The University of Maryland CLPsych 2015 Shared Task System

Philip Resnik[2,4], **William Armstrong**[1,4], **Leonardo Claudino**[1,4], **Thang Nguyen**[3]
[1]Computer Science, [2]Linguistics, [3]iSchool, and [4]UMIACS, University of Maryland
{resnik,armstrow}@umd.edu
{claudino,daithang}@cs.umd.edu

1 Introduction

The 2015 ACL Workshop on Computational Linguistics and Clinical Psychology included a shared task focusing on classification of a sample of Twitter users according to three mental health categories: users who have self-reported a diagnosis of depression, users who have self-reported a diagnosis of post-traumatic stress disorder (PTSD), and control users who have done neither (Coppersmith et al., 2015; Coppersmith et al., 2014). Like other shared tasks, the goal here was to assess the state of the art with regard to a challenging problem, to advance that state of the art, and to bring together and hopefully expand the community of researchers interested in solving it.

The core problem under consideration here is the identification of individuals who suffer from mental health disorders on the basis of their online language use. As Resnik et al. (2014) noted in their introduction to the first ACL Workshop on Computational Linguistics and Clinical Psychology, few social problems are more costly than problems of mental health, in every possible sense of cost, and identifying people who need help is a huge challenge for a variety of reasons, including the fear of social or professional stigma, an inability of people to recognize symptoms and report them accurately, and the lack of access to clinicians. Language technology has the potential to make a real difference by offering low-cost, unintrusive methods for early screening, i.e. to identify people who should be more thoroughly evaluated by a professional, and for ongoing monitoring, i.e. to help providers serve their patients better over the long-term continuum of care (Young et al., 2014).

This paper summarizes the University of Maryland contribution to the CLPsych 2015 shared task. More details of our approach appear in Resnik et al. (2015), where we also report results on preliminary experimentation using the CLPsych Hackathon data (Coppersmith, 2015).

2 System Overview

In our system, we build on a fairly generic supervised classification approach, using SVM with a linear or RBF kernel and making use of baseline lexical features with TF-IDF weighting.

2.1 Variations explored

The innovations we explore center on using topic models to develop features that capture latent structure in the dataset, going beyond "vanilla" latent Dirichlet allocation (Blei et al., 2003) to include supervised LDA (Blei and McAuliffe, 2008, sLDA) as well as a supervised variant of the "anchor" algorithm (Arora et al., 2013; Nguyen et al., 2015, sAnchor). Putting together various combinations in our experimentation — linear vs. RBF kernel, big vs. small vocabulary, and four feature configurations (namely sLDA, sAnchor, lexical TF-IDF, and all combined), we evaluated a total of 16 systems for each of the three shared tasks (discriminating depression vs. controls, depression vs. PTSD, and PTSD vs. controls) for a total of 48 systems in all.

In general below, systems are named simply by concatenating the relevant elements of the description. For example, *combobigvocabSVMlinear_1* is the name of the system that uses (a) an SVM with linear kernel (*SVMlinear*), (b) models computed using the big vocabulary (*bigvocab*, details below), and (c) the "all combined" feature configuration

Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 54–60,
Denver, Colorado, June 5, 2015.

(*combo*). The numerical suffix is for internal reference and can be ignored. The names of all systems are shown in the legends of Figure 1 grouped by each pair of conditions.

As an exception to our general scheme, we also explored using sLDA to make predictions directly rather than providing topic posterior features for the SVM, i.e. by computing the value of the regression variable as a function of the posterior topic distribution given the input document (Blei and McAuliffe, 2008, sLDA). These systems are simply referred to as *SLDA Prediction*.

2.2 SLDA and SAnchor topic features

We briefly describe the features we used based on sLDA and sAnchor; see Resnik et al. (2015) for more details, as well as sample topics induced by these models on the closely related CLPsych Hackathon dataset. For both topic models, we used posterior topic distributions, i.e. the vector of $\Pr(\text{topic}_k|\text{document}), k = 1..K$ in a K-topic model, as features for supervised learning.

SLDA posteriors with informed priors. To take full advantage of the shared task's labeled training data in a topic modeling setting, we opted to use *supervised* topic models (sLDA, introduced by Blei and McAuliffe (2008)), as a method for gaining both clinical insight and predictive ability. However, initial exploration with the training dataset provided noisy topics of variable quality, many of which seemed intuitively unlikely to be useful as predictive features in the mental health domain. Therefore we incorporated an informed prior based on a model that we knew to capture relevant latent structure.

Specifically, we followed Resnik et al. (2013) in building a 50-topic model by running LDA on stream-of-consciousness essays collected by Pennebaker and King (1999) — a young population that seems likely to be similar to many authors in the Twitter dataset. These 50 topics were used as informed priors for sLDA.

Tables 3 to 5 show the top words in the sLDA topics with the 5 highest and 5 lowest Z-normalized regression scores, sLDA having been run with parameters: number of topics $(k) = 50$, document-topic Dirichlet hyper-parameter $(\alpha) = 1$, topic-word Dirichlet hyper-parameter $(\beta) = 0.01$, Gaussian variance for document responses $(\rho) = 1$, Gaussian variance for topic's regression parameters $(\sigma) = 1$, Gaussian mean for topic's regression parameters $(\mu) = 0.0$, using binary variables for the binary distinction in each experimental task.

Supervised anchor (SAnchor) posteriors. The anchor algorithm (Arora et al., 2013) provides a fast way to learn topic models and also enhances interpretability by automatially identifying a single "anchor" word associated with each topic. For example, one of the topics induced from the Hackathon tweets is associated with the anchor word *fat* and is characterized by the following most-probable words in the topic:

> *fat eat hate body sleep weight girl bed skinny cry fast beautiful die perfect cross hair ugh week sick care*

Nguyen et al. (2015) introduce SANCHOR, a supervised version of the anchor algorithm which, like sLDA, jointly models text content along with a per-document regression variable. We did not explore informed priors with SANCHOR in these experiments; this is left as a question for future work.

2.3 Classifier details

The majority of our experiments used SVM classifiers with either a linear or an RBF kernel. Specifically, we used the python *scikit-learn* module (*sklearn.svm.SVC*), which interfaces with the widely-used *libsvm*. Default parameters were used throughout except for the *class_weight* parameter, which was set to *None*.

In the *SLDA Prediction* experiments, we followed Blei and McAuliffe (2008) in computing the response value for each test document from $\eta^\top \bar{z}$ where $\bar{z}$ is the document's posterior topic distribution and the ηs are the per-topic regression parameters. SLDAPrediction_1 and SLDAPrediction_2 were conducted with small and big vocabularies, respectively.

2.4 Data Preparation

Data organization: weekly aggregation. To overcome potential problems for topic modeling with documents that are too small (individual tweets) or too large (all tweets for an author) we grouped tweets together by the week they were posted. Thus each author corresponded to several documents, one for each week they tweeted one or

Notes	Valence	Top 20 words
high emotional valence	e	life live dream change future grow family goal mind rest decision marry chance choice successful career set regret support true
high emotional valence	e	love life happy heart amaze hurt perfect crazy beautiful lose smile cry boy true fall real sad relationship reason completely
relationship problems	n	time boyfriend friend relationship talk person break doe happen understand hard trust care spend reason san situation antonio date leave
transition to college	n	school college student semester university experience hard grade parent graduate freshman campus learn texas attend teacher expect challenge adjust education
self-doubt	n	question realize understand completely idea sense level bring issue concern simply situation lack honestly admit mention fear step feeling act
poor ego control	n	yeah suck wow haha stupid funny hmm crap crazy blah freak type ugh weird lol min gosh hey bore hmmm
feeling ignored/annoyed *	n	call talk phone doe stop bad ring message loud head homework answer cell mad forget annoy sound hurt suppose mine
somatic complaints	n	cold hot feel sick smell rain walk start weather bad window foot freeze nice wait throat day heat hate warm
emotional distress *	n	feel happy day sad depress feeling cry scar afraid lonely head moment emotion realize confuse hurt inside guilty fear upset
family of origin issues	n	mom dad family sister parent brother kid child mother father grow doctor baby hard cousin die age cry proud husband
negative affect *	n	damn hell doe shit fuck smoke woman hate drink piss sex drug kid god bitch time real break screw cigarette
anxiety over failure	n	worry hard study test class lot grade focus mind start nervous stress concentrate trouble reason easier hop harder fail constantly
negative affect*	n	hate doe bad stupid care understand time suck happen anymore mad don mess scar horrible smart matter hat upset fair
sleep disturbance*	n	sleep tire night morning wake bed day time late stay hour asleep nap fall start tomorrow sleepy haven awake lay
somatic complaints	n	hurt eye hear itch hand air sound tire nose arm loud leg leave noise finger smell neck stop light water
social engagement	p	game football team win ticket excite school weekend week texas run lose night season saturday sport dallas longhorn coach fan
exercise, good self-care	p	run day feel walk class wear lose weight buy gym gain short fat dress shop exercise campus clothe body shirt

Table 1: LDA topics from Pennebaker stream-of-consciousness essays identified by a clinician as most relevant for assessing depression. Topics with negative valence (n) were judged likely to be indicators for depression, those with positive valence (p) were judged likely to indicate absence of depression, and those labeled (e) have strong emotional valence without clearly indicating likely assessment. Asterisked topics were viewed as the strongest indicators. Many more of the 50 topics from this model were intuitively coherent but not judged as particularly relevant for the depression-assessment task. This table is reproduced from Resnik et al. (2015).

more times; each document was treated as being labeled by the author's individual-level label. In preliminary experimentation, we found that this temporal grouping greatly improved the performance of our models, though it should be noted that organizing the data in this way fails to account for the fact that an author's mental health can vary greatly from week to week. For instance, a user identified as having depression at some point may not be experiencing symptoms in any given week, yet that week's document would still be labeled as positive for depression. This could potentially be mitigated in future work by attempting to identify the time of diagnosis and increasing the label weight on documents near that time.

Token pre-processing and vocabularies. All systems went through the same basic pre-processing: we first removed words with non-alphanumeric characters, emoticon character codes, and stop words.[1] The remaining tokens were further lemmatized.

For SVM classification we explored using systems with both *small* and *big* vocabularies. For the former, we first filtered out documents with less than 50 tokens and then selected tokens that appeared more than 100 documents; the latter was obtained in a similar fashion, except setting the word-per-document cutoff to 10 rather than 50, and the document-per-word cutoff to 30 instead of 100.[2]

For *sLDA* prediction, we used an external vocabulary produced from the set of 6,459 stream-of-consciousness essays collected between 1997 and 2008 by Pennebaker and King (1999), who asked students to think about their thoughts, sensations, and feelings in the moment and "write your thoughts as they come to you". As discussed in Section 2, running LDA on this dataset provided informative priors for sLDA's learning process on the Twitter training data. The student essays average approximately 780 words each, and for uniformity, we pre-processed them with the same tools as the Twitter set.[3] We created a shared vocabulary for our models by taking the union of the vocabularies from the two datasets, resulting in a roughly 10-20% increase in vocabulary size over the Twitter dataset alone.

Author-level features. In order to arrive at a single feature vector for each author based on documents aggregated at the weekly level, we weight-averaged features across weeks, where weights corresponded to the fraction of the author's tweets associated with each week alone. In other words, the more an author posted in a week, the more important that week's features would be, compared to the

[1] Unicode emoticons were left in, converted to the token EMOJI.

[2] When referring to vocabulary size, we use the terms *short* and *small* interchangeably.

[3] With the exception of the document count filters, due to the different number and sizes of documents, which were adjusted accordingly.

other weeks.

Data splits. We did an 80-20 partition into training and development sets, respectively. Since we did not do any hyper-parameter tuning, the dev set was used primarily for sanity checking and to get a preliminary sense of system performance. We report test set results based on models that were trained on the training set alone.[4]

3 Results

3.1 Overall results and ROCs

The ROC curves for all our submitted systems on the shared tasks (Section 2) are shown in Figure 1. The area under curve (AUC) scores for TF-IDF (baseline) and all configurations of combined features (best systems) are shown in Table 2, from which we see that the 8 best-performing feature configurations achieved an average AUC of about 0.84. We obtained the best overall results when we used a big vocabulary, combined all features, and trained a linear SVM. We saw that bigger vocabularies improved performance of linear SVMs but not RBF SVMs, and that, in general, linear SVMs did better.

The order of difficulty for these discrimination problems seems to be DvP $>$ DvC $>$ PvC, judging from the performance of our top 8 systems. This suggests that there may be greater overlap of linguistic signal between tweets from people who have self-reported PTSD and those who have self-reported depression, which is not entirely surprising since the two conditions often co-occur. According to Tull (2015), “Depression is one of the most commonly occurring disorders in PTSD... [A]mong people who have or have had a diagnosis of PTSD, approximately 48% also had current or past depression ...People who have had PTSD at some point in their life are almost 7 times as likely as people without PTSD to also have depression.”

3.2 Qualitative discussion for sLDA

To get a sense of the role that supervised topic modeling may be playing, we take a brief qualitative look at the topics induced by sLDA on the training set. Tables 3,4, and 5 show the most polarized

[4]It is possible that modest improvements could be obtained by folding the dev set back into the training data, but we wished to avoid inspecting the dev set so that we can continue to use it for further development.

Feature configuration / Problem AUC	DvC	DvP	PvC
tfidfshortvocabSVMlinear	0.824	0.808	0.860
tfidfbigvocabSVMlineasr	0.845	0.827	0.884
tfidfshortvocabSVMrbf	0.831	0.812	0.872
tfidfbigvocabSVMrbf	0.815	0.798	0.855
comboshortvocabSVMlinear	0.841	0.832	0.879
combobigvocabSVMlinear	**0.860**	**0.841**	**0.893**
comboshortvocabSVMrbf	0.835	0.818	0.876
combobigvocabSVMrbf	0.830	0.811	0.869

Table 2: Area under curve (AUC) of selected feature configurations in Fig. 1 per each problem: depression vs. control (DvC), depression vs. PTSD (DvP) and PTSD vs. control (PvC). Boldface: big vocabulary, combined features, SVM linear. This setting was the best for all three tasks.

topics resulting from the sLDA models constructed on the DvC, DvP and PvC tasks respectively, where polarization is measured by the value of the sLDA regression variable for the topic. The topics we see are not as clean and coherent as the topics in Table 1, which is unsurprising since the latter topics came from LDA run on individually coherent documents (stream-of-consciousness essays) collected from a more uniform population (UT Austin college students) (Pennebaker and King, 1999), as contrasted with aggregations of tweets over time from a sample of Twitter users.

At the same time, there does seem to be interpretable signal distinguishing the high versus low polarity topics, at least in comparisons against controls. Comparing depression vs. control (Table 3), we see subdivisions of negative affect — for example, the most depression-oriented topic, as identified using positive regression values, is dominated by negatively oriented interjections (*fuck*, *shit*, *damn*, etc.), and the next most depression oriented topic appears to largely capture relationship discussion (*omg*, *cute*, *cry*, *guy*, *feel*, *hot*, *pretty*). Conversely, the least depression-oriented topics in the table, i.e. with the most negative regression values, contain not only many positive affect terms (*lol*, *haha*, etc.) but also activities related to family (*car*, *weekend*, *home*) and social activity (*food*, *tonight*, *party*, *dinner*, *weekend*).

In PTSD vs. control (Table 5), we see, among the topics more oriented toward PTSD users, topics that may be related to attention to veteran issues (*sign*, *support*, *homeless*, *petition*, *marine*), and possibly

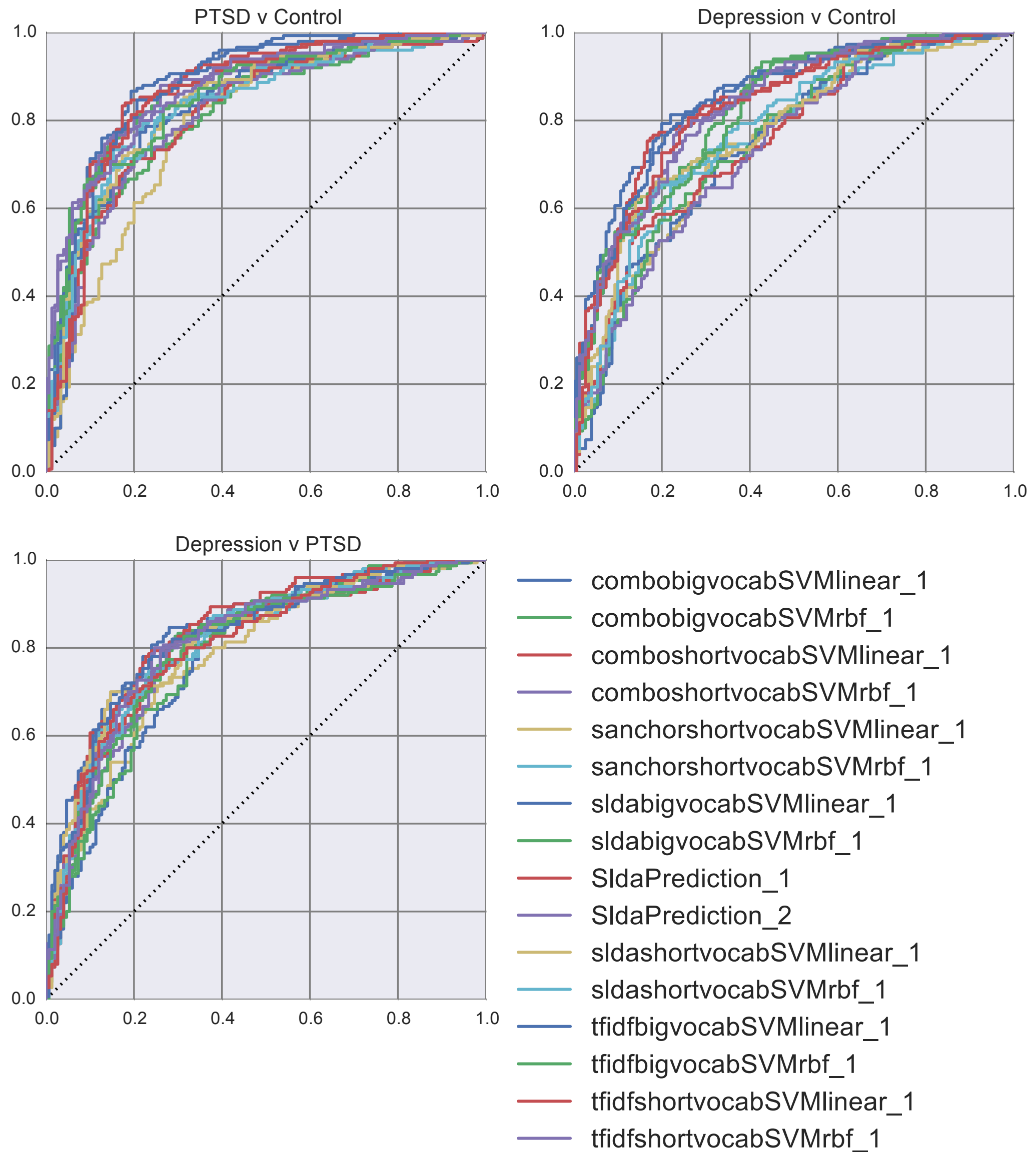

Figure 1: ROC curves of submitted systems.

Regression value	Top 20 words
5.362	fuck shit bitch sex smoke dick drink girl damn fuckin suck weed wanna life wtf hell gonna gay hate drug
4.702	omg cute cry gonna god guy demi idk literally feel wow hot pretty dont bye perfect pls ugh omfg laugh
4.204	line feel people cross friend comment doe start time link mental depression life live health submit deal talk lot issue
3.132	watch movie time episode read write season totally book favorite play character awesome scene star stuff cool horror start hug
2.877	week post baby inbox month day hey pain ago pregnant hun girl start doe bad boy feel time ive private
-1.689	food tonight truck night bring android party dinner tomorrow weekend awesome island game free wine lunch bar complete jack live
-1.87	nigga shit bitch hoe bout real tho gotta ima aint money lil wit bruh tryna mad yall damn ppl smh
-2.584	lol lmao damn smh yea gotta hell dude gon tho watch baby lmfao EMOJI wtf black bro idk boo funny
-2.966	car weekend home house drive summer miss week beach family rain weather run dog ready leave cancer race ride hour
-3.017	haha hahaha yeah hahahaha time night hahah wait watch ill love feel drink dad brother sleep phone sister eat miss

Table 3: Most extreme sLDA topics from Twitter training data (Depression (1) vs. Control (-1))

Regression value	Top 20 words
3.342	harry boy direction louis niall liam guy zayn demi fan tweet fandom laugh video tour day love concert people proud
2.984	EMOJI EMOJI night love EMOJI EMOJI EMOJI EMOJI EMOJI EMOJI EMOJI EMOJI tonight miss girl people EMOJI happy feel tomorrow
2.933	yeah pretty lot stuff play doe cool time send weird wait aww favourite kinda twitter awesome wow happen cat sound
2.708	bitch lmao nigga shit girl wanna hoe talk fuck dick bae damn baby lmfao pussy EMOJI text school boy lil
2.227	girl cute wanna boy guy friend love hate hair text life mom kiss hot feel fall relationship literally boyfriend date
-1.847	kid halloween call guy drink beer fun college throw sam hey dress pick scream play star remember walk porn doe
-2.11	child read change public agree abuse issue record system service kid pay refuse article response court lie business company doe
-2.357	obama tcot vote american ppl ebola america president gop gun country isi texas pay law lie idiot democrat military illegal
-2.568	food live beach town local fresh city coffee time life ago meet house chef fish street change nyc month san
-2.682	ptsd learn fear create canada meet experience speak positive step battle join voice awareness hear youth future world understand key

Table 4: Most extreme sLDA topics from Twitter training data (Depression (1) vs. PTSD (-1))

Regression value	Top 20 words
5.007	people woman doe call black white sex gay real kid word person twitter dude wrong lady marriage female marry tweet
3.581	sign support free share people day family time release send stand fight homeless petition marine pic hero home raise info
3.498	time doe cat lot tweet buy wife twitter feel haven move yep sit door house nice wear glad leave send
3.472	story child mother ptsd mom life son talk death surprise family mental parent woman care save daughter difference pls watch
3.238	feel day eat lose time fat body hard weight start run sleep gym workout fast cut stop food pain stay
-1.979	lol lmao ppl yea dat tho jus gotta wat smh kno dnt money yal dey damn cuz leo tht wen
-2.013	EMOJI love EMOJI EMOJI girl EMOJI day EMOJI EMOJI EMOJI EMOJI EMOJI EMOJI wanna miss people EMOJI EMOJI EMOJI night
-2.318	iphone apple player app phone bowl super youtube free update add ipad hand note box review pro game google play
-2.418	school class sleep tomorrow day feel hate bed tire home night hour homework study people teacher start wake boyfriend gonna
-2.743	haha hahaha yeah night love xxx sleep feel babe miss bed mum girl wait home ill bore boy phone tonight

Table 5: Most extreme sLDA topics from Twitter training data (PTSD (1) vs. Control (-1))

mental health issues including PTSD itself (*story*, *mother*, *ptsd*, *death*, *surprise*, *mental*).

Consistent with the lower performance on depression vs. PTSD (DvP), in Table 4 no topics jump out quite as forcefully as being polarized toward one condition or the other, except for the most PTSD-oriented topic, which appears as if it may concern efforts to draw attention to PTSD (*ptsd*, *learn*, *fear*, *speak*, *positive*, *step*, *battle*, *join*, *voice*, *awareness*). It may be, however, that in incorporating the depression vs. PTSD distinction, the model is actually capturing broader characteristics of relevant subpopulations: particularly in this dataset, people self-reporting a PTSD diagnosis may well be older on average than people self-reporting a depression diagnosis, if not chronologically than in terms of life experience. The topic with the most positive regression value in the table, i.e. leaning toward depression rather than PTSD, includes terms most likely related to youth/pop culture: *Niall* Horan, *Harry* Styles, *Liam* Payne, and *Louis* Tomlinson are the members of the pop boy band One Direction. Other positive- (i.e. depression-)leaning topics in the table also have a quality of disinhibition more characteristic of younger people. In contrast, the negative- (i.e. PTSD-)leaning topics in the table tend toward more mature topics, including, for example, politics and current affairs (*obama*, *tcot* (top conservatives on Twitter), *vote*, *ebola*).

Although our efforts are still in an early stage, our hope is that more sophisticated topic models can not only enhance predictive accuracy, as in Table 2, but also that observations like these about topics or themes might help create insight for clinicians. Examples like the ones in Tables 1 and 3-5 can help establish face validity with clinicians by showing that these models can capture things they already know about. Others can potentially lead to new questions worth asking, e.g. in Table 3, might the topic relating to entertainment (*watch*, *movie*, *episode*, *read*, *write*, *season*, *book*) suggest a closer look at social isolation (staying in watching movies, reading books) as a linguistically detectable online behavior that might correlate with increased likelihood of depression? If true, this would be consistent with, and complement, Choudhury et al. (2013), who look at non-linguistic measures of social engagement in Twitter among their potential depression-related attributes.[5]

4 Conclusions and Future Directions

In this paper we have briefly described the University of Maryland contribution to the CLPsych 2015 shared tasks. We found that TF-IDF features alone

[5] Although only an anecdotal observation involving two rather different datasets, the Depression v Control ROC curve in Figure 1 appears remarkably similar to the ROC curve in De Choudhury et al's Figure 4.

performed very well, perhaps surprisingly well, on all three tasks; TF-IDF combined with supervised topic model posteriors resulted in an even more predictive feature configuration.

In future work, we plan to conduct a thorough error analysis to see where the models go astray. We also plan to look at the extent to which our data organization may have influenced performance; in preliminary experimentation in Resnik et al. (2015), we found suggestive evidence that aggregating tweets by week, rather than as a single document per user, might make a significant difference, and that is the strategy we adopted here. This may not just be a question of document size — other time-based aggregations may be worth exploring, e.g. grouping tweets by time of day.

Acknowledgments

We are grateful to Rebecca Resnik for contributing her comments and clinical expertise, and we thank Glen Coppersmith, Mark Dredze, Jamie Pennebaker, and their colleagues for kindly sharing data and resources. This work was supported in part by NSF awards 1320538, 1018625, and 1211153. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the view of the sponsor.

References

Sanjeev Arora, Rong Ge, Yoni Halpern, David M. Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu. 2013. A practical algorithm for topic modeling with provable guarantees.

David Blei and Jon McAuliffe. 2008. Supervised topic models. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 121–128. MIT Press, Cambridge, MA.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.

Munmun De Choudhury, Scott Counts, Eric Horvitz, and Michael Gamon. 2013. Predicting depression via social media. In *AAAI*. AAAI, July.

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 51–60, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, USA, June. North American Chapter of the Association for Computational Linguistics.

Glen Coppersmith. 2015. [Un]Shared task: Computational linguistics and clinical psychology. `http://glencoppersmith.com/papers/CLPsych2015_hackathon_shared_task.pdf`.

Thang Nguyen, Jordan Boyd-Graber, Jeffrey Lund, Kevin Seppi, and Eric Ringger. 2015. Is your anchor going up or down? Fast and accurate supervised topic models. In *North American Chapter of the Association for Computational Linguistics*.

James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Philip Resnik, Anderson Garron, and Rebecca Resnik. 2013. Using topic modeling to improve prediction of neuroticism and depression in college students. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1348–1353, Seattle, Washington, USA, October. Association for Computational Linguistics.

Philip Resnik, Rebecca Resnik, and Margaret Mitchell, editors. 2014. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Association for Computational Linguistics, Baltimore, Maryland, USA, June.

Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*.

Matthew Tull. 2015. PTSD and Depression. `http://ptsd.about.com/od/relatedconditions/a/depressionPTSD.htm`.

Bill Young, Chris Clark, John Kansky, and Erik Pupo. 2014. Definition: Continuum of care, May. `http://www.himss.org/ResourceLibrary/genResourceDetailPDF.aspx?ItemNumber=30272`.