda_i - \bar{\lambda})^2 + (d_i, \bar{d})^2} \tag{17}$$

So far, we haven't mentioned the predicted versus actual scatterplots. We will use Figure 19 as an example. These plots are useful for showing if a model is having difficulty generalizing for a particular range of data, for example. All the summary statistics we've shown so far cannot capture this, unless we split the true $y$'s into distinct sets and calculate them individually. While we did not observe this phenomenon, the charts did pick up something else, which is that models trained by optimizing distance-based metrics tend to predict near the mean, as we can see the blue and green lines of best fit for predictions of MAE and MSE-trained models tend to lie flatter than their correlation counterparts, which tend to predict distances further than the mean, but giving a line of least square with a smaller cosine angle to that of the ideal model.

This spread is particularly evident for extraversion. If we observe the plots for extraversion in Figure 42, we see that using different subsets of features, the models fail to generalize in two ways: for TweetNLP features, correlation-metric-trained models are essentially trying to minimize the cosine angle of line of best fit, completely sacrificing proximity to the mean and ending up predicting essentially over a the entire range. For the LWIC + MRC features, the model simply learns to predict like a naive mean predictor, as all the least square lines are almost parallel with the mean line. Using this figure as an example, we see that by combining LWIC + MRC and TweetNLP features, these two plots average each other out, narrowing the prediction range of correlation-trained models, and encouraging the distance-based models to make predictions further away from the mean.

### 7.7.4 Comparison of results

Results comparison table

| Paper | n | agreeableness | | | conscientiousness | | | emotional stability | | | extraversion | | | intellect | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | PEAR | MSE | MAE | PEAR | MSE | MAE | PEAR | MSE | MAE | PEAR | MSE | MAE | PEAR |
| Us (2015) | 243 | 0.050 | 0.182 | 0.265 | 0.041 | 0.161 | 0.385 | 0.037 | 0.152 | 0.468 | 0.049 | 0.180 | 0.273 | 0.032 | 0.144 | 0.399 |
| Golbeck (2011)[7] | 279 | - | 0.130 | - | - | 0.146 | - | - | 0.160 | - | - | 0.182 | - | - | 0.119 | - |
| Quercia (2011)[10] | 335 | 0.049 | - | - | 0.0475 | - | - | 0.0531 | - | - | 0.055 | - | - | 0.043 | - | - |
| Liu (2015)[11] | 100 | * | * | 0.1516 | * | * | 0.2102 | * | * | 0.1779 | * | * | 0.108 | * | * | 0.1952 |

Figure 50: Comparison of results with prior work. MSE-scores are [0,1]-normalized. For our results, for each score, we use the model that optimizes that score, when trained on all (LWIC + MRC + TweetNLP + misc) `tf`-normalized features on the >200 tweets dataset. For other papers' results, we use the model that gives the best score for that metric if they have multiple models. We have explained the ambiguity of RMSE formulas in Section 6.2.1. Where normalization range is unclear, the cell is filled with *. Where data is unavailable, we denote it with -. Lower is better for MSE and MAE. Higher is better for PEAR.

To compare results with prior work, we first need to normalize their ranges to [0,1]. Where $Y = ky + c$,

$$RMSE(Y) = \frac{\sqrt{\sum k^2 (y-\hat{y})^2}}{n}$$

$$\frac{RMSE(Y)}{k} = RMSE(y) = \sqrt{MSE(y)} \tag{18}$$

$$MSE(Y) = \frac{\sum k^2 (y-\hat{y})^2}{n}$$

$$\frac{MSE(Y)}{k^2} = MSE(y) \tag{19}$$

$$MAE(Y) = \frac{\sum k|y-\hat{y}|}{n}$$

$$\frac{MAE(Y)}{k} = MAE(y) \tag{20}$$

From Figure 50, we see that our models perform better than Quercia (2011)[10] and Liu's (2015),[11] using the statistics they chose to evaluate their models with, except in agreeableness, where it underperforms by 2% compared to Quercia's model. Liu et. al.'s work is interesting. Assuming a range of [0,100] because they did not mention normalization, their worst score, emotional stability (MAE=0.582, RMSE=0.729), gives normalized MAE and RMSE scores an order of magnitude smaller than anything else we have seen so far. That would be a fantastic result, if not for the fact that using the Pearson correlation scores they reported, which, when compared to our model, ours perform between 75% (agreeableness) to 160% (emotional stability) better.

As for Golbeck et. al.'s result, their model performs significantly better in agreeableness, conscientiousness, and intellect when evaluated with MAE, whereas our model marginally beats them out in emotional stability and extraversion. We only have MAE to compare with, so this is the extent to which we can compare our results. They reported personality score averages on a normalized [0,1] scale, but did not mention the range when they reported the results. We calculated their MAE with the assumption that it is indeed [0,1]-normalized.

# 8 Conclusion

We have shown that our methods and features are at the very least competitive with the state of the art, using only simple ridge regression in conjunction with PCA. Considering that there are 15 comparative models (3 papers, and 5 personality scores), our models outperform 11 of them, 9 of them by a large margin.

Summarizing our results, in the context of our classification task (clean data, tritile classification), we determined that `tf` features performed the most consistently. TweetNLP features perform comparably to LWIC and MRC features when used alone, and when used together, it improves classification performance for intellect and emotional stability, compared to using LWIC + MRC features alone.

We have shown that the extraversion is among the most difficult Big-Five aspects to predict. This is completely consistent with the findings in other literature (Golbeck 2011, Quercia 2012, Liu 2015).[7,11,39] Models have markedly lower performance in both the classification and regression tasks.

In the context of our regression problem, when the dataset is clean and rich, models trained with TweetNLP outperform those trained using LWIC + MRC features by between 44% to 131%, when evaluated using the R2 statistic. When used together, performance is improved across the board, with extraversion getting a 150% boost, despite its predictive difficulty.

When the dataset is noisy, TweetNLP on its own only gives comparable performance to LWIC + MRC, and performs worse in predicting extraversion. However, when the noisy data is ranked, LWIC + MRC is very sensitive to that change and performance suffers greatly, whereas TweetNLP features actually become more predictive for agreeableness, conscientiousness, and emotional stability.

We have suggested that this is because TweetNLP depends on many clusters of words which give specific signals, which captures different aspects of personality than LWIC + MRC, but requires more words to do so. Its robustness to the change to ranked data suggests that the aspects it captures are more general than LWIC and MRC's definitions.

Comparing the results between clean and noisy data, we conclude that clean data gives better regression performance, even at the cost of a third of the sample population. We cleaned the data by removing users with fewer than 200 tweets from the set, based on Gou et. al.'s study (2013).[6]

As for the many evaluation metrics we brought up, we discussed the potential pitfalls when using them, and emphasized the fact that many are required to capture the essence of a model. We have shown that prediction-truth value scatterplots and the 3d plots show features that our selected statistics fail to capture. We then suggested a method for reconciling discrepancies between $\lambda$ and PCA dimensions when optimizing for multiple metrics, based on observations of these 3d plots.

Based on our results, we have shown that TweetNLP features are at least competitive with LWIC and MRC features in predicting personality from Twitter.

# 9 Further work

We have shown that TweetNLP clusters are useful in predicting personality from Twitter. They are constructed automatically via Brown clustering, without any need for personality labels, so an obvious way to proceed would be to try an construct (or derive from TweetNLP) an automatic system that gives more than then 1,000 clusters in the `50mpaths` dataset, and see how this changes predictive performance. In our tasks, PCA reduces the 1,000 dimensions to 300 or less, so this is not a numerically-efficient way to proceed. However, since this directly solves the problem of cost of labeling corpuses, it is a tantalizing way forward.

The alternative would be to proceed more traditionally, by constructing personality dictionaries from corpuses labeled with personality scores, which may not necessarily have to be tweets. This is in line with conventional sentiment analysis methods. Liu's (2015)[11] work is related to this line of thought, as they investigated active learning to try and minimize the labels required to increase the performance of their model while minimizing the number of labels needed to do so.

Another way forward would be to initiate a concerted effort such as the myPersonality Facebook application. Gou's (2013)[6] KnowMe experimental system is a smaller-scale deployment of a very similar idea, but they kept it internally within IBM. The alternative would be to pool collected labels and users, which would create a much larger dataset to perform research on. The good part about tweets is that they are simply text files; as long as they are anonymized and the labels are passed along, features could be generated independently. All of the features we generated are based on publicly-available dictionaries and tools, or calculated ourselves.

# 10 Bibliography

[1] Back, Stopfer, Vazire et. al. Facebook profiles reflect actual personality, nt self-idealization. *Psychological science* (2010).

[2] Bachrach, Kosinki, Graepel. Personality and Patterns of Facebook Usage. *ACM* (2011):24-32.

[3] Kosinki, Stillwell, Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 110.15 (2013): 5802-5805

[4] Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behavior, and Social Networking* 14.9 (2011): 483-488

[5] Bai, Hao, Li et. al. Predicting Big Five Personality Traits of Microblog Users *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on.* VOL. 1. IEEE, 2013

[6] Bou, Zhou, Yang et. al. KnowMe and ShareMe: Understanding Automatically Discovered Personality Traits from Social Media and User Sharing Preferences *IBM Research Report* (2013).

[7] Golbeck, Robles, Edmonson et. al. Predicting Personality from Twitter *IEEE Third international Conference on Social Computing* (2011).

[8] Oberlander, Nowson. Whose thumb is it anyway? Classifying author personality from weblog text. *Proceedings of the COLING/ACL on Main conference poster sessions. Association for Computational Linguistics* (2006).

[9] Qiu, Lin, Ramsay et. al. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*. 46.6 (2012):710-718.

[10] Quercia, Kosinki, Stillwell. Our Twitter profiles, our selves: Predicting personality with Twitter. *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.* IEEE, 2011.

[11] Liu, Nie, Bai et. al. Personality Prediction for Microblog Users with Active Learning Method. *Human Centered Computing.* Springer International Publishing, 2015. 41-54.

[12] Mairesse, Wallker, Mehl. et. al. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* (2007): 457-500.

[13] Brown, Desouza, Mercer et. al. Class-based n-gram models of natural language. *Computational linguistics*, 18.4 (1992):467-479.

[14] Gimpel, Schneider, O'Connor et. al. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers- Volume 2 (pp. 42-47). Association for Computational Linguistics* (2011).

[15] Owoputi, O'Connor, Dyer et. al. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. *Proceedings of NAACL* (2013).

[16] Brody, Samuel, Diakopoulos. Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. *Proceedings of the conference on empirical methods in natural language processing* (pp. 562-570). Association of Computational Linguistics (2011).

[17] Ng. L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning.* ACM, 2004.

[18] Vapnik. Estimation of dependences based on empirical data. Vol. 41. New York: Springer-Verlag, 1982.

[19] Lv, Zhai. Lower-bounding term frequency normalization. *Proceedings of the 20th ACM international conference on Information and knowledge management.* ACM, 2011.

[20] Davis, Goadrich. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd international conference on Machine learning.* ACM, 2006.

[21] Huang, Ling. Using AUC and accuracy in evaluating learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on,* 17.3 (2005): 299-310

[22] Logo, Valverde. AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography* 17.2 (2008): 145-151.

[23] Boyd, Costa, Davis et. al. Unachievable region in precision-recall space and its effect on empirical evaluation. *Machine learning: proceedings of the International Conference. International Conference on Machine Learning.* Vol. 2012. NIH Public Access, 2012.

[24] Forman, Scholz. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *ACM SIGKDD Explorations Newsletter* 12.1 (2010): 49-57.

[25] Corey, Dunlap, Burke. Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *The Journal of general psychology* 125.3 (1998): 245-261

[26] Choudhury, Munmun, Counts, et. al. Predicting Depression via Social media. *ICWSM*. 2013.

[27] Holmes. RMS Error. <http://statweb.stanford.edu/~susan/courses/s60/split/node60.html> Stanford University. (Accessed 2015).

[28] Collins. Natural Language Processing. <https://www.coursera.org/course/nlangp> Columbia University. (Accessed 2015).

[29] Manning, Raghavan. An Introduction to Information Retrieval. Cambridge University Press, 2009. p. 233.

[30] Hastie, Tibshirani, Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2010.

[31] Natasa Milic-Frayling. Real-time Ad Bidding and User Surveillance. Principal Researcher, Microsoft Research Cambridge. UCL web economics guest lecture (3rd March 2015).

[32] Pedregosa, Garoquaux, Gramfort et. al. Scikit-learn: Machine Learning in Python. *JMLR 12*, (2011):2825-2830.

[33] Pennebaker, Francis, Booth. Linguistic inquiry and word count: LIWC 2001 *Mahway: Lawrence Erlbaum Associates* 71 (2001): 2001.

[34] Coltheart. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* 33.4 (1981): 497-505

[35] Wilson. MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers* 20.1 (1988): 6-10

[36] Bolinger, Dwight. Forms of English: Accent, Morpheme, Order. Harvard University Press, Cambridge, Massachusetts, USA.

[37] McCrae, John. An introduction to the five-factor model and its applications. *Journal of personality*, 60.2 (1992):175-215.

[38] Digman. Personality structure: Emergence of the five-factor model. *Annual review of Psychology*, 41.1(1990):417-440.

[39] Quercia, Lambiotte, Stillwell et. al. The personality of popular Facebook users. *Proceedings of the ACM 2012 conference on computer supported cooperative work.* ACM, 2012.

[40] Naumann, Vazire, Rentfrow et. al. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin.* (2009)

[41] Kenny, Horner, Kashy et. al. Consensus at Zero Acquaintance: Replication, Behavioral Cues, and Stability. *Journal of Personality and Social Psychology.* 62.1 (1992): 88-97

[42] Rentfrow, Gosling. Message in a ballad: The Role of Music Preferences in Interpersonal Perception. *Psychological Science.* 17.3 (2006): 236-242

[43] Gillath, Bahns, Ge et. al. Shoes as a source of first impressions. *Journal of Research in Personality.* 46.4 (2012): 423-430

[44] Maltby, Macaskill, Day. Failure to forgive self and others: A replication and extension of the relationship between forgiveness, personality, social desirability and general health. *Personality and Individual Differences* 30.5 (2001): 881-885

[45] Kassarjian. Personality and consumer behaviour: A review. *Journal of Marketing Research* (1971): 409-418.

[46] Rook, Hoch. Consuming impulses. *Advances in consumer research.* 12.1 (1985): 23-27

# 11 Appendix

(*This page intentionally left blank*)

Table of correlations. Value on the left is the Pearson correlation coefficient, value on the right is the p-value of that correlation. These are correlations between original personality scores (not ranks) and tf features. tf features are raw feature scores (counts) divided by the total word count the user has. Generated from a set of 243 users, filtered down from 357 by selecting for users who have more than 200 tweets. Features with correlations of p-values >0.05 are omitted.

Significance or signal level matrix summarize different ranges of p-values:
1 : (0.01, 0.05]
2 : (0.005, 0.01]
3 : (0.0001, 0.005]
4 : [0.0001, 0]

| feature (tf) | sig | | | | | agreeableness | | conscientiousness | | emotional stability | | extraversion | | intellect | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Misc features | | | | | | | | | | | | | | | |
| capitalWords | | 3 | | | | -0.094819 | 0.140538 | 0.187370 | 0.003371 | 0.062970 | 0.328317 | 0.048696 | 0.449874 | -0.101935 | 0.112977 |
| friends | | 1 | 1 | | | 0.013499 | 0.834177 | 0.150598 | 0.018829 | 0.153089 | 0.016930 | -0.026118 | 0.685397 | 0.037028 | 0.565681 |
| letter_runs_numOfRuns | | | 3 | 1 | | 0.054442 | 0.398154 | -0.065743 | 0.307422 | -0.204750 | 0.001330 | -0.135678 | 0.034524 | 0.018584 | 0.773178 |
| tweets | | | 4 | | 1 | 0.010465 | 0.871069 | -0.083310 | 0.195589 | -0.249776 | 0.000083 | -0.031598 | 0.624029 | -0.150689 | 0.018757 |
| Lwic features | | | | | | | | | | | | | | | |
| lwic_Achiev | | 3 | 1 | | | 0.018872 | 0.769755 | 0.183253 | 0.004154 | 0.145459 | 0.023337 | -0.039752 | 0.537423 | 0.051127 | 0.427542 |
| lwic_Adverbs | | 3 | 1 | | | 0.000768 | 0.990496 | -0.217200 | 0.000652 | -0.158646 | 0.013286 | 0.079145 | 0.218952 | -0.108279 | 0.092150 |
| lwic_Affect | | 1 | | | 2 | 0.094894 | 0.140226 | -0.140151 | 0.028942 | -0.001148 | 0.985795 | 0.028360 | 0.660009 | 0.175644 | 0.006046 |
| lwic_Anger | | 1 | | | 1 | -0.023382 | 0.716864 | -0.127790 | 0.046598 | 0.057397 | 0.373009 | -0.051415 | 0.424946 | -0.141760 | 0.027133 |
| lwic_Article | | 3 | | 1 | | 0.003664 | 0.954688 | 0.217453 | 0.000642 | 0.061214 | 0.342011 | 0.135827 | 0.034325 | -0.030122 | 0.640327 |
| lwic_Assent | | 3 | 3 | | 1 | 0.010579 | 0.869684 | -0.198551 | 0.001870 | -0.237755 | 0.000183 | -0.032820 | 0.610677 | -0.152976 | 0.017013 |
| lwic_AuxVb | | 3 | 1 | | | -0.017235 | 0.789235 | -0.190042 | 0.002936 | -0.144252 | 0.024522 | -0.014640 | 0.820382 | -0.022128 | 0.731441 |
| lwic_Bio | | | 3 | | | 0.119885 | 0.062053 | -0.081143 | 0.207506 | -0.203736 | 0.001407 | 0.073158 | 0.255933 | -0.016172 | 0.801953 |
| lwic_Body | | | 4 | | 1 | 0.094944 | 0.140013 | -0.074183 | 0.249313 | -0.250475 | 0.000079 | 0.026759 | 0.678099 | -0.134839 | 0.035667 |
| lwic_Cause | | | | 3 | | -0.020343 | 0.752377 | -0.039483 | 0.540175 | 0.052825 | 0.412336 | -0.180891 | 0.004674 | 0.067937 | 0.291517 |
| lwic_Certain | 3 | | 4 | | | -0.210185 | 0.000979 | 0.074541 | 0.247030 | -0.252618 | 0.000068 | -0.022027 | 0.732621 | -0.031894 | 0.620779 |
| lwic_CogMech | 1 | | | | | 0.128803 | 0.044873 | -0.086823 | 0.177335 | -0.063401 | 0.325010 | 0.070485 | 0.273761 | -0.056077 | 0.384122 |
| lwic_Conj | | 3 | 3 | | 4 | 0.005612 | 0.930650 | -0.241905 | 0.000140 | 0.238159 | 0.000179 | -0.083100 | 0.196720 | 0.250767 | 0.000077 |
| lwic_Death | | | | | 1 | 0.062012 | 0.335738 | -0.102012 | 0.112705 | 0.006565 | 0.918904 | 0.100237 | 0.119137 | 0.139878 | 0.029260 |
| lwic_Excl | | 3 | 3 | | | 0.033020 | 0.608495 | -0.193931 | 0.002394 | -0.217472 | 0.000641 | 0.057576 | 0.371520 | 0.011124 | 0.863031 |
| lwic_Feel | | | 1 | | | -0.040346 | 0.531356 | 0.030674 | 0.634216 | -0.160651 | 0.012152 | -0.047799 | 0.458271 | -0.013373 | 0.835694 |
| lwic_Filler | | 3 | | | 2 | 0.062054 | 0.335415 | -0.207148 | 0.001163 | -0.113983 | 0.076155 | -0.065099 | 0.312195 | -0.165756 | 0.009640 |
| lwic_Friends | | | 4 | | | 0.033554 | 0.602717 | -0.105518 | 0.100806 | -0.259373 | 0.000043 | -0.035570 | 0.581086 | 0.021269 | 0.741497 |
| lwic_Funct | | 1 | | 3 | 3 | 0.015016 | 0.815856 | -0.127414 | 0.047251 | 0.085996 | 0.181512 | -0.195086 | 0.002252 | 0.191826 | 0.002675 |
| lwic_Future | | | 2 | | | -0.101651 | 0.113990 | 0.019269 | 0.765054 | -0.177964 | 0.005400 | -0.048145 | 0.455020 | 0.038029 | 0.555207 |
| lwic_Health | 3 | | | | | 0.182644 | 0.004283 | -0.106720 | 0.096963 | -0.120740 | 0.060199 | 0.048049 | 0.455925 | -0.073406 | 0.254315 |
| lwic_Hear | 3 | | | | | -0.180432 | 0.004782 | -0.011181 | 0.862339 | -0.063832 | 0.321721 | 0.002735 | 0.966164 | -0.043053 | 0.504148 |
| lwic_I | | 3 | | 1 | 1 | -0.060664 | 0.346375 | -0.241499 | 0.000144 | 0.094105 | 0.143569 | -0.144833 | 0.023945 | 0.133491 | 0.037572 |
| lwic_Inhib | | 4 | 3 | | | -0.058925 | 0.360400 | 0.258774 | 0.000044 | 0.181203 | 0.004603 | -0.068123 | 0.290200 | 0.016948 | 0.792660 |
| lwic_Insight | | | 1 | | | -0.002310 | 0.971427 | -0.068160 | 0.289938 | 0.137079 | 0.032686 | 0.047693 | 0.459268 | -0.065050 | 0.312561 |
| lwic_Ipron | | | | | 1 | 0.007471 | 0.907759 | -0.056575 | 0.379906 | 0.022185 | 0.730774 | 0.059549 | 0.355324 | 0.145101 | 0.023683 |
| lwic_Leisure | | | | | 1 | 0.070260 | 0.275294 | 0.100429 | 0.118425 | -0.002751 | 0.965975 | -0.070245 | 0.275400 | 0.140708 | 0.028304 |
| lwic_Negate | | | 3 | | | -0.008585 | 0.894086 | -0.077175 | 0.230675 | -0.179926 | 0.004904 | -0.040462 | 0.530166 | 0.032635 | 0.612686 |
| lwic_Negemo | | | 3 | | 3 | -0.020232 | 0.753684 | -0.125664 | 0.050397 | -0.207873 | 0.001117 | 0.021819 | 0.735051 | -0.191505 | 0.002721 |
| lwic_Percept | 1 | | | | | 0.127431 | 0.047221 | 0.095068 | 0.139495 | -0.017809 | 0.782393 | -0.066936 | 0.298705 | -0.114617 | 0.074526 |
| lwic_Posemo | 2 | | | | | 0.176482 | 0.005805 | -0.067700 | 0.293208 | 0.078520 | 0.222627 | 0.020482 | 0.750738 | 0.072651 | 0.259251 |
| lwic_Ppron | | 3 | 3 | | | 0.003736 | 0.953801 | -0.220708 | 0.000529 | -0.181954 | 0.004433 | -0.084274 | 0.190445 | 0.077526 | 0.228552 |
| lwic_Prep | | 3 | | | 1 | 0.026767 | 0.678006 | 0.199412 | 0.001785 | -0.059179 | 0.358330 | -0.084994 | 0.186671 | -0.139729 | 0.029434 |
| lwic_Present | | 3 | | 1 | | 0.002894 | 0.964207 | -0.185925 | 0.003629 | 0.017177 | 0.789924 | 0.140082 | 0.029023 | -0.026927 | 0.676196 |
| lwic_Pronoun | | 3 | 4 | | | 0.030529 | 0.635816 | -0.236860 | 0.000194 | -0.353099 | 0.000000 | 0.068279 | 0.289089 | -0.079717 | 0.215632 |
| lwic_Quant | | 1 | 1 | | | -0.060029 | 0.351450 | 0.151764 | 0.017918 | -0.144889 | 0.023890 | 0.064129 | 0.319469 | -0.099917 | 0.120324 |
| lwic_Relativ | | 3 | 3 | | | 0.100008 | 0.119986 | 0.184074 | 0.003986 | -0.184663 | 0.003869 | 0.068091 | 0.290424 | -0.109778 | 0.087707 |
| lwic_Sad | | | | 3 | | -0.070779 | 0.271758 | -0.045635 | 0.478900 | -0.084105 | 0.191340 | -0.196489 | 0.002089 | 0.033322 | 0.605225 |
| lwic_See | 1 | 1 | | 1 | | 0.136916 | 0.032896 | 0.137275 | 0.032435 | -0.037294 | 0.562890 | 0.132381 | 0.039202 | -0.000328 | 0.995937 |
| lwic_Sexual | | 1 | 1 | | | 0.094245 | 0.142971 | -0.134891 | 0.035596 | -0.159909 | 0.012561 | 0.053008 | 0.410715 | -0.011729 | 0.855665 |
| lwic_SheHe | 1 | | | 1 | 1 | -0.136043 | 0.034037 | -0.017130 | 0.790497 | 0.034342 | 0.594219 | 0.127764 | 0.046643 | 0.144377 | 0.024398 |
| lwic_Space | | 2 | | | 1 | 0.018038 | 0.779667 | 0.171860 | 0.007248 | 0.035733 | 0.579350 | 0.090822 | 0.158127 | 0.129831 | 0.043176 |
| lwic_Swear | | | | 3 | | -0.043498 | 0.499751 | -0.114403 | 0.075074 | -0.001674 | 0.979284 | 0.188175 | 0.003234 | -0.041323 | 0.521450 |
| lwic_They | | 3 | | | 1 | -0.009543 | 0.882343 | 0.197550 | 0.001974 | 0.048037 | 0.456039 | 0.025149 | 0.696477 | 0.135038 | 0.035393 |
| lwic_Verbs | | 2 | | 1 | | -0.109378 | 0.088876 | -0.165872 | 0.009588 | -0.059774 | 0.353504 | -0.158253 | 0.013520 | -0.115775 | 0.071622 |
| lwic_We | | | 4 | | | 0.068851 | 0.285061 | 0.071206 | 0.268867 | -0.287303 | 0.000005 | 0.086444 | 0.179241 | -0.055504 | 0.389006 |
| lwic_Work | | 3 | | 1 | | -0.036610 | 0.570078 | 0.228270 | 0.000334 | 0.079075 | 0.219363 | -0.134963 | 0.035497 | 0.123435 | 0.054657 |
| MRC features | | | | | | | | | | | | | | | |
| mrc_irreg_P | 1 | | 1 | | | -0.139806 | 0.029344 | 0.057210 | 0.374571 | -0.135952 | 0.034159 | 0.061730 | 0.337950 | -0.013337 | 0.836136 |
| mrc_status_Q | | | | | 1 | -0.059691 | 0.354179 | -0.053685 | 0.404757 | 0.121702 | 0.058170 | 0.038019 | 0.555318 | 0.126970 | 0.048034 |
| mrc_status_W | | 1 | | 3 | | 0.040867 | 0.526062 | 0.132043 | 0.039711 | -0.009330 | 0.884956 | 0.217735 | 0.000631 | -0.118830 | 0.064404 |
| mrc_status_E | | 2 | 3 | | 1 | -0.006923 | 0.914503 | 0.175017 | 0.006232 | -0.240386 | 0.000155 | 0.037964 | 0.555885 | -0.130868 | 0.041521 |
| mrc_status_P | 1 | | | | 1 | -0.131677 | 0.040267 | 0.057210 | 0.374571 | -0.100687 | 0.117479 | -0.068564 | 0.287080 | -0.148922 | 0.020208 |
| mrc_status_$ | | | 4 | | | 0.001299 | 0.983933 | 0.085043 | 0.186420 | -0.306971 | 0.000001 | 0.043079 | 0.503892 | -0.029891 | 0.642892 |
| mrc_status_S | | | 1 | | | -0.026581 | 0.680130 | -0.019472 | 0.762655 | 0.131750 | 0.040156 | -0.055446 | 0.389503 | -0.083655 | 0.193734 |
| TweetNLP label features | | | | | | | | | | | | | | | |
| tl_! | | 3 | 3 | | | 0.009614 | 0.881475 | -0.243474 | 0.000126 | 0.204688 | 0.001335 | 0.067845 | 0.292175 | -0.052742 | 0.413077 |
| tl_# | | | 1 | | | -0.041758 | 0.517068 | 0.119092 | 0.063813 | 0.139911 | 0.029221 | 0.003483 | 0.956928 | 0.026876 | 0.676772 |
| tl_& | 1 | 1 | | | | -0.144235 | 0.024539 | -0.133912 | 0.036967 | 0.001422 | 0.982407 | -0.124004 | 0.053543 | 0.006115 | 0.924446 |
| tl_, | | 1 | | | 1 | 0.100379 | 0.118613 | 0.135598 | 0.034633 | -0.008509 | 0.895021 | -0.054947 | 0.393790 | 0.148938 | 0.020195 |
| tl_@ | | | | | 1 | -0.030139 | 0.640139 | -0.065513 | 0.309116 | -0.041248 | 0.522208 | 0.041502 | 0.519640 | 0.134970 | 0.035487 |
| tl_^ | | 1 | 1 | | 3 | 0.034813 | 0.589161 | 0.140151 | 0.028943 | 0.158338 | 0.013469 | -0.106201 | 0.098606 | 0.241325 | 0.000145 |
| tl_A | | 1 | | 1 | 3 | 0.008102 | 0.900007 | 0.132610 | 0.038861 | 0.101166 | 0.115734 | -0.129962 | 0.042965 | 0.236197 | 0.000203 |
| tl_D | 1 | 3 | | | | 0.134947 | 0.035518 | 0.244866 | 0.000115 | -0.083135 | 0.196528 | -0.006562 | 0.918941 | 0.057757 | 0.370009 |
| tl_E | | | 4 | | 3 | -0.089371 | 0.164906 | -0.024304 | 0.706197 | -0.328777 | 0.000000 | -0.002149 | 0.973420 | -0.244760 | 0.000116 |
| tl_G | | | 1 | 2 | | 0.009451 | 0.883471 | 0.037285 | 0.562980 | -0.152913 | 0.017059 | 0.178026 | 0.005384 | -0.003913 | 0.951611 |
| tl_L | | 4 | | 2 | 1 | 0.023037 | 0.720863 | -0.276206 | 0.000012 | -0.007734 | 0.904529 | -0.165690 | 0.009669 | 0.142893 | 0.025918 |
| tl_N | | 3 | | | 1 | -0.083274 | 0.195782 | 0.234006 | 0.000233 | 0.075669 | 0.239929 | -0.023193 | 0.719049 | 0.129700 | 0.043390 |
| tl_O | 1 | 3 | 1 | 2 | | -0.156327 | 0.014714 | -0.227567 | 0.000349 | 0.135075 | 0.035343 | -0.176480 | 0.005806 | 0.002027 | 0.974917 |
| tl_P | | 1 | | | 1 | 0.010892 | 0.865855 | 0.143038 | 0.025766 | 0.110503 | 0.085622 | -0.088833 | 0.167474 | 0.164049 | 0.010423 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tl_R | | 3 | 3 | | | 0.013122 | 0.838735 | -0.191120 | 0.002776 | -0.181339 | 0.004572 | 0.066040 | 0.305231 | -0.077399 | 0.229319 |
| tl_T | | | | | 1 | -0.037432 | 0.561438 | 0.014570 | 0.821233 | 0.056640 | 0.379359 | -0.103526 | 0.107437 | 0.156239 | 0.014771 |
| tl_V | | | 2 | | | 0.094844 | 0.140436 | -0.038560 | 0.549694 | -0.165153 | 0.009910 | 0.030796 | 0.632868 | -0.034299 | 0.594677 |
| TweetNLP cluster features | | | | | | | | | | | | | | | |
| tc_0000 | | 1 | | 1 | 3 | -0.030599 | 0.635045 | -0.145389 | 0.023405 | 0.122779 | 0.055965 | -0.131110 | 0.041143 | 0.193301 | 0.002475 |
| tc_000100 | | | 1 | | | -0.107360 | 0.094965 | -0.018652 | 0.772360 | 0.132264 | 0.039377 | 0.089643 | 0.163621 | 0.028520 | 0.658213 |
| tc_000101 | 1 | 1 | | | | -0.148295 | 0.020746 | 0.139462 | 0.029748 | 0.055134 | 0.392176 | 0.065507 | 0.309165 | -0.068460 | 0.287811 |
| tc_001000 | 3 | | | | 1 | 0.191033 | 0.002789 | -0.077919 | 0.226194 | -0.055447 | 0.389493 | 0.042774 | 0.506912 | -0.129449 | 0.043800 |
| tc_001001 | 1 | | | | | -0.126183 | 0.049448 | 0.027809 | 0.666220 | 0.112356 | 0.080471 | 0.004035 | 0.950102 | -0.092173 | 0.152007 |
| tc_0010100 | | 3 | | | | -0.033507 | 0.603221 | -0.190637 | 0.002847 | 0.069631 | 0.279631 | -0.039002 | 0.545131 | -0.074346 | 0.248268 |
| tc_001010100 | | 2 | 1 | | | -0.124962 | 0.051708 | 0.166197 | 0.009446 | 0.152736 | 0.017189 | 0.031150 | 0.628957 | -0.027076 | 0.674502 |
| tc_001010101 | | | | 1 | 1 | 0.031747 | 0.622391 | -0.062080 | 0.335213 | 0.011055 | 0.863872 | 0.131144 | 0.041089 | -0.145413 | 0.023382 |
| tc_001010111 | | 1 | | | 1 | 0.080222 | 0.212726 | -0.135064 | 0.035359 | 0.090189 | 0.161058 | -0.083600 | 0.194032 | 0.154103 | 0.016207 |
| tc_00101100 | | 1 | | | 2 | -0.000669 | 0.991726 | -0.141351 | 0.027584 | -0.052100 | 0.418788 | -0.026442 | 0.681710 | 0.168714 | 0.008405 |
| tc_00101101 | | 1 | | 2 | | -0.018165 | 0.778147 | -0.154877 | 0.015673 | -0.036864 | 0.567405 | -0.169434 | 0.008127 | -0.078762 | 0.221200 |
| tc_00101110 | | 1 | 3 | | | 0.031403 | 0.626174 | -0.154200 | 0.016140 | -0.230162 | 0.000297 | 0.022817 | 0.723413 | -0.010741 | 0.867700 |
| tc_00101111010 | 1 | | | 3 | | -0.146115 | 0.022714 | 0.006745 | 0.916695 | 0.069178 | 0.282776 | -0.229815 | 0.000303 | 0.082870 | 0.197964 |
| tc_001011110111 | | | | | 1 | -0.016748 | 0.795058 | 0.003215 | 0.960233 | -0.116427 | 0.070028 | 0.057474 | 0.372369 | 0.160460 | 0.012256 |
| tc_001011111010 | | 3 | | | 2 | 0.028494 | 0.658503 | -0.180689 | 0.004722 | 0.103525 | 0.107440 | 0.003088 | 0.961801 | 0.168261 | 0.008585 |
| tc_0010111110110 | | 1 | | | 3 | -0.054427 | 0.398284 | -0.130885 | 0.041495 | -0.093845 | 0.144680 | 0.101294 | 0.115272 | -0.186995 | 0.003436 |
| tc_0010111110111 | | 4 | 3 | | | -0.017077 | 0.791130 | -0.281161 | 0.000009 | -0.193115 | 0.002500 | 0.008743 | 0.892145 | -0.102487 | 0.111031 |
| tc_001011111111 | | 1 | 3 | | | 0.039166 | 0.543434 | -0.134273 | 0.036457 | 0.215329 | 0.000727 | 0.060321 | 0.349112 | -0.047989 | 0.456488 |
| tc_0011000 | 1 | | | | | 0.161806 | 0.011538 | -0.107823 | 0.093539 | 0.008436 | 0.895907 | 0.007099 | 0.912330 | 0.114091 | 0.075876 |
| tc_0011011110 | | | 3 | | | -0.096002 | 0.135634 | 0.001115 | 0.986202 | -0.194905 | 0.002274 | -0.086302 | 0.179957 | -0.062476 | 0.332134 |
| tc_0011011111 | | | 3 | | | -0.061558 | 0.339302 | -0.111660 | 0.082375 | 0.186207 | 0.003577 | -0.055947 | 0.385228 | -0.009270 | 0.885685 |
| tc_001110100 | | | 1 | | | 0.026734 | 0.678384 | -0.011746 | 0.855453 | -0.146621 | 0.022243 | 0.028916 | 0.653771 | -0.044781 | 0.487171 |
| tc_0011101010 | | 1 | 1 | | | -0.013175 | 0.838094 | 0.132868 | 0.038479 | 0.128134 | 0.046006 | -0.019855 | 0.758122 | -0.050592 | 0.432399 |
| tc_00111010110 | 1 | | | | | 0.128699 | 0.045047 | 0.015191 | 0.813741 | 0.012993 | 0.840307 | 0.022278 | 0.729693 | 0.054651 | 0.396342 |
| tc_00111010111 | | | 3 | | | -0.076117 | 0.237146 | 0.117924 | 0.066479 | 0.191804 | 0.002678 | -0.030084 | 0.640744 | 0.012036 | 0.851921 |
| tc_001110110 | | | | 1 | | -0.103545 | 0.107371 | -0.035629 | 0.580459 | 0.102340 | 0.111548 | -0.146960 | 0.021933 | 0.099699 | 0.121139 |
| tc_001110111 | 1 | | | | | 0.130754 | 0.041701 | 0.064242 | 0.318618 | -0.104271 | 0.104918 | 0.120177 | 0.061415 | 0.053343 | 0.407766 |
| tc_00111100 | | 1 | 1 | | | -0.095308 | 0.138497 | -0.127246 | 0.047547 | -0.160935 | 0.011999 | -0.108062 | 0.092811 | -0.059468 | 0.355981 |
| tc_0011110110 | 2 | 1 | | 3 | | -0.169706 | 0.008024 | -0.131206 | 0.040994 | 0.061309 | 0.341260 | -0.186929 | 0.003448 | 0.002137 | 0.973566 |
| tc_00111110 | | | | | 1 | 0.042504 | 0.509606 | -0.082547 | 0.199727 | 0.003296 | 0.959237 | 0.004621 | 0.942872 | -0.129602 | 0.043549 |
| tc_001111111 | | | 2 | | | -0.037961 | 0.555916 | 0.101436 | 0.114762 | -0.165523 | 0.009744 | 0.026583 | 0.680096 | -0.021282 | 0.741342 |
| tc_01000010 | | | 2 | | | -0.045798 | 0.477321 | 0.001686 | 0.979141 | 0.172430 | 0.007054 | -0.054550 | 0.397218 | 0.118760 | 0.064561 |
| tc_010001001 | 2 | 3 | | | | -0.175006 | 0.006235 | -0.192318 | 0.002607 | 0.125643 | 0.050437 | 0.017411 | 0.787133 | 0.072517 | 0.260131 |
| tc_01000101010 | | 4 | | | 1 | -0.012253 | 0.849283 | -0.258556 | 0.000045 | -0.082779 | 0.198461 | -0.003029 | 0.962538 | -0.133887 | 0.037003 |
| tc_010001010110 | | | 3 | | | -0.086343 | 0.179753 | -0.074893 | 0.244794 | -0.220655 | 0.000531 | -0.042071 | 0.513934 | -0.006847 | 0.915441 |
| tc_0100010110 | | | 2 | | | 0.044365 | 0.491229 | 0.039391 | 0.541122 | -0.176580 | 0.005778 | -0.027881 | 0.665399 | -0.040406 | 0.530745 |
| tc_0100010111010 | | | | 1 | | 0.043658 | 0.498168 | -0.043338 | 0.501330 | 0.014283 | 0.824696 | 0.149714 | 0.019546 | -0.006963 | 0.914007 |
| tc_010001011110 | | | | 1 | | 0.036328 | 0.573056 | 0.124320 | 0.052932 | 0.014421 | 0.823031 | -0.157472 | 0.013993 | -0.061077 | 0.343093 |
| tc_010001011111 | | 1 | 4 | | | -0.114305 | 0.075326 | -0.162210 | 0.011330 | -0.318089 | 0.000000 | -0.013503 | 0.834126 | -0.038576 | 0.549533 |
| tc_0100011010 | | | 2 | 1 | | 0.112345 | 0.080499 | -0.005539 | 0.931551 | 0.173886 | 0.006580 | 0.135848 | 0.034297 | -0.080041 | 0.213763 |
| tc_0100011011 | | | 3 | | | -0.050702 | 0.431403 | -0.108543 | 0.091355 | -0.209946 | 0.000993 | -0.063281 | 0.325930 | -0.101945 | 0.112943 |
| tc_01000111010 | | | | | 1 | 0.065032 | 0.312688 | 0.024533 | 0.703563 | -0.104923 | 0.102750 | 0.085527 | 0.183914 | -0.150491 | 0.018915 |
| tc_01000111110 | | | | 1 | | 0.080173 | 0.213007 | 0.050740 | 0.431054 | 0.101582 | 0.114236 | -0.138620 | 0.030759 | 0.092899 | 0.148793 |
| tc_010001111110 | | | 3 | | | 0.123864 | 0.053815 | -0.019482 | 0.762538 | -0.187399 | 0.003366 | 0.024698 | 0.701664 | -0.033820 | 0.599834 |
| tc_0100100 | | | 3 | | | 0.058730 | 0.361994 | -0.116669 | 0.069443 | -0.183867 | 0.004028 | -0.056886 | 0.377284 | 0.023361 | 0.717099 |
| tc_01001010 | | | | | 2 | 0.040555 | 0.529222 | -0.037286 | 0.562968 | -0.093304 | 0.147023 | 0.096367 | 0.134143 | -0.178353 | 0.005299 |
| tc_01001011 | | | 1 | | | -0.034356 | 0.594067 | -0.079793 | 0.215195 | -0.127578 | 0.046964 | -0.029182 | 0.650800 | 0.081660 | 0.204618 |
| tc_010011000 | | | 1 | | | 0.001102 | 0.986359 | 0.119317 | 0.063308 | -0.139256 | 0.029994 | -0.008322 | 0.897306 | -0.081114 | 0.207672 |
| tc_01001101000 | | | | | 1 | -0.023313 | 0.717659 | 0.083922 | 0.192313 | -0.075575 | 0.240517 | 0.044454 | 0.490363 | -0.136734 | 0.033130 |
| tc_01001101001 | | | | 3 | | -0.124102 | 0.053353 | 0.091879 | 0.153323 | 0.004993 | 0.938282 | -0.210523 | 0.000960 | -0.039225 | 0.542829 |
| tc_010011010100 | | 3 | | | | -0.060827 | 0.345073 | -0.183107 | 0.004185 | -0.028739 | 0.655759 | 0.039308 | 0.541977 | 0.026898 | 0.676524 |
| tc_010011010111 | | | 1 | | | 0.084286 | 0.190381 | 0.018087 | 0.779085 | -0.143881 | 0.024896 | 0.072684 | 0.259031 | 0.057828 | 0.369427 |
| tc_010011011100 | | | | | 2 | 0.068829 | 0.285216 | 0.104493 | 0.104176 | 0.114960 | 0.073655 | -0.059876 | 0.352686 | 0.169431 | 0.008128 |
| tc_010011011101 | | 2 | 1 | | 2 | -0.121992 | 0.057571 | 0.168632 | 0.008437 | -0.159335 | 0.012887 | 0.040474 | 0.530050 | -0.169869 | 0.007963 |
| tc_0100110111101 | 2 | | | 3 | | -0.176589 | 0.005775 | 0.006800 | 0.916018 | 0.059527 | 0.355500 | -0.209148 | 0.001039 | -0.011485 | 0.858629 |
| tc_01001110011 | | | | 1 | | 0.065851 | 0.306621 | -0.061818 | 0.337260 | 0.062719 | 0.330248 | 0.126033 | 0.049721 | -0.045769 | 0.477599 |
| tc_010011110 | | | 1 | | | 0.049111 | 0.446015 | 0.099479 | 0.121968 | 0.137742 | 0.031844 | -0.052696 | 0.413481 | 0.007734 | 0.904531 |
| tc_0100111110 | | | | 3 | | 0.016372 | 0.799558 | -0.056268 | 0.382499 | 0.072084 | 0.262993 | -0.198063 | 0.001920 | 0.106144 | 0.098788 |
| tc_01001111111 | 1 | | | | | -0.147880 | 0.021109 | 0.103955 | 0.105981 | -0.027442 | 0.670357 | 0.055710 | 0.387240 | 0.033589 | 0.602337 |
| tc_01010000000 | | 1 | | | | 0.029937 | 0.642382 | 0.134479 | 0.036167 | 0.004221 | 0.947807 | 0.027626 | 0.668282 | -0.055717 | 0.387187 |
| tc_01010000001 | | | | | 1 | -0.098813 | 0.124499 | 0.105831 | 0.099793 | 0.024951 | 0.698756 | 0.108537 | 0.091375 | -0.146113 | 0.022717 |
| tc_010100000100 | | 1 | | | 3 | 0.044418 | 0.490715 | 0.152870 | 0.017090 | -0.058146 | 0.366793 | 0.003517 | 0.956497 | 0.227471 | 0.000351 |
| tc_010100001110 | | | | | 2 | -0.077235 | 0.230312 | -0.068079 | 0.290508 | -0.079359 | 0.217709 | -0.057410 | 0.372900 | -0.169137 | 0.008241 |
| tc_01010001100 | | 2 | | | 1 | -0.028745 | 0.655689 | 0.169047 | 0.008275 | -0.015066 | 0.815255 | 0.048969 | 0.447331 | 0.147473 | 0.021470 |
| tc_01010001111 | | 1 | | | 1 | 0.065721 | 0.307583 | 0.129877 | 0.043101 | 0.047698 | 0.459223 | 0.013064 | 0.839449 | -0.155788 | 0.015064 |
| tc_01010010010 | 2 | | | 3 | | -0.165149 | 0.009912 | 0.060626 | 0.346671 | -0.024435 | 0.704696 | -0.189411 | 0.003034 | 0.015463 | 0.810470 |
| tc_01010010011 | | | 4 | | | 0.123596 | 0.054339 | 0.000901 | 0.988850 | -0.290451 | 0.000004 | 0.025082 | 0.697255 | -0.067113 | 0.297429 |
| tc_01010011010 | | | | 2 | | -0.005416 | 0.933066 | 0.094915 | 0.140137 | -0.043200 | 0.502696 | -0.171196 | 0.007480 | 0.082141 | 0.201952 |
| tc_01010011100 | | 1 | 3 | | | -0.003928 | 0.951425 | 0.129413 | 0.043860 | 0.199154 | 0.001810 | -0.016093 | 0.802909 | -0.081559 | 0.205178 |
| tc_010100111101 | 1 | | | | | -0.140264 | 0.028812 | 0.025642 | 0.690831 | 0.027635 | 0.668178 | -0.090187 | 0.161070 | 0.072257 | 0.261847 |
| tc_010101000010 | | 1 | | 1 | | 0.083132 | 0.196544 | 0.128152 | 0.045975 | -0.043630 | 0.498448 | 0.144856 | 0.023923 | -0.002594 | 0.967908 |
| tc_010101000011 | | | | 2 | | -0.076793 | 0.232998 | 0.089919 | 0.162324 | -0.071375 | 0.267729 | -0.177343 | 0.005567 | -0.066848 | 0.299340 |
| tc_010101000110 | | | 1 | | | 0.001627 | 0.979875 | 0.087007 | 0.176414 | -0.131165 | 0.041058 | -0.091617 | 0.154502 | 0.108350 | 0.091939 |
| tc_01010100100 | | | 2 | 1 | | -0.109033 | 0.089894 | -0.082287 | 0.201149 | 0.170137 | 0.007863 | -0.137081 | 0.032683 | 0.002997 | 0.962928 |
| tc_01010100110 | 1 | 3 | | | 2 | 0.161071 | 0.011926 | -0.203832 | 0.001400 | -0.093557 | 0.145922 | 0.112989 | 0.078769 | 0.171718 | 0.007297 |
| tc_010101001110 | | | | | 2 | 0.111491 | 0.082843 | -0.069896 | 0.277797 | -0.086050 | 0.181239 | -0.106082 | 0.098986 | -0.178357 | 0.005297 |
| tc_01010101001 | | | | 1 | | -0.100900 | 0.116703 | 0.013862 | 0.829776 | -0.124713 | 0.052180 | -0.142213 | 0.026642 | -0.085163 | 0.185794 |
| tc_010101011110 | | 1 | | | 1 | -0.098979 | 0.123863 | 0.130360 | 0.042326 | 0.103002 | 0.109238 | -0.079277 | 0.218185 | 0.158762 | 0.013219 |
| tc_010101101 | | 1 | | | | 0.093122 | 0.147815 | -0.128599 | 0.045216 | -0.062128 | 0.334834 | 0.069331 | 0.281712 | 0.063230 | 0.326314 |
| tc_0101011110 | | | 1 | | | -0.116098 | 0.070829 | -0.093423 | 0.146504 | 0.147446 | 0.021495 | 0.021184 | 0.742489 | -0.101768 | 0.113571 |
| tc_0101100001 | 1 | | | | | -0.134286 | 0.036439 | -0.033063 | 0.608036 | 0.021025 | 0.744356 | -0.050297 | 0.435092 | -0.018197 | 0.777775 |
| tc_01011000110 | | | 1 | | | 0.011868 | 0.853974 | -0.071833 | 0.264666 | 0.129786 | 0.043249 | -0.059733 | 0.353837 | -0.032560 | 0.613509 |
| tc_010110010010 | | | 3 | | | -0.004027 | 0.950206 | -0.054821 | 0.394880 | 0.182795 | 0.004251 | -0.031477 | 0.625356 | 0.045801 | 0.477294 |
| tc_010110010011 | | 1 | 2 | | | -0.001457 | 0.981978 | -0.131803 | 0.040075 | 0.171962 | 0.007213 | -0.054048 | 0.401582 | -0.019173 | 0.766187 |
| tc_010110010110 | | | 2 | | | -0.091765 | 0.153838 | 0.003458 | 0.957228 | 0.169812 | 0.007984 | 0.102093 | 0.112417 | 0.019416 | 0.763317 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tc_010110011000 | 1 | | | | | 0.155253 | 0.015419 | 0.098884 | 0.124228 | 0.023174 | 0.719276 | 0.017494 | 0.786145 | -0.004088 | 0.949452 |
| tc_0101100110010 | | 1 | | | 1 | -0.106964 | 0.096195 | -0.153869 | 0.016371 | 0.059772 | 0.353523 | -0.021943 | 0.733609 | -0.157564 | 0.013937 |
| tc_010110011011 | | 1 | | | 3 | 0.010424 | 0.871570 | 0.137899 | 0.031648 | 0.065709 | 0.307672 | -0.056270 | 0.382485 | 0.206418 | 0.001212 |
| tc_0101100111011 | | | | | 1 | -0.018232 | 0.777354 | 0.102139 | 0.112255 | 0.023010 | 0.721177 | 0.037937 | 0.556170 | 0.136934 | 0.032873 |
| tc_0101100111101 | | | 1 | | | -0.016038 | 0.803563 | 0.021610 | 0.737501 | 0.133306 | 0.037839 | 0.122931 | 0.055660 | -0.106892 | 0.096422 |
| tc_0101100111111 | | | 1 | 1 | | -0.070487 | 0.273743 | 0.103116 | 0.108843 | 0.133898 | 0.036987 | -0.135671 | 0.034534 | 0.116316 | 0.070299 |
| tc_01011011100 | | 1 | | 1 | 1 | -0.085335 | 0.184904 | 0.164488 | 0.010217 | 0.024348 | 0.705689 | -0.129062 | 0.044441 | -0.136990 | 0.032799 |
| tc_0101101111000 | | | 1 | | | 0.069977 | 0.277238 | -0.048692 | 0.449904 | -0.142307 | 0.026541 | -0.028289 | 0.660806 | 0.119104 | 0.063784 |
| tc_0101101111010 | | | 1 | | | -0.001166 | 0.985568 | 0.083043 | 0.197027 | 0.143842 | 0.024937 | 0.014072 | 0.827240 | 0.051608 | 0.423204 |
| tc_0101101111011 | 1 | | | | | -0.126527 | 0.048826 | 0.104953 | 0.102652 | -0.054616 | 0.396651 | -0.049905 | 0.438688 | -0.087038 | 0.176262 |
| tc_01011110010 | | | | | 1 | 0.063571 | 0.323711 | -0.036929 | 0.566721 | 0.021213 | 0.742153 | 0.046481 | 0.470776 | -0.130132 | 0.042691 |
| tc_01011110011 | | 1 | 4 | | | 0.007446 | 0.908067 | -0.128639 | 0.045147 | -0.280722 | 0.000009 | -0.104164 | 0.105276 | -0.007773 | 0.904056 |
| tc_010111110 | | | 1 | | 3 | 0.046446 | 0.471109 | 0.026520 | 0.680818 | 0.131196 | 0.041008 | 0.112508 | 0.080058 | 0.222690 | 0.000470 |
| tc_01011111111 | | | | 1 | | -0.048954 | 0.447474 | 0.047733 | 0.458893 | -0.049578 | 0.441697 | -0.154678 | 0.015809 | 0.108490 | 0.091516 |
| tc_011000001 | | 1 | 1 | | | 0.079406 | 0.217435 | -0.141669 | 0.027233 | -0.137935 | 0.031603 | 0.029439 | 0.647928 | 0.040783 | 0.526906 |
| tc_0110000111 | 2 | 3 | 3 | | | -0.169537 | 0.008087 | -0.200299 | 0.001700 | -0.196122 | 0.002131 | -0.047313 | 0.462856 | 0.012147 | 0.850580 |
| tc_0110001011 | | | 1 | | | 0.119944 | 0.061923 | -0.075263 | 0.242466 | -0.144391 | 0.024383 | 0.023045 | 0.720766 | 0.095862 | 0.136207 |
| tc_011000110 | | 3 | | | | -0.097906 | 0.128011 | -0.214913 | 0.000745 | -0.044143 | 0.493404 | 0.002658 | 0.967125 | 0.057705 | 0.370447 |
| tc_0110001111 | | 2 | | 2 | 1 | 0.035536 | 0.581444 | -0.167259 | 0.008994 | 0.045842 | 0.476902 | -0.171711 | 0.007300 | 0.134163 | 0.036612 |
| tc_0110010001 | | | | 1 | 1 | -0.044152 | 0.493312 | -0.115869 | 0.071390 | 0.054311 | 0.399291 | -0.131801 | 0.040078 | 0.147807 | 0.021173 |
| tc_0110010010 | | | 2 | | | -0.076042 | 0.237611 | -0.053451 | 0.406817 | 0.174063 | 0.006525 | 0.017009 | 0.791943 | 0.001947 | 0.975915 |
| tc_011001010 | | | | | 1 | -0.123823 | 0.053896 | -0.021826 | 0.734968 | 0.067299 | 0.296087 | -0.022902 | 0.722429 | 0.137559 | 0.032074 |
| tc_011010000 | | 1 | | | | -0.058554 | 0.363437 | -0.145706 | 0.023101 | 0.015397 | 0.811265 | 0.034051 | 0.597344 | 0.058119 | 0.367020 |
| tc_0110100010 | | 1 | 1 | | | -0.040945 | 0.525266 | -0.146147 | 0.022684 | -0.155470 | 0.015274 | -0.076642 | 0.233923 | -0.075874 | 0.238652 |
| tc_011010001100 | 1 | | | 1 | | -0.139977 | 0.029145 | -0.021999 | 0.732950 | 0.073897 | 0.251148 | -0.151978 | 0.017756 | 0.074647 | 0.246354 |
| tc_0110100011011 | | | 3 | | | 0.066263 | 0.303604 | -0.115936 | 0.071226 | -0.200360 | 0.001695 | -0.018824 | 0.770326 | 0.097601 | 0.129209 |
| tc_011010001110 | | | | | 1 | -0.029170 | 0.650928 | -0.051762 | 0.421818 | 0.057504 | 0.372118 | -0.054090 | 0.401220 | 0.127868 | 0.046463 |
| tc_0110100011110 | | | 1 | | | 0.103696 | 0.106859 | -0.089807 | 0.162851 | -0.163021 | 0.010922 | 0.091890 | 0.153276 | 0.043185 | 0.502838 |
| tc_0110100011111 | | 1 | | 1 | | -0.056780 | 0.378178 | -0.137746 | 0.031839 | 0.023398 | 0.716679 | 0.154612 | 0.015854 | -0.031388 | 0.626333 |
| tc_01101001010 | | 3 | | | | 0.048270 | 0.453853 | -0.187022 | 0.003431 | -0.059695 | 0.354141 | 0.096104 | 0.135216 | -0.085847 | 0.182274 |
| tc_01101001011 | | 3 | 1 | | | 0.065872 | 0.306468 | -0.208396 | 0.001084 | 0.154085 | 0.016220 | -0.071404 | 0.267535 | 0.003183 | 0.960630 |
| tc_0110100110 | | | 2 | | | -0.089042 | 0.166474 | -0.020893 | 0.745904 | 0.177043 | 0.005649 | -0.078005 | 0.225684 | 0.058941 | 0.360265 |
| tc_0110100111 | 1 | 1 | 1 | | | -0.144542 | 0.024233 | -0.149890 | 0.019401 | 0.126267 | 0.049295 | 0.114099 | 0.075855 | 0.036891 | 0.567111 |
| tc_011010100 | | | 1 | | | -0.087643 | 0.173260 | 0.027376 | 0.671104 | -0.142385 | 0.026457 | -0.032960 | 0.609155 | -0.002228 | 0.972432 |
| tc_0110101110 | | | | 2 | 4 | 0.024296 | 0.706294 | -0.079091 | 0.219267 | 0.025180 | 0.696131 | -0.167868 | 0.008743 | 0.282130 | 0.000008 |
| tc_01101100001 | | | 3 | | | -0.081752 | 0.204103 | -0.032542 | 0.613699 | -0.185587 | 0.003692 | -0.043282 | 0.501883 | 0.030463 | 0.636552 |
| tc_011011000101 | | | | | 3 | 0.099889 | 0.120430 | -0.071971 | 0.263746 | -0.050238 | 0.435637 | 0.081508 | 0.205464 | -0.221852 | 0.000494 |
| tc_011011000110 | | | 2 | | | -0.028235 | 0.661413 | -0.101910 | 0.113066 | -0.176202 | 0.005885 | 0.011640 | 0.856745 | 0.027074 | 0.674531 |
| tc_0110110001110 | | 1 | | | 1 | 0.082409 | 0.200478 | 0.132005 | 0.039768 | -0.070382 | 0.274464 | -0.045413 | 0.481038 | 0.127188 | 0.047648 |
| tc_0110110010 | | | 3 | | | 0.079867 | 0.214765 | 0.017211 | 0.789528 | -0.217060 | 0.000657 | 0.114691 | 0.074338 | -0.121688 | 0.058199 |
| tc_0110110110 | | | | | 1 | 0.116456 | 0.069959 | -0.124796 | 0.052022 | 0.107284 | 0.095200 | 0.019881 | 0.757818 | 0.134751 | 0.035789 |
| tc_0110110111 | 1 | | 1 | 1 | | 0.155918 | 0.014979 | -0.096391 | 0.134046 | -0.135960 | 0.034147 | 0.145858 | 0.022957 | 0.083637 | 0.193833 |
| tc_011011100 | | 3 | 3 | | | -0.111819 | 0.081937 | -0.198159 | 0.001910 | -0.196820 | 0.002053 | 0.067868 | 0.292010 | -0.081335 | 0.206430 |
| tc_0110111011 | 1 | | | | | 0.145446 | 0.023350 | 0.102115 | 0.112342 | 0.029390 | 0.648471 | -0.116018 | 0.071026 | 0.092068 | 0.152479 |
| tc_011011110 | | | | | 4 | -0.028595 | 0.657370 | 0.013953 | 0.828680 | -0.084918 | 0.187071 | 0.029315 | 0.649317 | -0.262192 | 0.000035 |
| tc_011011111010 | 2 | | | | | -0.170537 | 0.007716 | 0.038214 | 0.553290 | 0.103314 | 0.108164 | 0.004035 | 0.950105 | 0.008940 | 0.889734 |
| tc_011011111011 | | | 3 | 1 | | 0.047746 | 0.458775 | 0.047988 | 0.456499 | -0.238865 | 0.000171 | 0.129842 | 0.043159 | -0.056182 | 0.383228 |
| tc_0110111111000 | | | | 1 | | -0.093580 | 0.145824 | -0.056379 | 0.381560 | -0.021613 | 0.737459 | -0.139105 | 0.030174 | 0.108898 | 0.090295 |
| tc_01101111110101 | | 3 | | | | -0.098751 | 0.124736 | 0.203812 | 0.001402 | 0.061184 | 0.342244 | 0.107191 | 0.095488 | 0.004019 | 0.950304 |
| tc_01101111110110 | 1 | | | | | 0.139211 | 0.030046 | 0.047746 | 0.458767 | -0.063470 | 0.324480 | 0.086063 | 0.181171 | 0.070104 | 0.276364 |
| tc_011011111101110 | | | | | 1 | -0.072202 | 0.262210 | 0.112031 | 0.081355 | 0.080384 | 0.211804 | -0.002086 | 0.974197 | 0.153260 | 0.016807 |
| tc_011011111101111 | | | | 1 | | -0.056614 | 0.379572 | -0.081679 | 0.204510 | -0.037399 | 0.561785 | -0.156042 | 0.014898 | 0.002305 | 0.971490 |
| tc_01101111111011 | | | | | 1 | -0.032665 | 0.612361 | 0.056438 | 0.381058 | 0.067535 | 0.294391 | -0.005779 | 0.928594 | -0.155726 | 0.015105 |
| tc_0110111111110 | | | 3 | | | -0.067513 | 0.294552 | 0.010133 | 0.875125 | -0.185447 | 0.003718 | -0.088350 | 0.169805 | -0.006223 | 0.923121 |
| tc_01110000 | | | | 2 | 2 | -0.052251 | 0.417441 | -0.033390 | 0.604488 | -0.050791 | 0.430593 | -0.174575 | 0.006366 | 0.172431 | 0.007054 |
| tc_011100010 | | 1 | | | 1 | 0.066919 | 0.298826 | 0.145333 | 0.023459 | -0.070239 | 0.275441 | -0.041411 | 0.520565 | 0.153883 | 0.016362 |
| tc_011100011 | | | 3 | | | 0.041317 | 0.521514 | -0.015546 | 0.809481 | -0.212303 | 0.000867 | 0.015441 | 0.810737 | 0.037583 | 0.559866 |
| tc_01110010 | | 1 | 4 | 1 | | 0.085810 | 0.182463 | -0.132363 | 0.039229 | -0.247714 | 0.000095 | 0.134107 | 0.036690 | -0.031374 | 0.626491 |
| tc_011100111110 | | 1 | | | 1 | 0.009822 | 0.878927 | 0.133780 | 0.037155 | 0.088468 | 0.169236 | 0.117016 | 0.068614 | -0.128314 | 0.045698 |
| tc_01110100 | | 1 | | 1 | 1 | -0.010670 | 0.868568 | -0.137876 | 0.031677 | -0.027788 | 0.666456 | -0.155246 | 0.015424 | 0.141238 | 0.027710 |
| tc_011101010 | | 2 | 1 | | | -0.042718 | 0.507470 | 0.170679 | 0.007665 | 0.154849 | 0.015692 | 0.067996 | 0.291101 | 0.002605 | 0.967775 |
| tc_011101011 | | | 1 | | | -0.027969 | 0.664416 | -0.058644 | 0.362698 | 0.131825 | 0.040041 | 0.012445 | 0.846958 | 0.059976 | 0.351878 |
| tc_01110111000 | | | 1 | | | -0.000436 | 0.994604 | -0.009261 | 0.885804 | 0.127866 | 0.046466 | 0.106172 | 0.098699 | -0.097121 | 0.131111 |
| tc_01110111001 | | | | | 3 | 0.049372 | 0.443602 | 0.110100 | 0.086778 | -0.071535 | 0.266655 | -0.018863 | 0.769866 | -0.207323 | 0.001152 |
| tc_01110111010 | | 1 | 2 | 1 | | 0.033794 | 0.600121 | 0.128598 | 0.045217 | -0.168258 | 0.008586 | 0.128101 | 0.046062 | -0.049786 | 0.439782 |
| tc_011101110111 | | | 2 | | | -0.053833 | 0.403461 | -0.049613 | 0.441373 | 0.170045 | 0.007897 | 0.026439 | 0.681741 | 0.018026 | 0.779801 |
| tc_0111011110 | | | 1 | | | -0.112478 | 0.080139 | -0.043848 | 0.496300 | 0.153499 | 0.016635 | -0.037641 | 0.559260 | -0.063398 | 0.325031 |
| tc_01110111110 | | | 1 | | | -0.047388 | 0.462149 | 0.063049 | 0.327704 | -0.136072 | 0.033999 | 0.002159 | 0.973292 | -0.080847 | 0.209177 |
| tc_011101111110 | | 1 | 2 | | | -0.114902 | 0.073802 | -0.162211 | 0.011329 | -0.172847 | 0.006916 | 0.025457 | 0.692957 | -0.031458 | 0.625565 |
| tc_0111100000 | | | 1 | | 2 | 0.052475 | 0.415449 | -0.022387 | 0.728430 | -0.153609 | 0.016556 | 0.022626 | 0.725642 | -0.171094 | 0.007516 |
| tc_01111000010 | | 1 | 1 | | | -0.056662 | 0.379169 | 0.157903 | 0.013730 | 0.145608 | 0.023194 | -0.102867 | 0.109705 | -0.011525 | 0.858144 |
| tc_01111000011 | | 1 | | | | -0.009560 | 0.882136 | 0.128004 | 0.046229 | 0.076132 | 0.237055 | -0.051883 | 0.420730 | -0.014481 | 0.822305 |
| tc_0111100010100 | 1 | | 1 | | 3 | -0.146689 | 0.022181 | 0.031768 | 0.622169 | -0.159448 | 0.012822 | -0.078984 | 0.219894 | -0.216418 | 0.000682 |
| tc_0111100010111 | | | 1 | | | 0.017563 | 0.785327 | 0.005529 | 0.931672 | -0.145277 | 0.023513 | 0.003652 | 0.954836 | -0.120565 | 0.060576 |
| tc_011110001100 | | 2 | 3 | | | -0.054546 | 0.397251 | -0.175567 | 0.006069 | -0.237163 | 0.000190 | 0.085729 | 0.182876 | -0.089335 | 0.165076 |
| tc_01111001000010 | | 1 | 1 | | | -0.064446 | 0.317082 | 0.134149 | 0.036632 | 0.141043 | 0.027927 | -0.015772 | 0.806755 | -0.008129 | 0.899674 |
| tc_0111100100011 | | 1 | | | | -0.033324 | 0.605201 | 0.152925 | 0.017050 | 0.123285 | 0.054954 | -0.102400 | 0.111336 | -0.068658 | 0.286419 |
| tc_011110010010 | | | | | 1 | -0.048547 | 0.451265 | 0.029799 | 0.643920 | -0.062038 | 0.335537 | -0.027016 | 0.675182 | 0.140111 | 0.028989 |
| tc_0111100100110 | | | 3 | | | 0.081569 | 0.205121 | -0.117804 | 0.066758 | 0.229387 | 0.000311 | -0.005303 | 0.934461 | 0.035981 | 0.576728 |
| tc_0111100100111 | 1 | | | | | 0.138494 | 0.030914 | -0.068886 | 0.284816 | 0.021256 | 0.741649 | -0.066842 | 0.299386 | -0.002647 | 0.967262 |
| tc_011110010100 | | | | 1 | | -0.057335 | 0.373522 | 0.044016 | 0.494650 | 0.037286 | 0.562965 | -0.130941 | 0.041407 | -0.024568 | 0.703154 |
| tc_0111100101010 | | | | 1 | 4 | 0.061891 | 0.336684 | 0.059389 | 0.356625 | -0.005804 | 0.928274 | -0.148223 | 0.020809 | 0.254610 | 0.000059 |
| tc_0111100101011 | 2 | | | 1 | | -0.173101 | 0.006832 | 0.057388 | 0.373081 | -0.019563 | 0.761570 | -0.126996 | 0.047987 | -0.005645 | 0.930245 |
| tc_0111100101110 | | | | 1 | | -0.083456 | 0.194802 | 0.047644 | 0.459733 | -0.051691 | 0.422460 | -0.130368 | 0.042313 | -0.061590 | 0.339049 |
| tc_0111100111000 | 1 | | | | | -0.157114 | 0.014215 | 0.015081 | 0.815066 | 0.006014 | 0.925696 | -0.044023 | 0.494582 | -0.007182 | 0.911314 |
| tc_011110011110 | | 3 | 3 | | | 0.050555 | 0.432743 | 0.192644 | 0.002563 | -0.182767 | 0.004257 | 0.029403 | 0.648335 | 0.091767 | 0.153828 |
| tc_0111100111110 | | | 3 | | 1 | -0.003728 | 0.953895 | 0.124695 | 0.052214 | 0.226479 | 0.000373 | -0.060322 | 0.349104 | 0.127378 | 0.047314 |
| tc_0111100111111 | 1 | | | | | 0.143474 | 0.025313 | 0.019405 | 0.763439 | 0.071710 | 0.265484 | -0.055888 | 0.385727 | -0.012617 | 0.844866 |
| tc_011110100010 | | | | | 1 | 0.008723 | 0.892386 | -0.053455 | 0.406778 | 0.110314 | 0.086161 | 0.055085 | 0.392597 | -0.147050 | 0.021852 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tc_01111010010 | | | 1 | | 3 | 0.053529 | 0.406123 | -0.065943 | 0.305949 | -0.154383 | 0.016012 | -0.110185 | 0.086533 | -0.224781 | 0.000414 |
| tc_011110100110 | | | 3 | | | -0.084605 | 0.188705 | -0.048713 | 0.449712 | -0.195894 | 0.002157 | 0.068775 | 0.285598 | -0.114114 | 0.075817 |
| tc_011110100111 | | | 1 | | | -0.011780 | 0.855037 | -0.040752 | 0.527228 | 0.144114 | 0.024661 | -0.049529 | 0.442146 | 0.016837 | 0.793999 |
| tc_011110101100 | | 3 | | | 1 | 0.055057 | 0.392844 | 0.180719 | 0.004715 | -0.074406 | 0.247888 | 0.036484 | 0.571402 | -0.145791 | 0.023020 |
| tc_011110101110 | | | 1 | | | 0.005944 | 0.926558 | -0.028624 | 0.657048 | 0.161977 | 0.011449 | -0.067170 | 0.297016 | -0.031528 | 0.624797 |
| tc_01111010111110 | | 3 | 1 | | | -0.085681 | 0.183123 | 0.195797 | 0.002168 | 0.133034 | 0.038235 | 0.026974 | 0.675664 | 0.073119 | 0.256187 |
| tc_011110110010 | | | | | 1 | -0.096028 | 0.135527 | -0.044113 | 0.493695 | -0.046911 | 0.466675 | -0.074015 | 0.250388 | -0.154848 | 0.015693 |
| tc_0111101101100 | | 1 | | | | -0.066098 | 0.304808 | 0.144666 | 0.024111 | -0.101221 | 0.115535 | 0.050265 | 0.435390 | 0.047321 | 0.462783 |
| tc_011110110111011 | | | | | 3 | -0.033696 | 0.601179 | -0.086561 | 0.178653 | 0.081196 | 0.207211 | 0.004709 | 0.941779 | 0.200182 | 0.001711 |
| tc_01111011011111 | | 1 | 1 | | | 0.014951 | 0.816639 | 0.159430 | 0.012832 | -0.154227 | 0.016120 | -0.034325 | 0.594400 | -0.011001 | 0.864527 |
| tc_0111101110010 | | 2 | 3 | | | 0.002545 | 0.968511 | -0.172024 | 0.007192 | -0.225616 | 0.000393 | -0.017775 | 0.782797 | -0.095103 | 0.139350 |
| tc_0111101110110 | | | 2 | | | -0.002703 | 0.966562 | 0.060262 | 0.349581 | -0.166798 | 0.009188 | -0.010684 | 0.868403 | 0.018183 | 0.777942 |
| tc_0111101110111 | 1 | | 1 | | | -0.141278 | 0.027665 | 0.008197 | 0.898845 | 0.145507 | 0.023291 | 0.038302 | 0.552368 | -0.038565 | 0.549641 |
| tc_0111101111100 | | | | | 1 | 0.040862 | 0.526106 | -0.055800 | 0.386480 | -0.051375 | 0.425300 | 0.061522 | 0.339579 | -0.159151 | 0.012992 |
| tc_01111011111100 | 1 | | 3 | | | -0.131254 | 0.040919 | 0.098291 | 0.126513 | -0.213892 | 0.000791 | 0.022680 | 0.725012 | -0.120927 | 0.059801 |
| tc_01111011111110 | | 1 | | 1 | | 0.076397 | 0.235425 | 0.145024 | 0.023759 | -0.071571 | 0.266416 | 0.142605 | 0.026223 | -0.040123 | 0.533622 |
| tc_011111011110 | 1 | | | | | 0.130508 | 0.042090 | 0.003898 | 0.951797 | -0.078474 | 0.222899 | 0.023486 | 0.715650 | 0.048708 | 0.449763 |
| tc_011111011111 | | 1 | | 1 | | -0.103318 | 0.108151 | -0.142399 | 0.026442 | -0.011991 | 0.852468 | -0.162491 | 0.011187 | -0.034050 | 0.597356 |
| tc_0111111000 | | | | | 1 | 0.024584 | 0.702969 | 0.009188 | 0.886696 | -0.100933 | 0.116582 | -0.012665 | 0.844278 | 0.127730 | 0.046701 |
| tc_01111110011 | 1 | | | | 3 | 0.149270 | 0.019915 | 0.061636 | 0.338683 | 0.003403 | 0.957914 | -0.021265 | 0.741539 | 0.204438 | 0.001354 |
| tc_0111111010110 | 1 | 2 | | | | -0.139856 | 0.029285 | -0.170364 | 0.007779 | -0.043798 | 0.496789 | -0.110117 | 0.086729 | 0.097877 | 0.128125 |
| tc_0111111010111 | | | 1 | 1 | | -0.021328 | 0.740805 | -0.026256 | 0.683821 | 0.161795 | 0.011544 | 0.126532 | 0.048817 | -0.044525 | 0.489665 |
| tc_01111110111 | | | 1 | | | 0.093249 | 0.147260 | -0.034720 | 0.590158 | -0.162800 | 0.011032 | -0.039005 | 0.545094 | -0.004679 | 0.942154 |
| tc_011111110 | | | 1 | | | 0.026244 | 0.683963 | -0.069786 | 0.278554 | 0.155480 | 0.015268 | -0.029850 | 0.643350 | 0.002698 | 0.966629 |
| tc_0111111111100 | | | | 1 | 3 | -0.072143 | 0.262605 | 0.094672 | 0.141160 | 0.049816 | 0.439500 | -0.138021 | 0.031496 | 0.246179 | 0.000105 |
| tc_0111111111110 | | | | 3 | | -0.100062 | 0.119785 | -0.031049 | 0.630072 | 0.008767 | 0.891853 | -0.199437 | 0.001782 | 0.118682 | 0.064738 |
| tc_01111111111111 | 1 | | 1 | | | 0.133232 | 0.037946 | -0.050187 | 0.436101 | 0.138381 | 0.031052 | 0.034062 | 0.597231 | 0.108912 | 0.090254 |
| tc_10001101 | | 1 | 2 | | | 0.009617 | 0.881435 | 0.146728 | 0.022146 | 0.174839 | 0.006286 | -0.077550 | 0.228411 | 0.025436 | 0.693196 |
| tc_10001111 | | 1 | | 2 | | -0.102205 | 0.112024 | 0.131817 | 0.040053 | 0.067539 | 0.294366 | -0.178340 | 0.005302 | 0.061132 | 0.342660 |
| tc_1001100 | | | 3 | | | 0.025628 | 0.690993 | -0.063828 | 0.321757 | -0.190769 | 0.002827 | -0.020693 | 0.748255 | 0.118587 | 0.064955 |
| tc_100110110 | | | | | 1 | -0.025511 | 0.692328 | 0.012089 | 0.851284 | 0.028718 | 0.655992 | 0.021436 | 0.739532 | 0.137509 | 0.032138 |
| tc_100110111 | 1 | 1 | | | | 0.136704 | 0.033169 | -0.159651 | 0.012707 | 0.040384 | 0.530966 | 0.016242 | 0.801118 | 0.120172 | 0.061426 |
| tc_1001110111 | 1 | 1 | | | 2 | 0.157987 | 0.013679 | 0.130756 | 0.041698 | -0.048805 | 0.448861 | 0.050240 | 0.435619 | -0.172430 | 0.007054 |
| tc_10011110 | | 1 | 3 | | 1 | -0.065391 | 0.310026 | 0.136836 | 0.032998 | -0.219688 | 0.000562 | -0.001676 | 0.979266 | -0.132086 | 0.039646 |
| tc_10110 | | | 1 | 1 | | 0.016352 | 0.799804 | 0.125287 | 0.051098 | 0.138281 | 0.031175 | -0.133422 | 0.037671 | 0.076200 | 0.236635 |
| tc_1011100 | | 1 | | | 1 | 0.100506 | 0.118143 | 0.146904 | 0.021984 | -0.007505 | 0.907338 | 0.091864 | 0.153391 | 0.141979 | 0.026894 |
| tc_1011101000 | | 2 | | 1 | | -0.043826 | 0.496515 | 0.173771 | 0.006617 | 0.112486 | 0.080120 | -0.128824 | 0.044837 | 0.055205 | 0.391567 |
| tc_1011101001 | | | 1 | | | -0.052817 | 0.412405 | 0.049860 | 0.439096 | -0.140619 | 0.028406 | -0.014738 | 0.819206 | 0.000751 | 0.990711 |
| tc_1011101010 | | | | | 1 | 0.028183 | 0.662005 | 0.054108 | 0.401066 | -0.084123 | 0.191244 | -0.028090 | 0.663048 | 0.159335 | 0.012887 |
| tc_1011101110 | | 2 | | | 3 | 0.124649 | 0.052302 | 0.173726 | 0.006631 | -0.123509 | 0.054512 | -0.023500 | 0.715489 | -0.183837 | 0.004034 |
| tc_101110111100 | | | | | 3 | 0.014989 | 0.816179 | -0.046812 | 0.467614 | -0.032246 | 0.616931 | 0.073147 | 0.256002 | -0.213021 | 0.000832 |
| tc_101110111110 | 1 | | | | | -0.164163 | 0.010369 | 0.039883 | 0.536074 | -0.084013 | 0.191831 | 0.010391 | 0.871980 | 0.057188 | 0.374758 |
| tc_10111011111100 | | | 3 | | 1 | -0.048958 | 0.447435 | 0.092356 | 0.151190 | -0.205871 | 0.001250 | -0.003167 | 0.960831 | 0.129720 | 0.043358 |
| tc_10111011111101 | | 1 | | | | -0.094640 | 0.141294 | 0.139349 | 0.029883 | 0.084509 | 0.189210 | -0.066186 | 0.304165 | 0.001334 | 0.983489 |
| tc_10111011111111 | | | 1 | | 1 | 0.013175 | 0.838094 | 0.124662 | 0.052278 | 0.128529 | 0.045333 | -0.060844 | 0.344944 | 0.146066 | 0.022761 |
| tc_10111110 | | 1 | 1 | | 4 | 0.050322 | 0.434869 | 0.164142 | 0.010379 | 0.147147 | 0.021764 | -0.036894 | 0.567088 | 0.270121 | 0.000020 |
| tc_10111111 | | 3 | | | 2 | -0.081045 | 0.208059 | 0.200048 | 0.001724 | -0.089651 | 0.163585 | 0.004662 | 0.942364 | -0.172794 | 0.006933 |
| tc_11000 | | 1 | 4 | | 1 | -0.051391 | 0.425159 | 0.134319 | 0.036392 | 0.255206 | 0.000057 | -0.090601 | 0.159146 | 0.127332 | 0.047394 |
| tc_11001 | | | | 1 | | -0.028291 | 0.660789 | 0.041585 | 0.518809 | 0.037562 | 0.560081 | 0.132664 | 0.038781 | -0.035624 | 0.580513 |
| tc_110100 | | 3 | | | | 0.066940 | 0.298673 | 0.234670 | 0.000223 | 0.060039 | 0.351370 | 0.075202 | 0.242847 | -0.036620 | 0.569965 |
| tc_110110 | | | | 1 | | 0.033769 | 0.600383 | -0.085701 | 0.183020 | 0.049183 | 0.445351 | -0.141156 | 0.027800 | -0.019771 | 0.759112 |
| tc_11011100 | | | 1 | | | 0.113707 | 0.076875 | -0.058355 | 0.365069 | -0.159058 | 0.013046 | 0.024920 | 0.699107 | 0.060473 | 0.347895 |
| tc_11011101 | | | 1 | | | 0.012880 | 0.841677 | -0.088228 | 0.170400 | 0.150486 | 0.018918 | 0.025613 | 0.691163 | 0.021195 | 0.742357 |
| tc_11011110 | | 1 | | 1 | | -0.038003 | 0.555480 | 0.132123 | 0.039589 | 0.050449 | 0.433708 | -0.152268 | 0.017537 | 0.049567 | 0.441796 |
| tc_110111110 | | 2 | | 1 | 1 | 0.081404 | 0.206045 | 0.172424 | 0.007056 | -0.090883 | 0.157847 | 0.134016 | 0.036820 | -0.133921 | 0.036954 |
| tc_110111111 | 2 | | | | | -0.165948 | 0.009555 | 0.074393 | 0.247970 | 0.079754 | 0.215418 | -0.036462 | 0.571635 | -0.031624 | 0.623747 |
| tc_111000000 | | | | | 1 | -0.045179 | 0.483305 | -0.069415 | 0.281123 | -0.041431 | 0.520358 | -0.085116 | 0.186037 | 0.137769 | 0.031810 |
| tc_11100100 | | 1 | | | 1 | -0.012480 | 0.846525 | 0.134944 | 0.035523 | 0.068408 | 0.288178 | -0.055305 | 0.390711 | 0.126332 | 0.049176 |
| tc_1110010101 | 1 | | | | | -0.140777 | 0.028226 | 0.024699 | 0.701651 | -0.122967 | 0.055587 | 0.070033 | 0.276857 | -0.012306 | 0.848644 |
| tc_11100101100 | | | | 1 | | -0.038540 | 0.549900 | 0.085265 | 0.185270 | -0.059653 | 0.354483 | -0.152871 | 0.017090 | -0.016286 | 0.800589 |
| tc_11100101101 | 3 | | | | | 0.187994 | 0.003264 | -0.074249 | 0.248890 | -0.000140 | 0.998263 | 0.078871 | 0.220556 | 0.068584 | 0.286942 |
| tc_11100101110 | 1 | | | | | -0.128994 | 0.044554 | 0.026800 | 0.677634 | -0.023964 | 0.710125 | -0.110221 | 0.086429 | 0.055508 | 0.388971 |
| tc_11100101111001 | 1 | | 3 | | | -0.164709 | 0.010114 | -0.079730 | 0.215555 | 0.246138 | 0.000106 | -0.046847 | 0.467282 | 0.013743 | 0.831224 |
| tc_11100101111010 | 3 | | | 1 | | -0.199143 | 0.001811 | -0.069480 | 0.280671 | 0.071832 | 0.264669 | -0.148295 | 0.020746 | 0.069785 | 0.278561 |
| tc_11100101111011 | | | 3 | | | 0.048056 | 0.455854 | -0.062138 | 0.334761 | -0.187850 | 0.003288 | 0.010185 | 0.874496 | 0.075488 | 0.241059 |
| tc_1110010111110 | | 1 | 2 | | | -0.036719 | 0.568925 | -0.161144 | 0.011886 | -0.177593 | 0.005499 | 0.055090 | 0.392561 | -0.008997 | 0.889027 |
| tc_1110010111111 | | 4 | 1 | | | 0.003068 | 0.962056 | -0.271623 | 0.000018 | -0.129847 | 0.043151 | 0.123516 | 0.054496 | -0.017323 | 0.788189 |
| tc_111001100010 | | 1 | | 1 | | -0.004936 | 0.938978 | 0.131677 | 0.040268 | -0.006104 | 0.924589 | -0.135459 | 0.034819 | 0.015285 | 0.812613 |
| tc_1110011001100 | | | 4 | | 1 | 0.037597 | 0.559717 | -0.012006 | 0.852292 | -0.283917 | 0.000007 | -0.051148 | 0.427352 | -0.151217 | 0.018340 |
| tc_1110011001110 | | 2 | | | 1 | -0.038129 | 0.554165 | -0.167367 | 0.008949 | -0.035674 | 0.579984 | -0.010041 | 0.876252 | 0.126145 | 0.049516 |
| tc_11100110100 | | | | 2 | | -0.102205 | 0.112024 | -0.065806 | 0.306957 | 0.067539 | 0.294366 | -0.178340 | 0.005302 | 0.061132 | 0.342660 |
| tc_111001101111 | | | 2 | | | -0.075596 | 0.240381 | -0.117443 | 0.067602 | 0.166988 | 0.009107 | -0.122069 | 0.057411 | 0.072199 | 0.262230 |
| tc_111010100100 | | 2 | | | 3 | -0.074404 | 0.247902 | -0.167535 | 0.008879 | 0.015184 | 0.813834 | 0.065486 | 0.309318 | -0.192558 | 0.002574 |
| tc_111010100101 | | 1 | 2 | | | -0.083602 | 0.194020 | -0.149235 | 0.019944 | 0.166863 | 0.009160 | 0.025476 | 0.692730 | -0.037756 | 0.558056 |
| tc_111010100110 | | 2 | | 1 | | 0.041853 | 0.516119 | -0.176501 | 0.005800 | -0.010218 | 0.874090 | -0.130054 | 0.042816 | 0.039884 | 0.536066 |
| tc_11101010011100 | | 3 | 1 | | 3 | -0.086260 | 0.180170 | -0.182261 | 0.004366 | -0.132814 | 0.038559 | 0.049568 | 0.441790 | -0.192248 | 0.002617 |
| tc_111010100111010 | 1 | 1 | | 2 | | 0.127735 | 0.046693 | -0.142380 | 0.026463 | -0.089950 | 0.162176 | 0.169150 | 0.008235 | -0.086938 | 0.176759 |
| tc_111010100111011 | | 2 | 1 | | | -0.051059 | 0.428160 | -0.177275 | 0.005585 | -0.140468 | 0.028579 | -0.012640 | 0.844582 | 0.098934 | 0.124037 |
| tc_11101010011110 | | 1 | | | 1 | -0.013957 | 0.828633 | -0.144838 | 0.023941 | 0.088803 | 0.167618 | -0.045403 | 0.481132 | 0.134644 | 0.035938 |
| tc_11101010011111 | | 3 | 1 | | | -0.108849 | 0.090441 | -0.186796 | 0.003471 | 0.132715 | 0.038705 | 0.031990 | 0.619735 | 0.070648 | 0.272647 |
| tc_1110101010010 | | 3 | | | 1 | -0.014954 | 0.816604 | -0.232559 | 0.000255 | -0.056777 | 0.378203 | 0.056269 | 0.382490 | -0.143768 | 0.025012 |
| tc_1110101010100 | | | 1 | | | 0.013849 | 0.829940 | -0.093052 | 0.148120 | 0.145181 | 0.023606 | 0.008276 | 0.897871 | 0.037134 | 0.564559 |
| tc_1110101010110 | | 3 | | 1 | 3 | 0.069935 | 0.277529 | -0.246806 | 0.000101 | 0.118863 | 0.064328 | 0.137776 | 0.031802 | -0.231023 | 0.000281 |
| tc_11101010101110 | | | 1 | | 2 | 0.088007 | 0.171476 | -0.090304 | 0.160523 | -0.128511 | 0.045364 | -0.040682 | 0.527933 | 0.173230 | 0.006790 |
| tc_111010101011110 | | 3 | 1 | | | 0.094987 | 0.139836 | -0.206393 | 0.001214 | -0.156676 | 0.014491 | 0.004514 | 0.944189 | 0.042801 | 0.506652 |
| tc_11101010111001 | 1 | | | 2 | | -0.163516 | 0.010679 | -0.038940 | 0.545769 | 0.065594 | 0.308523 | -0.177779 | 0.005450 | -0.057443 | 0.372626 |
| tc_111010101110101 | | 3 | 3 | | 3 | 0.029908 | 0.642707 | -0.179849 | 0.004922 | -0.217005 | 0.000659 | -0.002512 | 0.968923 | -0.203102 | 0.001458 |
| tc_111010101110111 | | 1 | | | 3 | 0.030517 | 0.635954 | -0.134465 | 0.036187 | -0.090741 | 0.158502 | 0.116423 | 0.070037 | -0.186118 | 0.003593 |
| tc_1110101011110 | | | | | 3 | -0.002250 | 0.972167 | -0.032198 | 0.617454 | 0.118678 | 0.064746 | -0.097156 | 0.130971 | 0.225348 | 0.000400 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tc_1110101011111 | 1 | 3 | 1 | | 3 | 0.144847 | 0.023932 | -0.230012 | 0.000300 | -0.154415 | 0.015990 | 0.069776 | 0.278624 | -0.191366 | 0.002740 |
| tc_11101011001011 | 1 | | | | | -0.132990 | 0.038299 | 0.021007 | 0.744569 | -0.046008 | 0.475304 | 0.101435 | 0.114764 | -0.012351 | 0.848097 |
| tc_111010110101 | | 3 | | | 3 | 0.004145 | 0.948749 | 0.180853 | 0.004683 | 0.043047 | 0.504210 | -0.046917 | 0.466617 | 0.208998 | 0.001048 |
| tc_11101011011100 | | | | 2 | 3 | 0.084988 | 0.186705 | 0.038269 | 0.552715 | -0.051179 | 0.427074 | 0.166857 | 0.009162 | 0.186709 | 0.003487 |
| tc_11101011011101 | | 1 | | | 2 | 0.047369 | 0.462334 | 0.134676 | 0.035893 | -0.036073 | 0.575751 | 0.015579 | 0.809081 | -0.167350 | 0.008956 |
| tc_111010110111110 | 1 | | | 1 | | 0.164829 | 0.010059 | -0.007621 | 0.905922 | -0.049831 | 0.439368 | -0.137243 | 0.032476 | 0.046273 | 0.472764 |
| tc_1110101101111110 | | | 1 | | | 0.032438 | 0.614839 | 0.013307 | 0.836492 | 0.157620 | 0.013902 | 0.068486 | 0.287628 | 0.080206 | 0.212820 |
| tc_111010111001100 | 1 | 1 | | | | -0.135264 | 0.035085 | 0.135205 | 0.035165 | 0.004755 | 0.941211 | -0.065853 | 0.306607 | -0.040778 | 0.526962 |
| tc_111010111001101 | | | | | 3 | 0.036686 | 0.569275 | 0.094003 | 0.144002 | 0.121597 | 0.058390 | 0.025956 | 0.687240 | -0.215964 | 0.000701 |
| tc_111010111001111 | | 1 | 3 | | | -0.047181 | 0.464115 | 0.147092 | 0.021813 | -0.224455 | 0.000422 | -0.043678 | 0.497967 | 0.055270 | 0.391010 |
| tc_1110101111110 | | 1 | 4 | | | -0.000939 | 0.988381 | 0.137885 | 0.031666 | -0.252814 | 0.000067 | 0.014534 | 0.821663 | -0.048309 | 0.453488 |
| tc_111010111111100 | | | 1 | | 3 | 0.001089 | 0.986530 | 0.120723 | 0.060235 | -0.141737 | 0.027158 | -0.112832 | 0.079187 | -0.195722 | 0.002177 |
| tc_1110101111111010 | | | | | 1 | -0.067515 | 0.294536 | 0.084793 | 0.187723 | -0.037591 | 0.559781 | -0.074720 | 0.245891 | 0.132794 | 0.038589 |
| tc_1110101111111111 | | 3 | 1 | | | -0.071144 | 0.269283 | 0.180925 | 0.004667 | -0.160301 | 0.012344 | -0.072705 | 0.258896 | -0.108525 | 0.091410 |
| tc_111100000111 | | 1 | | 3 | | 0.065462 | 0.309499 | 0.164771 | 0.010085 | 0.020519 | 0.750305 | 0.184619 | 0.003878 | 0.034145 | 0.596330 |
| tc_11110001000 | | 1 | 3 | | | -0.003719 | 0.954010 | 0.147796 | 0.021183 | -0.189627 | 0.003000 | -0.053397 | 0.407291 | -0.073098 | 0.256318 |
| tc_1111000111010 | | | 3 | | | -0.025545 | 0.691949 | -0.040377 | 0.531038 | -0.240599 | 0.000152 | 0.046181 | 0.473645 | -0.121105 | 0.059424 |
| tc_1111000111101 | 3 | | 1 | | | 0.211997 | 0.000882 | -0.050422 | 0.433957 | -0.136003 | 0.034090 | 0.082308 | 0.201032 | -0.118056 | 0.066173 |
| tc_11110010000 | | | 1 | | 1 | -0.003708 | 0.954139 | -0.056407 | 0.381324 | -0.132582 | 0.038902 | 0.121915 | 0.057729 | -0.137274 | 0.032436 |
| tc_111100100101 | | | 3 | | | -0.027033 | 0.674992 | 0.031266 | 0.627680 | -0.181882 | 0.004449 | 0.059013 | 0.359682 | -0.086790 | 0.177502 |
| tc_1111001001100 | | | | 4 | | -0.091793 | 0.153709 | 0.066850 | 0.299329 | 0.040034 | 0.534539 | -0.259553 | 0.000042 | 0.097946 | 0.127853 |
| tc_1111001001110 | | | | | 4 | 0.046424 | 0.471322 | 0.110901 | 0.084493 | 0.099377 | 0.122353 | 0.038304 | 0.552351 | 0.261200 | 0.000037 |
| tc_111100101010 | 1 | | | | | -0.158354 | 0.013459 | 0.039413 | 0.540898 | 0.037968 | 0.555847 | -0.060589 | 0.346974 | 0.026896 | 0.676542 |
| tc_111100101011 | 1 | | | | 1 | -0.128646 | 0.045136 | 0.030726 | 0.633633 | -0.077170 | 0.230706 | 0.021383 | 0.740157 | -0.161221 | 0.011846 |
| tc_1111001011000 | | | 4 | | | 0.019568 | 0.761516 | -0.063770 | 0.322192 | -0.257383 | 0.000049 | 0.000502 | 0.993786 | 0.091055 | 0.157060 |
| tc_11110010110011 | | 1 | | | 1 | -0.000620 | 0.992331 | 0.125956 | 0.049862 | 0.077809 | 0.226855 | -0.114519 | 0.074775 | 0.126568 | 0.048752 |
| tc_11110010110101 | | 2 | | 1 | | 0.067405 | 0.295326 | -0.169944 | 0.007934 | -0.085915 | 0.181927 | -0.153026 | 0.016976 | 0.123836 | 0.053870 |
| tc_11110010111010 | | | | 3 | | -0.034308 | 0.594581 | -0.057550 | 0.371734 | 0.039556 | 0.539430 | -0.181375 | 0.004563 | 0.056597 | 0.379718 |
| tc_1111001011110 | | | | 1 | | 0.049754 | 0.440077 | -0.100442 | 0.118380 | 0.008190 | 0.898936 | 0.163988 | 0.010452 | 0.072705 | 0.258892 |
| tc_11110010111111 | | 1 | | | | 0.119377 | 0.063175 | -0.157738 | 0.013830 | 0.107994 | 0.093016 | 0.034342 | 0.594215 | 0.019462 | 0.762772 |
| tc_111100110000 | | | | | 3 | 0.003450 | 0.957330 | -0.025482 | 0.692668 | 0.099921 | 0.120310 | -0.052237 | 0.417570 | -0.190349 | 0.002890 |
| tc_111100110001 | | | 1 | | | -0.045192 | 0.483180 | -0.011098 | 0.863344 | 0.163058 | 0.010903 | -0.071702 | 0.265538 | 0.069194 | 0.282665 |
| tc_1111001100100 | | | | | 2 | 0.107680 | 0.093976 | -0.045204 | 0.483063 | -0.054281 | 0.399557 | 0.046637 | 0.469281 | 0.175569 | 0.006068 |
| tc_1111001100101 | | | 1 | | 1 | 0.101184 | 0.115669 | 0.121559 | 0.058468 | -0.129873 | 0.043109 | 0.035781 | 0.578849 | -0.128384 | 0.045579 |
| tc_1111001100110 | | 1 | | | 1 | -0.060381 | 0.348633 | -0.151619 | 0.018030 | 0.099234 | 0.122896 | 0.070240 | 0.275433 | -0.163269 | 0.010800 |
| tc_111100110100 | | 3 | | | | 0.080941 | 0.208646 | 0.193150 | 0.002495 | -0.125488 | 0.050725 | 0.054176 | 0.400468 | 0.036617 | 0.570004 |
| tc_1111001101010 | | 1 | 3 | | | 0.104365 | 0.104604 | 0.147934 | 0.021062 | -0.233707 | 0.000238 | 0.112705 | 0.079528 | -0.112279 | 0.080678 |
| tc_1111001101011 | 3 | | 1 | | | -0.191965 | 0.002656 | 0.030315 | 0.638188 | 0.141689 | 0.027211 | 0.043115 | 0.503535 | 0.015928 | 0.804889 |
| tc_11110011011010 | | 2 | 3 | | | -0.005670 | 0.929930 | -0.175241 | 0.006165 | 0.191401 | 0.002735 | -0.041563 | 0.519028 | 0.006065 | 0.925058 |
| tc_111100110111011 | | 3 | | | 3 | 0.088176 | 0.170653 | 0.182330 | 0.004351 | 0.085561 | 0.183742 | -0.016289 | 0.800558 | 0.205979 | 0.001242 |
| tc_111100110111101 | | | | 1 | | -0.023783 | 0.712216 | -0.027439 | 0.670393 | 0.006371 | 0.921291 | -0.136007 | 0.034085 | 0.081440 | 0.205844 |
| tc_111100111010 | 1 | 1 | | | | -0.128907 | 0.044698 | 0.162713 | 0.011075 | 0.047253 | 0.463424 | 0.070291 | 0.275086 | 0.070336 | 0.274778 |
| tc_1111001110110 | | | 1 | | | -0.080533 | 0.210953 | -0.023085 | 0.720301 | 0.141875 | 0.027008 | -0.119947 | 0.061917 | 0.016725 | 0.795328 |
| tc_11110011101110 | | | | 3 | 1 | -0.122525 | 0.056479 | 0.022250 | 0.730021 | 0.069614 | 0.279743 | -0.237492 | 0.000186 | 0.128814 | 0.044854 |
| tc_11110011101111 | | | | 1 | | 0.069613 | 0.279755 | 0.105115 | 0.102120 | 0.083959 | 0.192115 | 0.139857 | 0.029284 | -0.047021 | 0.465628 |
| tc_1111001111000 | | | | | 1 | 0.032291 | 0.616444 | -0.005963 | 0.926326 | 0.025803 | 0.688994 | -0.071642 | 0.265941 | 0.156675 | 0.014492 |
| tc_11110011110010 | | | 2 | | 1 | -0.007605 | 0.906112 | 0.100430 | 0.118422 | 0.172334 | 0.007087 | -0.006972 | 0.913901 | 0.158505 | 0.013370 |
| tc_111100111100110 | | | | | 1 | -0.028684 | 0.656379 | -0.075949 | 0.238188 | 0.011321 | 0.860631 | 0.007981 | 0.901491 | 0.132369 | 0.039220 |
| tc_11110011110100 | | 1 | | | | -0.060003 | 0.351659 | 0.130558 | 0.042010 | -0.046635 | 0.469306 | -0.007535 | 0.906972 | -0.041949 | 0.515149 |
| tc_11110011110101 | | | | 3 | | 0.025744 | 0.689670 | -0.004532 | 0.943973 | 0.084244 | 0.190605 | -0.184711 | 0.003859 | 0.076319 | 0.235903 |
| tc_1111001111100 | | 1 | | 1 | | -0.088870 | 0.167300 | 0.145406 | 0.023388 | 0.025868 | 0.688255 | -0.148424 | 0.020634 | 0.067663 | 0.293472 |
| tc_1111001111101100 | | | | | 2 | 0.060943 | 0.344152 | 0.000530 | 0.993438 | -0.050386 | 0.434283 | 0.089654 | 0.163567 | -0.169420 | 0.008132 |
| tc_1111001111101101 | | | | | 1 | 0.072369 | 0.261109 | -0.002479 | 0.969328 | -0.032311 | 0.616224 | -0.010491 | 0.870755 | -0.137608 | 0.032013 |
| tc_1111010000 | | 1 | 2 | | | 0.051232 | 0.426596 | 0.134269 | 0.036462 | 0.165285 | 0.009851 | -0.061364 | 0.340824 | -0.010161 | 0.874785 |
| tc_11110100011 | | 1 | | | | -0.032771 | 0.611203 | -0.136972 | 0.032823 | 0.073869 | 0.251323 | 0.034159 | 0.596181 | 0.008598 | 0.893918 |
| tc_11110100110 | | | 1 | | | -0.102729 | 0.110185 | -0.018626 | 0.772674 | -0.127309 | 0.047435 | -0.024467 | 0.704327 | -0.045753 | 0.477755 |
| tc_111101001111 | 1 | | | 3 | 1 | 0.157518 | 0.013965 | 0.009412 | 0.883944 | -0.027404 | 0.670787 | 0.185832 | 0.003646 | -0.135664 | 0.034544 |
| tc_1111010101000 | | | 1 | | | -0.009352 | 0.884681 | -0.028911 | 0.653833 | 0.151734 | 0.017941 | -0.032276 | 0.616600 | 0.085509 | 0.184007 |
| tc_1111010101001 | | | 1 | | | 0.031801 | 0.621799 | -0.033051 | 0.608167 | 0.128700 | 0.045045 | -0.043725 | 0.497506 | 0.078147 | 0.224835 |
| tc_1111010101010 | | 2 | | 1 | | -0.029976 | 0.641945 | 0.173859 | 0.006589 | -0.028107 | 0.662859 | -0.143797 | 0.024982 | 0.007806 | 0.903645 |
| tc_1111010101011 | | 1 | | | 3 | -0.010833 | 0.866584 | 0.156115 | 0.014850 | 0.044842 | 0.486573 | -0.100055 | 0.119812 | 0.187250 | 0.003391 |
| tc_1111010101100 | | | 3 | | | -0.019008 | 0.768147 | 0.111452 | 0.082952 | 0.224578 | 0.000419 | -0.011012 | 0.864399 | 0.029439 | 0.647933 |
| tc_1111010101110 | | 1 | | 1 | 3 | 0.015219 | 0.813415 | 0.149490 | 0.019731 | 0.100362 | 0.118675 | -0.152623 | 0.017273 | 0.201181 | 0.001620 |
| tc_11110101101001 | | | 1 | 1 | | -0.011375 | 0.859968 | 0.078187 | 0.224599 | -0.141577 | 0.027335 | -0.136260 | 0.033751 | 0.099111 | 0.123362 |
| tc_11110101101110 | | | | 1 | 3 | 0.079708 | 0.215685 | 0.078214 | 0.224436 | 0.089502 | 0.164287 | -0.159411 | 0.012843 | 0.244683 | 0.000117 |
| tc_111101011101 | 1 | | | | | -0.126379 | 0.049092 | 0.019521 | 0.762069 | -0.041425 | 0.520424 | -0.038948 | 0.545685 | 0.004697 | 0.941934 |
| tc_11110101111011 | 3 | | 1 | | 2 | 0.192491 | 0.002583 | -0.056854 | 0.377551 | -0.162524 | 0.011170 | 0.037700 | 0.558638 | -0.169532 | 0.008090 |
| tc_11110101111110 | | | | | 1 | -0.081495 | 0.205533 | 0.067859 | 0.292074 | 0.010218 | 0.874090 | -0.022181 | 0.730820 | 0.131275 | 0.040886 |
| tc_111101100010 | | 1 | | | | 0.050709 | 0.431333 | -0.132436 | 0.039120 | 0.081748 | 0.204129 | 0.000398 | 0.995076 | 0.007079 | 0.912583 |
| tc_1111011001111 | | | 4 | | | -0.089780 | 0.162975 | 0.032235 | 0.617054 | -0.301456 | 0.000002 | -0.008212 | 0.898665 | -0.074189 | 0.249276 |
| tc_1111011010 | 1 | | | | 1 | 0.143331 | 0.025461 | 0.124590 | 0.052415 | 0.031747 | 0.622400 | 0.112512 | 0.080047 | 0.150078 | 0.019248 |
| tc_111101101100 | | | | 2 | 1 | -0.098304 | 0.126462 | 0.035539 | 0.581421 | 0.024498 | 0.703968 | -0.166537 | 0.009299 | 0.140780 | 0.028224 |
| tc_1111011011110 | | | | | 1 | 0.026748 | 0.678223 | -0.028614 | 0.657158 | 0.023564 | 0.714750 | 0.103088 | 0.108941 | 0.152533 | 0.017339 |
| tc_11110111000101 | | 1 | | | | 0.056140 | 0.383582 | -0.149070 | 0.020083 | 0.101560 | 0.114317 | -0.000317 | 0.996073 | -0.007061 | 0.912808 |
| tc_11110111001000 | | | | | 1 | -0.007916 | 0.902296 | 0.085253 | 0.185331 | 0.114702 | 0.074309 | -0.055321 | 0.390571 | 0.136885 | 0.032936 |
| tc_11110111001011 | | | 3 | | 3 | -0.022404 | 0.728225 | -0.085185 | 0.185679 | -0.242666 | 0.000133 | -0.000119 | 0.998532 | 0.221139 | 0.000516 |
| tc_11110111001100 | | 1 | | 3 | 1 | -0.084361 | 0.189989 | 0.144328 | 0.024447 | 0.059478 | 0.355897 | -0.221059 | 0.000518 | 0.136259 | 0.033752 |
| tc_111101110011010 | 1 | | | | | -0.126139 | 0.049526 | -0.082619 | 0.199334 | 0.021509 | 0.738674 | 0.030216 | 0.639280 | -0.007359 | 0.909133 |
| tc_11110111010010 | | | | 1 | 3 | -0.043346 | 0.501242 | 0.090726 | 0.158569 | 0.054499 | 0.397666 | -0.162982 | 0.010941 | 0.192962 | 0.002520 |
| tc_111101110100110 | | | | | 1 | 0.058133 | 0.366902 | 0.064267 | 0.318430 | 0.112293 | 0.080641 | -0.056386 | 0.381505 | 0.151755 | 0.017926 |
| tc_11110111010101 | | | | | 1 | -0.061128 | 0.342688 | -0.009937 | 0.877521 | -0.035874 | 0.577858 | -0.077528 | 0.228543 | 0.140813 | 0.028186 |
| tc_111101110101110 | | | | | 2 | -0.025396 | 0.693646 | 0.107704 | 0.093902 | -0.098631 | 0.125196 | -0.125300 | 0.051074 | -0.168660 | 0.008426 |
| tc_11110111011001 | | | | | 1 | 0.085731 | 0.182868 | 0.002181 | 0.973016 | 0.034776 | 0.589557 | -0.037810 | 0.557490 | 0.139506 | 0.029697 |
| tc_111101110110101 | | | | | 3 | 0.077208 | 0.230476 | 0.056619 | 0.379535 | 0.036611 | 0.570060 | -0.002770 | 0.965737 | -0.203413 | 0.001433 |
| tc_111101110111000 | 1 | 1 | 4 | | | 0.150241 | 0.019116 | 0.138251 | 0.031211 | -0.353052 | 0.000000 | 0.031579 | 0.624239 | -0.111762 | 0.082093 |
| tc_1111011101110010 | | | | 1 | | 0.006295 | 0.922235 | 0.097206 | 0.130772 | 0.119002 | 0.064015 | -0.158134 | 0.013591 | 0.038733 | 0.547908 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tc_111101110111011 | 1 | | 2 | | 3 | 0.157718 | 0.013842 | -0.068720 | 0.285979 | -0.167035 | 0.009087 | 0.047969 | 0.456676 | -0.225217 | 0.000403 |
| tc_11110111011110 | | | | | 1 | 0.042395 | 0.510690 | -0.002938 | 0.963653 | -0.087161 | 0.175646 | -0.028067 | 0.663308 | -0.154536 | 0.015906 |
| tc_111101110111110 | | 2 | | 1 | 1 | -0.019477 | 0.762589 | 0.178701 | 0.005209 | -0.054130 | 0.400869 | -0.129852 | 0.043142 | 0.146351 | 0.022494 |
| tc_11110111100110 | | 1 | | | | 0.002730 | 0.966229 | 0.152129 | 0.017642 | 0.087679 | 0.173084 | -0.109758 | 0.087765 | 0.030949 | 0.631171 |
| tc_11110111001110 | | 1 | 1 | 1 | 1 | -0.027781 | 0.666529 | 0.155920 | 0.014978 | -0.131454 | 0.040611 | -0.131233 | 0.040952 | 0.127651 | 0.046839 |
| tc_11110111011010 | | | 3 | | | 0.106895 | 0.096412 | 0.031202 | 0.628384 | -0.194364 | 0.002340 | -0.022118 | 0.731558 | 0.012672 | 0.844197 |
| tc_11110111011011 | | | 3 | | 1 | -0.065704 | 0.307706 | -0.017574 | 0.785189 | 0.198636 | 0.001861 | -0.002444 | 0.969766 | 0.145804 | 0.023008 |
| tc_111101110111001 | 1 | | 3 | | | 0.147615 | 0.021344 | -0.019074 | 0.767367 | -0.234729 | 0.000223 | 0.037533 | 0.560388 | -0.123684 | 0.054167 |
| tc_111101110111010 | | | | 3 | 3 | -0.080302 | 0.212270 | 0.029010 | 0.652725 | 0.057509 | 0.372076 | -0.216134 | 0.000694 | 0.202047 | 0.001545 |
| tc_111101110111101 | | | 1 | | | 0.112653 | 0.079668 | 0.034694 | 0.590441 | 0.134249 | 0.036491 | -0.068480 | 0.287672 | 0.124440 | 0.052702 |
| tc_111101110111110 | | | 1 | | 1 | 0.037701 | 0.558635 | 0.039320 | 0.541849 | -0.134979 | 0.035474 | 0.058890 | 0.360686 | -0.130279 | 0.042454 |
| tc_11110111100010 | | 1 | | | 3 | 0.084175 | 0.190969 | 0.130448 | 0.042186 | 0.003900 | 0.951768 | -0.081552 | 0.205215 | 0.215377 | 0.000725 |
| tc_11110111100101 | | 1 | 1 | | | -0.022819 | 0.723390 | 0.137868 | 0.031687 | 0.132913 | 0.038413 | -0.109582 | 0.088279 | 0.092890 | 0.148831 |
| tc_11110111100110 | | | | | 3 | -0.049385 | 0.443478 | 0.069816 | 0.278350 | 0.004297 | 0.946867 | -0.022636 | 0.725519 | 0.215104 | 0.000737 |
| tc_11110111100111 | | 3 | | | 1 | -0.019521 | 0.762067 | 0.186204 | 0.003578 | 0.099114 | 0.123348 | -0.120879 | 0.059904 | 0.133622 | 0.037383 |
| tc_111101111010110 | | | 1 | | | -0.009308 | 0.885228 | 0.090093 | 0.161507 | 0.131819 | 0.040050 | 0.057970 | 0.368244 | -0.099795 | 0.120780 |
| tc_111101111010111 | | 2 | | | 1 | 0.035466 | 0.582193 | 0.172516 | 0.007026 | 0.067766 | 0.292740 | 0.013403 | 0.835333 | 0.158206 | 0.013547 |
| tc_11110111101110 | | 1 | | | | 0.003256 | 0.959724 | -0.137368 | 0.032316 | -0.022142 | 0.731282 | -0.076816 | 0.232856 | 0.046952 | 0.466288 |
| tc_1111011111000 | | | | 1 | | -0.016616 | 0.796634 | 0.001218 | 0.984936 | 0.021977 | 0.733208 | 0.150286 | 0.019080 | 0.035202 | 0.585003 |
| tc_11110111110100 | | | | 1 | | -0.001339 | 0.983431 | 0.051028 | 0.428439 | 0.022692 | 0.724874 | 0.135004 | 0.035440 | -0.023671 | 0.713515 |
| tc_11110111110101 | | 1 | 3 | | | -0.055897 | 0.385653 | 0.127319 | 0.047417 | -0.184196 | 0.003961 | -0.085855 | 0.182235 | -0.032466 | 0.614530 |
| tc_11111000000 | | | | | 1 | -0.009783 | 0.879404 | 0.072365 | 0.261131 | 0.115839 | 0.071464 | -0.057837 | 0.369348 | 0.128481 | 0.045414 |
| tc_111110001010 | | 2 | 1 | | | -0.080878 | 0.208997 | 0.178764 | 0.005193 | -0.155279 | 0.015402 | 0.039445 | 0.540564 | 0.024319 | 0.706024 |
| tc_111110001011 | | | 1 | | | 0.005440 | 0.932765 | -0.048588 | 0.450878 | -0.128803 | 0.044872 | -0.098464 | 0.125841 | 0.116846 | 0.069020 |
| tc_111110001100 | | 1 | | | 1 | 0.081965 | 0.202925 | 0.129056 | 0.044451 | -0.008599 | 0.893909 | 0.061852 | 0.336990 | -0.125983 | 0.049812 |
| tc_111110001101 | | | 1 | | | -0.046180 | 0.473656 | 0.083433 | 0.194924 | 0.132820 | 0.038550 | -0.108470 | 0.091576 | 0.073014 | 0.256871 |
| tc_1111100011100 | | | | | 3 | 0.012557 | 0.845589 | -0.053080 | 0.410079 | 0.001310 | 0.983796 | -0.088772 | 0.167770 | 0.223247 | 0.000454 |
| tc_1111100011110 | | | | | 1 | 0.040121 | 0.533643 | -0.068823 | 0.285256 | 0.104520 | 0.104086 | -0.091831 | 0.153540 | 0.145429 | 0.023366 |
| tc_11111000111110 | | 1 | 1 | 2 | | -0.050144 | 0.436493 | 0.140377 | 0.028682 | -0.140986 | 0.027991 | -0.175997 | 0.005943 | 0.100395 | 0.118550 |
| tc_11111000111111 | 1 | | 1 | | | -0.144804 | 0.023975 | 0.095931 | 0.135924 | -0.129329 | 0.043997 | 0.033159 | 0.606994 | 0.001196 | 0.985197 |
| tc_11111001001010 | 1 | 1 | | | | -0.140301 | 0.028770 | -0.135135 | 0.035261 | -0.004376 | 0.945892 | -0.018027 | 0.779793 | -0.020142 | 0.754736 |
| tc_11111001001011 | | | 1 | | | 0.018188 | 0.777881 | -0.081690 | 0.204447 | 0.126240 | 0.049343 | 0.031262 | 0.627723 | -0.018704 | 0.771747 |
| tc_1111100100110 | | 1 | | | 1 | -0.056328 | 0.381991 | -0.164362 | 0.010275 | 0.045157 | 0.483522 | 0.057625 | 0.371108 | 0.139983 | 0.029138 |
| tc_11111001001111 | | | | | 3 | -0.011758 | 0.855304 | -0.081936 | 0.203085 | -0.105771 | 0.099986 | 0.102204 | 0.112027 | 0.208530 | 0.001076 |
| tc_11111001010 | 2 | | 2 | | | -0.175204 | 0.006176 | -0.063882 | 0.321341 | -0.177665 | 0.005480 | 0.022952 | 0.721849 | 0.017243 | 0.789138 |
| tc_111110010110 | | | | 2 | | -0.122174 | 0.057196 | -0.062394 | 0.332770 | -0.048321 | 0.453374 | -0.172625 | 0.006989 | 0.061797 | 0.337425 |
| tc_11111001011110 | | 1 | 3 | | | -0.080626 | 0.210426 | -0.136851 | 0.032979 | -0.205776 | 0.001256 | -0.080713 | 0.209932 | -0.099598 | 0.121519 |
| tc_11111001011111 | | | | | 3 | -0.071005 | 0.270228 | -0.071268 | 0.268453 | -0.087011 | 0.176396 | -0.030033 | 0.641319 | -0.214387 | 0.000768 |
| tc_111110011110 | | 1 | | | 3 | 0.026419 | 0.681968 | -0.130795 | 0.041635 | 0.031571 | 0.624331 | -0.065934 | 0.306009 | 0.185598 | 0.003690 |
| tc_1111100111111 | | 1 | | 1 | | 0.060539 | 0.347370 | 0.150895 | 0.018594 | -0.047597 | 0.460179 | 0.148021 | 0.020985 | -0.016717 | 0.795431 |
| tc_1111101001 | | | 3 | | | 0.037527 | 0.560444 | 0.105598 | 0.100544 | -0.241782 | 0.000141 | -0.012734 | 0.843450 | -0.049603 | 0.441464 |
| tc_1111101010 | | | | | 2 | 0.003836 | 0.952559 | -0.022367 | 0.728657 | -0.117553 | 0.067345 | -0.004234 | 0.947646 | -0.166544 | 0.009296 |
| tc_1111101101 | 3 | | | | | 0.210115 | 0.000983 | 0.019143 | 0.766541 | 0.002703 | 0.966563 | 0.036004 | 0.576482 | 0.071026 | 0.270081 |
| tc_1111101111001 | 1 | | | | | 0.156363 | 0.014690 | 0.067954 | 0.291400 | 0.020368 | 0.752083 | -0.002446 | 0.969741 | -0.011410 | 0.859542 |
| tc_1111101111100 | | | | 2 | | -0.004189 | 0.948204 | 0.007850 | 0.903105 | 0.101355 | 0.115052 | 0.178486 | 0.005264 | -0.044118 | 0.493650 |
| tc_11111011111110 | | 1 | | | 1 | 0.028213 | 0.661661 | 0.150329 | 0.019045 | -0.037355 | 0.562250 | -0.008912 | 0.890069 | 0.148873 | 0.020250 |
| tc_111111000010 | | | | 1 | | -0.006952 | 0.914138 | -0.005254 | 0.935064 | 0.075077 | 0.243639 | 0.152268 | 0.017537 | -0.040426 | 0.530540 |
| tc_1111110001010 | | 1 | | | 1 | -0.033922 | 0.598732 | 0.131630 | 0.040339 | 0.063034 | 0.327826 | -0.054218 | 0.400105 | 0.149855 | 0.019430 |
| tc_11111100010110 | | | | | 3 | 0.004344 | 0.946284 | 0.046379 | 0.471748 | 0.085983 | 0.181580 | -0.070050 | 0.276739 | 0.197833 | 0.001944 |
| tc_11111100010111 | | | | | 1 | 0.005368 | 0.933652 | 0.094262 | 0.142896 | 0.040722 | 0.527531 | 0.034852 | 0.588747 | -0.138935 | 0.030378 |
| tc_1111110001100 | | 1 | 1 | | | 0.007365 | 0.909064 | 0.149257 | 0.019926 | 0.161035 | 0.011945 | -0.062568 | 0.331419 | 0.058892 | 0.360667 |
| tc_11111100011011 | 1 | | 1 | | | -0.137419 | 0.032252 | -0.035367 | 0.583245 | 0.161966 | 0.011455 | -0.044026 | 0.494551 | -0.060410 | 0.348395 |
| tc_111111001000 | | 1 | | | | -0.065067 | 0.312427 | 0.133167 | 0.038041 | 0.053532 | 0.406105 | -0.065018 | 0.312799 | -0.070252 | 0.275349 |
| tc_1111110010010 | | | | 1 | | -0.091201 | 0.156393 | 0.068791 | 0.285484 | 0.082677 | 0.199014 | -0.148176 | 0.020850 | 0.118554 | 0.065030 |
| tc_1111110010011 | | | | | 2 | 0.120355 | 0.061028 | -0.052623 | 0.414130 | 0.049682 | 0.440739 | 0.002552 | 0.968435 | 0.170114 | 0.007872 |
| tc_1111110010110 | 1 | 2 | | 1 | | -0.162623 | 0.011120 | 0.169765 | 0.008001 | 0.125368 | 0.050947 | -0.162644 | 0.011110 | 0.105045 | 0.102349 |
| tc_1111110010111 | | | | 1 | | -0.033257 | 0.605925 | -0.013888 | 0.829460 | 0.047264 | 0.463327 | -0.158747 | 0.013227 | 0.109412 | 0.088778 |
| tc_1111110011000 | | 3 | 1 | | | -0.053777 | 0.403957 | 0.201201 | 0.001619 | 0.160729 | 0.012110 | -0.070496 | 0.273683 | 0.037821 | 0.557383 |
| tc_1111110011001 | | 3 | | | | 0.032593 | 0.613147 | 0.180894 | 0.004674 | 0.092083 | 0.152409 | -0.072020 | 0.263417 | -0.002150 | 0.973397 |
| tc_11111100110110 | | 1 | 1 | | | -0.022924 | 0.722172 | 0.136917 | 0.032894 | -0.127610 | 0.046909 | -0.053212 | 0.408913 | -0.017981 | 0.780342 |
| tc_1111110011101 | | | | | 1 | -0.014858 | 0.817750 | -0.109017 | 0.089943 | 0.018337 | 0.776102 | 0.035914 | 0.577433 | -0.131308 | 0.040835 |
| tc_11111100111110 | | | | | 1 | 0.026324 | 0.683046 | 0.016662 | 0.796091 | 0.076665 | 0.233778 | 0.017337 | 0.788025 | 0.140146 | 0.028949 |
| tc_111111010000 | | 1 | | | | -0.039407 | 0.540959 | -0.155947 | 0.014960 | 0.106564 | 0.097455 | -0.076688 | 0.233638 | 0.050132 | 0.436605 |
| tc_111111010001 | | | 2 | | | -0.035525 | 0.581567 | 0.040364 | 0.531164 | 0.168468 | 0.008502 | 0.041382 | 0.520851 | -0.007879 | 0.902746 |
| tc_111111010010 | | 1 | 1 | | | -0.001205 | 0.985086 | 0.128320 | 0.045689 | -0.135520 | 0.034738 | 0.041080 | 0.523902 | -0.058137 | 0.366865 |
| tc_1111110100110 | | | 1 | | | 0.025352 | 0.694160 | -0.067927 | 0.291593 | 0.129927 | 0.043021 | 0.092703 | 0.149653 | 0.103359 | 0.108010 |
| tc_111111010100 | 1 | | | | 1 | -0.140146 | 0.028949 | 0.072490 | 0.260308 | 0.066289 | 0.303413 | -0.056259 | 0.382572 | 0.131674 | 0.040272 |
| tc_1111110101110 | | 1 | 1 | | | -0.044191 | 0.492932 | 0.149273 | 0.019912 | 0.144543 | 0.024232 | 0.113145 | 0.078354 | 0.077680 | 0.227628 |
| tc_11111101011110 | | | 1 | | | 0.027300 | 0.671966 | 0.123167 | 0.055189 | 0.135841 | 0.034306 | 0.032510 | 0.614048 | -0.111210 | 0.083626 |
| tc_1111110111001 | | | 1 | | 3 | 0.068437 | 0.287978 | 0.027275 | 0.672247 | -0.130699 | 0.041787 | -0.058448 | 0.364302 | 0.240798 | 0.000151 |
| tc_111111011101011 | | | 1 | | | 0.043515 | 0.499577 | 0.079970 | 0.214176 | -0.148796 | 0.020315 | -0.040621 | 0.528557 | 0.031732 | 0.622563 |
| tc_11111101110110 | | | 4 | | 2 | 0.110795 | 0.084793 | 0.033495 | 0.603348 | -0.260600 | 0.000039 | 0.051522 | 0.423982 | -0.178958 | 0.005143 |
| tc_111111011011110 | | 1 | 1 | | | -0.016618 | 0.796609 | 0.147069 | 0.021834 | 0.142631 | 0.026194 | -0.032592 | 0.613151 | -0.008846 | 0.890879 |
| tc_1111110111110 | | | 1 | | | -0.047215 | 0.463784 | 0.002877 | 0.964415 | 0.126079 | 0.049637 | -0.033140 | 0.607193 | -0.053503 | 0.406359 |
| tc_1111111000 | | | 3 | | | -0.066886 | 0.299067 | 0.042447 | 0.510168 | -0.182943 | 0.004219 | -0.044732 | 0.487644 | -0.099956 | 0.120181 |
| tc_111111101100 | | 1 | | | 3 | 0.020561 | 0.749803 | 0.157068 | 0.014244 | 0.003590 | 0.955596 | 0.092956 | 0.148541 | 0.236114 | 0.000204 |
| tc_111111101101 | | 2 | 3 | | | -0.121956 | 0.057643 | 0.176182 | 0.005890 | 0.188792 | 0.003133 | 0.000237 | 0.997071 | 0.087659 | 0.173182 |
| tc_111111101110 | | | 1 | | 1 | 0.029985 | 0.641848 | 0.043557 | 0.499166 | 0.127413 | 0.047252 | -0.000400 | 0.995053 | 0.152641 | 0.017259 |
| tc_111111101111 | | | | 3 | | -0.112424 | 0.080285 | -0.009146 | 0.887205 | 0.092198 | 0.151894 | -0.198071 | 0.001919 | 0.119104 | 0.063786 |
| tc_111111110110 | | | | | 1 | 0.068872 | 0.284914 | -0.010742 | 0.867688 | -0.056773 | 0.378234 | -0.053979 | 0.402189 | 0.137692 | 0.031908 |
| tc_11111111011110 | | | 3 | | | -0.119022 | 0.063970 | 0.059391 | 0.356608 | 0.192020 | 0.002648 | 0.036939 | 0.566616 | -0.042775 | 0.506910 |
| tc_11111111011111 | | | | 2 | | -0.026874 | 0.676793 | 0.035763 | 0.579034 | 0.036567 | 0.570532 | -0.166854 | 0.009164 | -0.006141 | 0.924132 |
| tc_1111111110 | | 1 | | | 1 | -0.010536 | 0.870200 | -0.151675 | 0.017986 | 0.065339 | 0.310411 | -0.047560 | 0.460521 | -0.133119 | 0.038111 |
| tc_111111111100 | | | | | 1 | -0.034288 | 0.594792 | 0.030442 | 0.636783 | -0.100757 | 0.117222 | 0.062231 | 0.334031 | -0.140190 | 0.028897 |
| tc_111111111101 | | 1 | 3 | | | 0.061483 | 0.339889 | 0.148187 | 0.020840 | -0.194957 | 0.002267 | -0.050968 | 0.428980 | 0.061664 | 0.338467 |
| tc_111111111110 | | 1 | | | 1 | 0.011572 | 0.857572 | 0.148650 | 0.020440 | 0.020650 | 0.748760 | -0.097890 | 0.128074 | 0.161050 | 0.011937 |
| tc_1111111111110 | | | 1 | | | -0.087178 | 0.175564 | -0.013525 | 0.833861 | 0.128463 | 0.045445 | -0.103700 | 0.106844 | 0.058365 | 0.364989 |
| tc_1111111111111 | | | 1 | | | -0.036209 | 0.574304 | 0.106462 | 0.097777 | -0.131724 | 0.040195 | -0.116691 | 0.069392 | -0.075852 | 0.238788 |

- End of document -