

Predictive Linguistic Features of Schizophrenia

Efsun Sarioglu Kayi¹, Mona Diab¹, Luca Pauselli², Michael Compton², Glen Coppersmith³

¹Department of Computer Science, George Washington University
{efsun, mtdiab}@gwu.edu

²Medical Center, Columbia University
{mtc2176@cumc.columbia.edu, pausell@nyspi.columbia.edu}

³Qntfy
glen@qntfy.com

Abstract

Schizophrenia is one of the most disabling and difficult to treat of all human medical/health conditions, ranking in the top ten causes of disability worldwide. It has been a puzzle in part due to difficulty in identifying its basic, fundamental components. Several studies have shown that some manifestations of schizophrenia (e.g., the *negative symptoms* that include blunting of speech prosody, as well as the *disorganization* symptoms that lead to disordered language) can be understood from the perspective of linguistics. However, schizophrenia research has not kept pace with technologies in computational linguistics, especially in semantics and pragmatics. As such, we examine the writings of schizophrenia patients analyzing their syntax, semantics and pragmatics. In addition, we analyze tweets of (self proclaimed) schizophrenia patients who publicly discuss their diagnoses. For writing samples dataset, syntactic features are found to be the most successful in classification whereas for the less structured Twitter dataset, a combination of features performed the best.

1 Introduction

Schizophrenia is an etiologically complex, heterogeneous, and chronic disorder. It imposes major impairments on affected individuals, can be devastating to families, and it diminishes the productivity of communities. Furthermore, schizophrenia is associated with remarkably high direct and indirect health care costs. Persons with schizophrenia often have multiple medical comorbidities, have a tragically reduced life expectancy, and are often treated without the benefits of sophisticated *measurement-based care*.

Similar to other psychoses, schizophrenia has been studied extensively on the neurological and behavioral levels. Covington et al. (2005) note the existence of many language abnormalities (in syntactic, semantic, pragmatic, and phonetic domains of linguistics) comparing patients to controls. They observed the following:

- reduction in syntax complexity (Fraser et al., 1986);
- impaired semantics, such as the organization of individual propositions into larger structures (Rodriguez-Ferrera et al., 2001);
- abnormalities in pragmatics which is a level obviously disordered in schizophrenia (Covington et al., 2005);
- phonetic anomalies like flattened intonation (aprosody), more pauses, and constricted pitch/timbre (Stein, 1993).

A few studies have used computational methods to assess acoustic parameters (e.g., pauses, prosody) that correlate with negative symptoms, but schizophrenia research has not kept pace with technologies in computational linguistics, especially in semantics and pragmatics. Accordingly, we analyze the predictive power of linguistic features in a comprehensive manner by computing and analyzing many syntactic, semantic and pragmatic features. This sort of analysis is particularly useful for finding meaningful signals that help us better understand the mental health conditions. To this end, we compute part-of-speech (POS) tags and dependency parses to capture the syntactic information in patients' writings. For semantics, we derive topic based representations and semantic role labels of writings. In addition, we add more semantics by adding dense features using clusters that are trained on online resources. For pragmatics, we consider the sentiment that exists in writings, i.e. *positive* vs. *negative* and its intensity. To

the best of our knowledge, no previous work has conducted comprehensive analysis of schizophrenia patients' writings from the perspective of syntax, semantics and pragmatics, collectively.

2 Predictive Linguistic Features of Schizophrenia

2.1 Dataset

The first dataset called LabWriting consists of 93 patients with schizophrenia who were recruited from sites in both Washington, D.C. and New York City. This includes patients that have a diagnosis of schizophreniform disorder or first-episode or early-course patients with a psychotic disorder not otherwise specified. All patients were native English-speaking patients, aged 18-50 years and cognitively intact enough to understand and participate in the study. A total of 95 eligible controls were also native English speakers aged 18-50. Patients and controls did not differ by age, race, or marital status, however, patients were more likely to be male and had completed fewer years of education. All study participants were assessed for their ability to give consent, and written informed consent was obtained using Institutional Review Board-approved processes. Patients and controls were asked to write two paragraph-length essays: one about their average Sunday and the second about what makes them the angriest. The total number of writing samples collected from both patients and controls is 373. Below is a sample response from this dataset (text from patients rendered verbatim as is including typos):

The one thing that probably makes me the most angry is when good people receive the bad end of the draw. This includes a child being struck for no good reason. A person who is killed but was an innocent bystander. Or even when people pour their heart and soul into a job which pays them peanuts but they cannot sustain themselves without this income. Just in general a good person getting the raw end of deal. For instance people getting laid off because their company made bad investments. the Higher ups keep their jobs while the worker ants get disposed of. How about people who take advantage of others and build an Empire off it like insurance or drug companies. All these good decent people not getting what they deserved. Yup that makes me angry.

In addition, we evaluated social media messages with self-reported diagnoses of schizophrenia using the Twitter API. This dataset includes 174 users with apparently genuine self-stated diagnosis of a schizophrenia-related condition and 174 age and gender matched controls. Schizophrenia users were selected via regular expression on *schizo* for a close phonetic approximation. Each diagnosis was examined by a human annotator to verify that it seems genuine. For each schizophrenia user, a control that had the same gender label and was closest in age was selected. The average number of tweets per user is around 2,800. Detailed information on this dataset can be found in (Mitchell et al., 2015). Below are some tweets from this dataset (they have been rephrased to preserve anonymity):

this is my first time being unemployed. please forgive me. i'm crazy. #schizophrenia

i'm in my late 50s. i worry if i have much time left as they say people with #schizophrenia die 15-20 years younger

#schizophrenia takes me to devil-like places in my mind

2.2 Approach and Experimental Design

We cast the problem as a supervised binary classification task where a system should discriminate between a patient and a control. To classify schizophrenia patients from controls, we trained support vector machines (SVM) with linear kernel and Random Forest classifiers. We used Weka (Hall et al., 2009) to conduct the experiments with 10-fold stratified cross validation. We report Precision, Recall, F-Score, and Area Under Curve (AUC) value which is the area under receiver operating characteristics curve (ROC).

2.2.1 Syntactic Features

To capture the syntactic information from writings, we produce the POS tags and dependency parse trees using Stanford Core NLP (Manning et al., 2014). To use these as features to the classifier, we calculate the frequency of each POS tag and dependencies from parse trees. For the Twitter dataset, we use a parser (Kong et al., 2014) and POS tagger (Gimpel et al., 2011) that are specifically trained for social media data.

2.2.2 Semantic Features

To analyze the semantics of the writings, we consider several sources of information. As a first approach, we use semantic role labeling (SRL). Specifically, we use Semafor (Das et al., 2010) tool to generate semantic role labels of the writings and then calculate the frequency of the labels as features for the classifier. For Twitter dataset, due to its short form and poor syntax, we were not able to compute SRL features.

In addition to SRL, we analyzed the topic distribution of writings using Latent Dirichlet Allocation (LDA) (Blei et al., 2003). With this approach, we want to see the possibility of different themes emerging in the writings of patients vs. controls. Using LDA, we represent each writing as a topic distribution where each topic is automatically learned as a distribution over the words of the vocabulary. We use the MALLET tool (McCallum, 2002) to train the topic model and empirically choose number of topics based on best classification performance on a validation set. The best performing number of topics is 20 for LabWriting dataset and 40 for Twitter dataset.

Finally, we compute dense semantic features by computing clusters based on global word vectors. Specifically, for LabWriting dataset, we use word vectors trained on Wikipedia 2014 dump and Gigaword 5 (Parker, 2011) which are generated based on global word-word co-occurrence statistics (Pennington et al., 2014). For Twitter dataset, we use Twitter models trained on 2 billion tweets.¹ We, then, create clusters of these word vectors using the K-means algorithm (K= 100, empirically chosen) for both datasets. Then, for each writing, we calculate the frequency of each cluster by checking the existence of each word of the document in the cluster. With this cluster based representation, we aim to capture the effect of semantically related words on the classification.

2.2.3 Pragmatic Features

For pragmatics, we wanted to see whether patients exhibit more negative sentiment than controls. For that purpose, we use the Stanford Sentiment Analysis tool (Socher et al., 2013). Given a sentence, it predicts its sentiment at five possible levels: *very negative*, *negative*, *neutral*, *positive*, and *very positive*. For each writing, we calculate the frequency of sentiment levels. Additionally, sentiment inten-

sities are produced at the phrase level. Rather than categorical values, this intensity encodes the magnitude of the sentiment more explicitly. As such, we calculate the total intensity for each document as sum of its phrases' intensities at each level. For Twitter dataset, we use a sentiment classifier that was trained for social media data (Radeva et al., 2016). Its output includes three levels of sentiment *negative*, *neutral*, and *positive* without intensity information.

2.2.4 Feature Analysis

To be able to better evaluate best performing features, we analyze them based on two feature selection algorithms: Information Gain (IG) for Random Forest and Recursive Feature Elimination (RFE) algorithm for SVM (Guyon et al., 2002). The Information Gain measure selects the attributes that decrease the entropy the most. The RFE algorithm, on the other hand, selects features based on their weights based on the fact that the larger weights correspond to the more informative features.

3 Results

The list of syntactic, semantic and pragmatic features are presented in Table 1 for both datasets. Tables 2 and 3 illustrate our results for the LabWriting dataset and Twitter dataset, respectively. The majority baseline F-Score is 34.39 for the LabWriting and 32.11 for Twitter dataset. The top performance for each dataset and classifier is shown in bold. The corresponding ROC plots for features are shown in Figures 1 and 2 for LabWriting and Twitter datasets respectively. In each ROC plot, true positive rate (recall) is plotted against true negative rate where SVM is shown in magenta and Random Forest is shown in blue. The diagonal line from bottom left to upper right represents random guess and better performing results are closer to upper left corner. Overall, Random Forest performs better than SVM even though for some feature combinations, SVM's performance is higher. This could be due to bootstrapping of samples that takes place in Random Forest since both of the datasets are on the smaller side. For LabWriting dataset, the best performing features according to F-Score are syntactic: POS+Parse (syntax) for SVM and syntax + pragmatics features for Random Forest. According to AUC, best performing feature is POS for both classifiers. For Twitter dataset, the best performing features according to

¹<http://nlp.stanford.edu/projects/glove/>

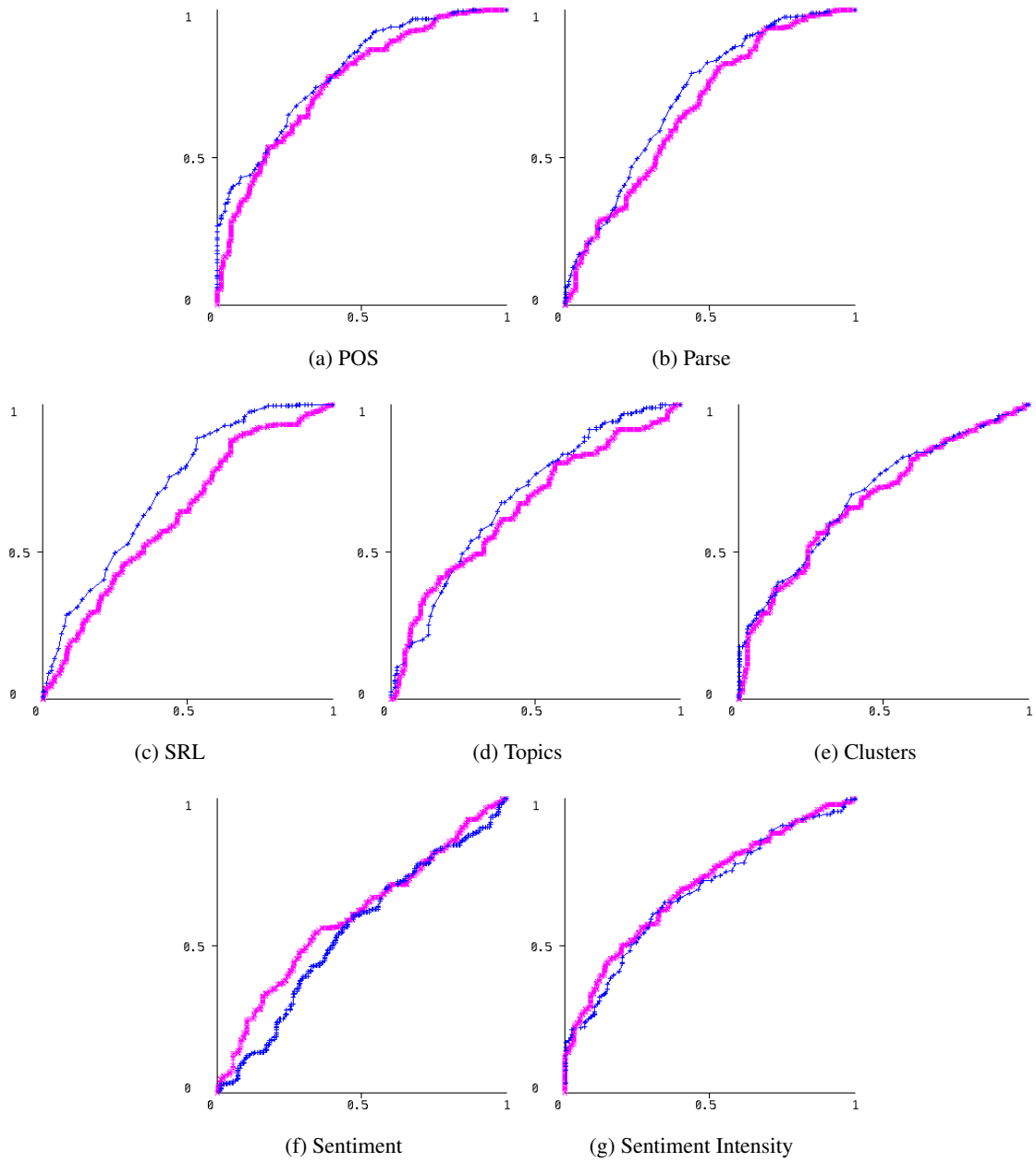


Figure 1: ROC Plots for LabWriting Dataset

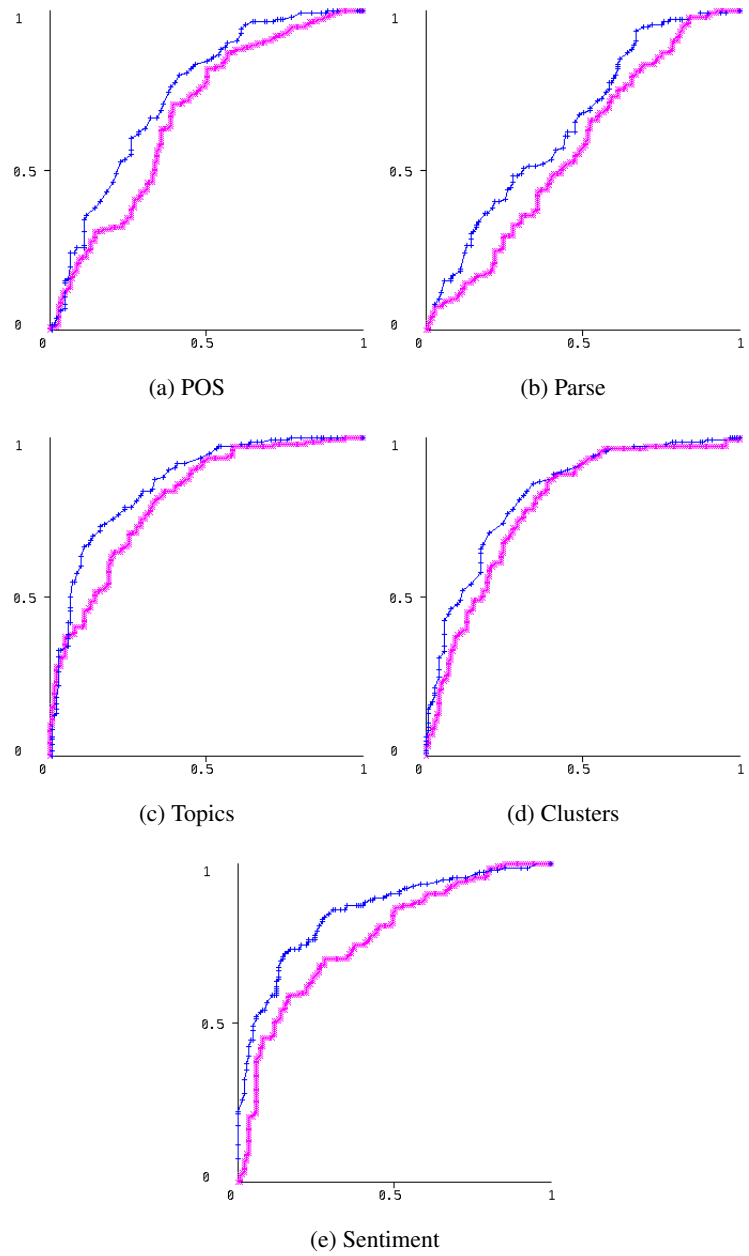


Figure 2: ROC Plots for Twitter Dataset

Table 1: Feature Categories

Category	Writing Samples	Twitter
Syntactic	POS, Dependency Parse	POS, Dependency Parse
Semantic (Sem.)	SRL, Topics, Clusters	Topics , Clusters
Pragmatic (Prag.)	Sentiment, Sentiment Intensity	Sentiment

Table 2: Classification Performance of LabWriting Dataset

Features	SVM				Random Forest			
	AUC	Precision	Recall	F-Score	AUC	Precision	Recall	F-Score
POS	75.72	68.89	68.07	68.48	78.92	70.11	69.41	69.76
Parse	65.34	60.15	59.23	59.69	66.68	66.74	65.16	65.94
SRL	64.25	58.65	58.24	58.44	70.62	65.22	64.64	64.93
Topics	66.49	63.52	63.17	63.34	68.26	62.77	62.34	62.55
Clusters	69.68	65.12	64.85	64.98	68.43	65.38	64.62	65.00
Sentiment	60.23	54.99	53.83	54.40	56.27	57.97	57.66	57.81
Sentiment Intensity	69.98	64.20	64.07	64.13	69.39	65.31	64.62	64.96
Syntax	74.17	69.38	68.38	68.88	75.78	69.25	68.09	68.67
Semantics	66.46	61.59	61.05	61.32	69.16	64.39	63.72	64.05
Pragmatics	68.95	62.67	62.45	62.56	69.59	67.57	66.78	67.17
Syntax + Sem.	68.24	66.36	66.12	66.24	76.60	71.24	69.36	70.29
Syntax + Prag.	73.43	68.14	67.28	67.71	78.75	71.74	70.52	71.12
Sem.+ Prag.	68.01	63.98	63.46	63.72	72.09	65.75	64.81	65.28
All	70.11	66.94	66.66	66.80	77.57	71.18	69.91	70.54

both F-Score and AUC are the ones that include most of the combination of features: semantics + pragmatics for Random Forest and all features for SVM. Typically, essays, such as the ones in LabWriting dataset, are expected to have better syntax than informal tweets and as such syntactic features were not as predictive for tweets. We also analyze top performing features according to Information Gain measure and SVM RFE algorithm in Sections 3.1, 3.3, 3.2 and explain the differences of results for the two datasets in Section 3.4.

3.1 Top Syntactic Features

Syntactic features perform well mainly for LabWriting dataset. Between POS tags and dependence parses, the former perform better for both datasets. For LabWriting dataset, the top POS tag is *FW*, (Foreign Word). When we look at the words that were tagged *FW*, they correspond to misspelled words. Even though this could be considered a criterion for schizophrenia patients, it may also depend on patients and controls' education and language skills which we expect it to be similar but it may still show some differences. Another top POS tag is *LS*, (List item marker),

which was assigned to small case *i* which in reality refers to pronoun *I*. This could imply that the patients prefer to talk about themselves. This coincides with several other studies (Rude et al., 2004; Chung and Pennebaker, 2007) which found that use of first person singular is associated with negative affective states such as depression. Because of the likelihood of comorbidity of mental illnesses, this requires further investigation as to whether this is specific to schizophrenia patients or not. Finally, another top POS tag is *RP*, adverbial particle and top parse tag is *advmod*, adverb modifier. This could mean the ratio of adverbs used could be a characteristic of patients. Finally for Twitter dataset, the top POS tag is *#* corresponding to hash tags. This could be an important discriminative feature between patients and controls as patients use less hashtags than controls.

3.2 Top Semantic Features

For classification using semantic features, clusters, topics and SRL perform comparably. For LabWriting dataset, top SRL features consist of general categories and some specific ones that could be relevant for schizophrenia patients. General la-

Table 3: Classification Performance of Twitter Dataset

Features	SVM				Random Forest			
	AUC	Precision	Recall	F-Score	AUC	Precision	Recall	F-Score
POS	69.34	67.83	59.14	63.19	75.17	69.48	68.92	69.20
Parse	58.63	44.76	54.06	48.97	63.72	60.94	60.63	60.78
Topics	79.88	74.85	73.66	74.25	83.48	79.38	78.77	79.07
Clusters	78.02	73.87	71.27	72.55	82.54	74.80	73.76	74.28
Sentiment	75.79	65.15	60.50	62.74	85.28	80.00	79.50	79.75
Syntax	74.87	68.16	62.13	65.01	74.35	67.72	67.47	67.59
Semantics	80.29	70.81	70.34	70.57	85.62	75.00	74.72	74.86
Syntax+Sem.	81.47	73.02	72.55	72.78	86.35	78.30	78.02	78.16
Syntax+Prag.	82.16	75.61	72.22	73.88	83.55	75.89	75.52	75.70
Sem.+Prag.	80.99	74.52	74.05	74.28	88.98	82.01	81.30	81.65
All	82.58	75.18	74.39	74.78	88.01	78.81	78.40	78.60

bels are *Quantity* and *Social Event*. More specific labels are *Morality Evaluation*, *Catastrophe*, *Manipulate into Doing* and *Being Obligated*. Words that are labeled as such are listed in Table 4. These two different sets of labels could be due to the type of questions asked to the patients. One question is neutral in nature talking about their daily life whereas the other is about the things that make them angry and more emotionally charged. A second semantic feature is the topic distributions of writings. The top words from the most informative topics are listed in Table 5. For LabWriting dataset, one of the top topics consist of words about typical Sunday activities corresponding to one of the questions asked. The second top topic, on the other hand, consist of words that show the anger of the author. For Twitter dataset, one of the topics consist of schizophrenia-related words and the other consist of hate words. Again, the top topics seem to contain relevant information in analyzing schizophrenia patients’ writings and classification using topic features perform comparably well. As a final semantic feature, we use dense cluster features. The classification performance of cluster features is similar to classification performance using topics. However, cluster features’ analysis is not as interpretable as topics, since they are formed from massive online data resources.

3.3 Top Pragmatic Features

When it comes to pragmatic features, top sentiment features are *neutral*, *negative* and *very negative* (LabWriting only). For sentiment intensity, *neutral intensity*, *negative intensity* and *very negative intensity* are more informative which is con-

sistent with sentiment categorical analysis. In general, neutral sentiment is the most common for a given text and for patients, we would expect to see more negative sentiment and this was confirmed by this analysis. However, negative sentiment could also be prominent in other psychiatric diseases such as post-traumatic stress disorder (PTSD)(Coppersmith et al., 2015), as such, by itself, it may not be a discriminatory feature for schizophrenia patients. For classification purposes, sentiment intensity features performed better than sentiment features. This could be due to the fact that intensity values are more specific and collected at word/phrase level in contrast to sentence level.

3.4 Effect of Datasets’ Characteristics

The two datasets have some commonalities and differences and present different challenges. The LabWriting dataset was collected in a more controlled manner and follows a structure that can be expected from a short essay. Accordingly, NLP tools applied to these writings are successful. On the other hand, the Twitter dataset consists of combinations of short text that include many abbreviations that are not standard, e.g. users’ own solutions to fixed length limit imposed by Twitter. It is also very informal in nature and thus lacks proper grammar and syntax more frequently than LabWriting. Hence, some machine learning approaches for NLP analysis of these tweets are limited even though social media specific tools were used such as POS tagger (Gimpel et al., 2011), dependency parser (Kong et al., 2014), sentiment analysis tool (Radeva et al., 2016), and Twitter

Table 4: **Top Discriminative SRL Features**

Label	Sample Words/Phrases
Quantity	several, both, all, a lot, many, a little, a few, lots
Social Event	social, dinner, picnics, hosting, dance
Morality Evaluation	wrong, depraved, foul, evil, moral
Catastrophe	incident, tragedy, suffer
Manipulate into Doing	harassing, bullying
Being Obligated	duty, job, have to, had to, should, must, responsibility, entangled, task

Table 5: **Discriminative Topics' Top Words**

Method	Dataset	Top Words
IG	Writing	church sunday wake god service pray sing worship bible spending thanking
IG&RFE	Writing	i'm can't trust upset person lie real feel honest lied lies judge lying steal
IG&RFE	Twitter	god jesus mental schizophrenic schizophrenia illness paranoid medical evil
IG&RFE	Twitter	don love people fuck life feel fucking hate shit stop god person sleep bad die

models for dense clusters. For instance, even though we were able to compute POS tags and parse trees for tweets, the tag set is much smaller than PennTree Bank tags set. Similarly, some approaches such as SRL were not successful on tweets. On the other hand, both datasets consist of patients and controls with similar demographics (age, gender, etc), thus we largely expect patients and controls to have similar linguistic capabilities. In addition, for LabWriting dataset, patients and controls were recruited from the same neighborhoods. We have no such explicit guarantees for the Twitter dataset, though they were excluded if they did not primarily tweet in English. Accordingly, any differentiation these classification methods found can largely be attributed to the illness. Finally, LabWriting dataset had many spelling errors. We elected not to employ any spelling correction techniques (since misspelling may very well be a feature meaningful to schizophrenia). This likely negatively influenced the calculation of some of the features which depend on correct spelling such as SRL.

4 Related Work

To date, some studies have investigated applying Latent Semantic Analysis (LSA) to the problem (Elvevag et al., 2007) of lexical coherence and they found significant distinctions between schizophrenia patients and controls. The work of (Bedi et al., 2015) extends this approach by incorporating syntax, i.e., phrase level LSA measures

and POS tags. In the latter related work, several measures based on LSA representation were developed to capture the possible incoherence in patients. In our study, we used LDA to capture possible differences in themes between patients and controls. LDA is a more descriptive technique than LSA since topics are represented as distributions over vocabulary and top words for topics provide a way to understand the theme that they represent. We also incorporated syntax to our analysis with POS tags and additionally dependency parses. Another work by (Howes et al., 2013) predicts outcomes by analyzing doctor-patient communication in therapy using LDA. Even though manual analysis of LDA topics with manual topics seems promising, classification using topics does not perform as successful unless otherwise additional features are incorporated such as doctors' and patients' information. Although, we had detailed demographic information for LabWriting dataset and derived age and sex information for Twitter dataset, we chose not to incorporate them to the classification process be able focus solely on writings' characteristics.

The work of Mitchell et al. (2015) is, in many respects, similar to ours by examining schizophrenia using LDA, clustering and sentiment analysis. Their sentiment analysis is lexicon-based using Linguistic Inquiry Word Count (LIWC) (Tausczik and Pennebaker, 2010) categories. In our approach to sentiment analysis, we utilized a machine learning approach. Lexicon-based ap-

proaches generally have higher precision at the cost of lower recall. Having coverage of more of the content may be beneficial for analysis and interpretation, so we opt to use a more generalizable machine learning approach. For clustering, they used Brown clustering; whereas, we used clusters trained on global word vectors which were learned from large amounts of online data. This has the advantage that we could capture words and/or semantics that may not be learned from our dataset. Finally, their use of LDA is similar to our approach, i.e. representing documents as topic distributions, and their analysis does not include syntactic and dense cluster features. They had their best performance with an accuracy value of 82.3 using a combination of topic based representation and their version of sentiment features. In our analysis, combination of semantic and pragmatic features performed the best with an accuracy value of 81.7. Due to possible differences in preprocessing, parameter selection, and randomness that exist in the experiments, the results are not directly comparable, however, this also shows that the difficulty of applying more advanced machine learning based NLP techniques for Twitter dataset.

5 Conclusion

Computational assessment models of schizophrenia may provide ways for clinicians to monitor symptoms more effectively and a deeper understanding of schizophrenia and the underpinning cognitive biases could benefit affected individuals, families, and society at large. Objective and passive assessment of schizophrenia symptoms (e.g., delusion or paranoia) may provide clarity to clinical assessments, which currently rely on patients' self-reporting symptoms. Furthermore, the techniques discussed here hold some potential for early detection of schizophrenia. This would be greatly beneficial to young people and first-degree relatives of schizophrenia patients who are prodromal (clinically appearing to be at high risk for schizophrenia) but not yet delusional/psychotic, since it would allow targeted early interventions.

Among the linguistic features considered for this study, syntactic fetures provide the biggest boost in classification performance for LabWriting dataset. For Twitter dataset, combination of features such as semantics and pragmatics for SVM and syntax, semantics and pragmatics for Random Forest have the best performance.

In the future, we will be focusing on the features that showed the most promise in this study and also add new features such as level of committed belief for pragmatics. Finally, we are collecting more data and we will expand our analysis to more mental health datasets.

References

- Gillinder Bedi, Facundo Carrillo, Guillermo A Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B Mota, Sidarta Ribeiro, Daniel C Javitt, Mauro Copelli, and Cheryl M Corcoran. 2015. Automated Analysis of Free Speech Predicts Psychosis Onset in High-Risk Youths. *Npj Schizophrenia* 1.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3:993–1022.
- Cindy K Chung and James W Pennebaker. 2007. Social Communication .
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses. In *NAACL Workshop on Computational Linguistics and Clinical Psychology*.
- Michael A Covington, Congzhou He, Cati Brown, Lorina Naci, Jonathan T McClain, Bess Sirmon Fjordbak, James Semple, and John Brown. 2005. Schizophrenia and the Structure of Language: The Linguist's View. *Schizophr Res* 77(1):85–98.
- Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic Frame-semantic Parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. pages 948–956.
- Brita Elvevag, Peter W Foltz, Daniel R Weinberger, and Terry E Goldberg. 2007. Quantifying Incoherence in Speech: An Automated Methodology and Novel Application to Schizophrenia. *Schizophr Res* 93(1-3):304–316.
- W I Fraser, K M King, P Thomas, and R E Kendell. 1986. The Diagnosis of Schizophrenia by Language Analysis. *Br J Psychiatry* 148:275–278.
- Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech Tagging for Twitter: Annotation, Features, and Experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*. Association for Computational Linguistics, Stroudsburg, PA, USA, HLT '11, pages 42–47.

- Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene Selection for Cancer Classification Using Support Vector Machines. *Mach. Learn.* 46(1-3):389–422.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.* 11(1):10–18.
- Christine Howes, Matthew Purver, and Rose McCabe. 2013. Using Conversation Topics for Predicting Therapy Outcomes in Schizophrenia. *Biomed Inform Insights* 6(Suppl 1):39–50. <https://doi.org/10.4137/BII.S11661>.
- Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archana Bhatia, Chris Dyer, and Noah A Smith. 2014. A Dependency Parser for Tweets .
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*. pages 55–60.
- Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. [Http://mallet.cs.umass.edu](http://mallet.cs.umass.edu).
- Margaret Mitchell, Kristy Hollingshead, and Glen Coppersmith. 2015. Quantifying the Language of Schizophrenia in Social Media. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. pages 11–20.
- Robert et al. Parker. 2011. English Gigaword Fifth Edition LDC2011T07.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. pages 1532–1543.
- Axinia Radeva, Mohammad Rasooli, and Kathleen McKeown. 2016. Columbia Language Independent Sentiment System. Technical report, Columbia University.
- S Rodriguez-Ferrera, R A McCarthy, and P J McKenna. 2001. Language in Schizophrenia and Its Relationship to Formal Thought Disorder. *Psychol Med* 31(2):197–205.
- Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language Use of Depressed and Depression-Vulnerable College Students. *Cognition and Emotion* 18(8):1121–1133.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Conference on Empirical Methods in Natural Language Processing (EMNLP)* .
- J Stein. 1993. Vocal Alterations in Schizophrenic Speech. *J Nerv Ment Dis* 181(1):59–62.
- Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* pages 24–54.