Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Proc Conf Assoc Comput Linguist Meet. Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:

Proc Conf Assoc Comput Linguist Meet. 2016 August ; 2016: 1044–1053.

Understanding Discourse on Work and Job-Related Well-Being in Public Social Media

Tong Liu[1*], **Christopher M. Homan**[1*], **Cecilia Ovesdotter Alm**[1$], **Ann Marie White**[2], **Megan C. Lytle**[2], and **Henry A. Kautz**[3]

Tong Liu: tl8313@rit.edu; Christopher M. Homan: cmh@cs.rit.edu; Cecilia Ovesdotter Alm: coagla@rit.edu; Ann Marie White: annmarie_white@urmc.rochester.edu; Megan C. Lytle: megan_lytle@urmc.rochester.edu; Henry A. Kautz: kautz@cs.rochester.edu

[1*]Golisano College of Computing and Information Sciences, Rochester Institute of Technology

[1$]College of Liberal Arts, Rochester Institute of Technology

[2]University of Rochester Medical Center

[3]Department of Computer Science, University of Rochester

Abstract

We construct a humans-in-the-loop supervised learning framework that integrates crowdsourcing feedback and local knowledge to detect job-related tweets from individual and business accounts. Using data-driven ethnography, we examine discourse about work by fusing language-based analysis with temporal, geospatial, and labor statistics information.

1 Introduction

Work plays a major role in nearly every facet of our lives. Negative and positive experiences at work places can have significant social and personal impacts. Employment condition is an important social determinant of health. But how exactly do jobs influence our lives, particularly with respect to well-being? Many theories address this question (Archambault and Grudin, 2012; Schaufeli and Bakker, 2004), but they are hard to validate as well-being is influenced by many factors, including geography as well as social and institutional support.

Can computers help us understand the complex relationship between work and well-being? Both are broad concepts that are difficult to capture objectively (for instance, the unemployment rate as a statistic is continually redefined) and thus challenging subjects for computational research.

Our first contribution is to propose a classification framework for such broad concepts as work that alternates between humans-in-the-loop annotation and machine learning over multiple iterations to simultaneously clarify human understanding of these concepts and automatically determine whether or not posts from public social media sites are about work. Our framework balances the effectiveness of crowdsourced workers with local experience, evaluates the degree of subjectivity throughout the process, and uses an iterative post-hoc

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

evaluation method to address the problem of discovering gold standard data. Our performance (on an open-domain problem) demonstrates the value of our humans-in-the-loop approach which may be of special relevance to those interested in discourse understanding, particularly settings characterized by high levels of subjectivity, where integrating human intelligence into active learning processes is essential.

Our second contribution is to use our classifiers to study job-related discourse on social media using data-driven ethnography. Language is fundamentally a social phenomenon, and social media gives us a lens through which to observe a very particular form of discourse in real time. We add depth to the NLP analysis by gathering data from specific geographical regions to study discourse along a broad spectrum of interacting social groups, using work as a framing device, and we fuse language-based analysis with temporal, geospatial and labor statistics dimensions.

2 Background and Related Work

Though not the first study of job-related social media, prior ones used data from large companies' internal sites, whose users were employees (De Choudhury and Counts, 2013; Yardi et al., 2008; Kolari et al., 2007; Brzozowski, 2009). An obvious limitation in that case is it excludes populations without access to such restricted networks. Moreover, workers may not disclose true feelings about their jobs on such sites, since their employers can easily monitor them. On the other hand, we show that on Twitter, it is quite common for tweets to relate negative feelings about work ("*I don't wanna go to work today*"), unprofessional behavior ("*Got drunk as hell last night and still made it to work*"), or a desire to work elsewhere ("*I want to go work at Disney World so bad*").

Nonetheless, these studies inform our work. DeChoudhury et al. (2013) investigated the landscape of emotional expression of the employees via enterprise internal microblogging. Yardi et al. (2008) examined temporal aspects of blogging usage within corporate internal blogging community. Kolari et al. (2007) characterized comprehensively how behaviors expressed in posts impact a company's internal social networks. Brzozowski (2009) described a tool that aggregated shared internal social media which when combined with its enterprise directory added understanding the organization and employees connections.

From a theoretical perspective, the Job Demands-Resources Model (Schaufeli and Bakker, 2004) suggests that job demands (e.g., overworked, dissonance, and conflict) lead to burnout and disengagement while resources (e.g., appreciation, cohesion, and safety) often result in satisfaction and productivity. Although burnout and engagement have an inverse relationship, these states fluctuate and can vary over time. In 2014, more than two-thirds of U.S. workers were disengaged at work (Gallup, 2015a) and this disconnection costs the U.S. up to $398 billion annually in lost work and medical treatment (Gallup, 2015b). Indeed, job dissatisfaction poses serious health risks and has even been linked to suicide (Hazards Magazine, 2014). Thus, examining social media for job-related messages provides a novel opportunity to study job discourse and associated demands and resources. Moreover, the declarative and affective tone of these tweets may have important implications for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

understanding the relationship between burnout and engagement with such public health concerns as mental health.

3 Humans-in-the-Loop Classification

From July 2013 to June 2014 we collected over 7M geo-tagged tweets from around 85,000 public accounts in a 15-county around a midsized city using DataSift[1]. We removed punctuation and special characters, and used the Internet Slang Dictionary[2] to normalize nonstandard terms.

Figure 1 shows our humans-in-the-loop framework for learning classifiers to identify job-related posts. It consists of four rounds of machine classification – similar to that of Li et al. (2014) except that our rounds are not as uniform – where the classifier in each round acts as a filter on our training data, providing human annotators a sample of Twitter data to label and (except for the final round) using these labeled data to train the classifiers in later rounds.

The initial classifier C_0 is a simple term-matching filter; see Table 1 (number options were considered for some terms). The other classifiers (C_1, C_2, C_3) are SVMs that use a feature space of n-grams from the training set.

Round 1

We ran C_0 on our dataset. Approximately 40K tweets having at least five tokens passed this filter. We call them *Job-Likely tweets*. We randomly chose around 2,000 *Job-Likely* tweets and split them evenly into 50 AMT Human Intelligence Tasks (HITs), and further randomly duplicated five tweets in each HIT to evaluate each worker's consistency. Five crowdworkers assigned to each HIT[4] answered, for each tweet, the question: *Is this tweet about job or employment?* All crowdworkers lived in the U.S. and had an approval rating of 90% or better. They were paid $1.00 per HIT[5]. We assessed inter-annotator reliability among the five annotators in each HIT using Geertzen's tool (Geertzen, 2016).

This yielded 1,297 tweets where all 5 annotators agreed on the same label (Table 2). To balance our training data, we added 757 tweets chosen randomly from tweets outside the *Job-Likely* set that we labeled not job-related. C_1 trained on this set.

Round 2

Our goal was to collect 4,000 more labeled tweets that, when combined with the Round 1 training data, would yield a class-balanced set. Using C_1 to perform regression, we ranked the tweets in our dataset by the confidence score (Chang and Lin, 2011). We then spot-checked the tweets to estimate the frequency of job-related tweets as the confidence score increases. We discovered that among the top-ranked tweets about half, and near the

[1] http://datasift.com/
[2] http://www.noslang.com/dictionary
[4] This is based on empirical insights for crowdsourced annotation tasks (Callison-Burch, 2009; Evanini et al., 2010).
[5] We consulted with Turker Nation (http://www.turkernation.com) to ensure that the workers were treated and compensated fairly for their tasks. We also rewarded annotators based on the qualities of their work.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

separating hyperplane (i.e., where the confidence scores are near zero) almost none, are job-related.

Based on these estimates, we randomly sampled 2,400 tweets from those in the top 80th percentile of confidence scores (*Type-1*). We then randomly sampled about 800 tweets each from the first deciles of tweets greater and lesser than zero, respectively (*Type-2*).

The rationale for drawing from these two groups was that the false Type-1 tweets represent those on which the C_1 classifier most egregiously fails, and the Type-2 tweets are those closest to the feature vectors and those toward which the classifier is most sensitive.

Crowdworkers again annotated these tweets in the same fashion as in Round 1 (see Table 3), and cross-round comparisons are in Tables 2 and 4. We trained C_2 on all tweets from Round 1 and 2 with unanimous labels (bold in Table 2).

Round 3

Two coauthors with prior experience from the local community reviewed instances from Round 1 and 2 on which crowd-workers disagreed (highlighted in Table 5) and provided labels. Cohen's kappa agreement was high: $\kappa = 0.80$. Combined with all labeled data from the previous rounds this yielded 2,670 gold-standard-labeled job-related and 3,250 not job-related tweets. We trained C_3 on this entire set. Since it is not strictly class-balanced, we grid-searched on a range of class weights and chose the estimator that optimized F1 score, using 10-fold cross validation[6]. Table 6 shows C_3's top-weighted features, which reflect the semantic field of work for the job-related class.

Discovering Businesses

Manual examination of job-related tweets revealed patterns like: *Panera Bread: Baker – Night (#LOCATION) http://URL #Hospitality #VeteranJob #Job #Jobs #TweetMyJobs*. Nearly all tweets that contained at least one of these hashtags: *#veteranjob*, *#job*, *#jobs*, *#tweetmyjobs*, *#hiring*, *#retail*, *#realestate*, *#hr* also included a URL, which spot-checking revealed nearly always led to a recruitment website (see Table 7). This led to an effective heuristic to separate individual from business accounts only for posts that have first been classified as job-related: if an account had more job-related tweets with any of the above hashtags + URL patterns, we labeled it *business*; otherwise *individual*.

4 Results and Discussion

Crowdsourced Validation

The fundamental difficulty in open-domain classification problems such as this one is there is no gold-standard data to hold out at the beginning of the process. To address this, we adopted a post-hoc evaluation where we took balanced sets of labeled tweets from each classifier (C_0, C_1,C_2 and C_3) and asked AMT workers to label a total of 1,600 samples, taking the majority votes (where at least 3 out of 5 crowdworkers agreed) as reference

[6]These scores were determined respectively using the mean score over the cross-validation folds. The parameter settings that gave the best results on the left out data were a linear kernel with penalty parameter $C = 0.1$ and class weight ratio of 1:1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

labels. Our results (Table 8) show that C_3 performs the best, and significantly better than C_0 and C_1.

Estimating Effective Recall

The two machine-labeled classes in our test data are roughly balanced, which is not the case in real-world scenarios. We estimated the effective recall under the assumption that the error rates in our test samples are representative of the entire dataset. Let y be the total number of the classifier-labeled "positive" elements in the entire dataset and n be the total of "negative" elements. Let y_t be the number of classifier-labeled "positive" tweets in our 1, 600-samples test set and let n_t = 1, 600 - y_t. Then the estimated effective recall.

$$\hat{R}=\frac{y \cdot n_t \cdot R}{y \cdot n_t \cdot R+n \cdot y_t \cdot (1-R)}.$$

Assessing Business Classifier

For Table 8's tweets labeled by $C_0 - C_3$ as job-related, we asked AMT workers: *Is this tweet more likely from a personal or business account?* Table 9 shows that this method was quite accurate.

Our explanation for the strong performance of the business classifier is that the class of job-related tweets is relatively rare, and so by applying the classifier only to job-related tweets we simplify the individual-or-business problem dramatically. Another, perhaps equally effective, simplification is that our tweets are geo-specific and so we automatically filter out business tweets from, e.g., national media.

Generalizability Tests

Can our best model C_3 discover job-related tweets from other geographical regions, even though it was trained on data from one specific region? We repeated the tests above on 400 geo-tagged tweets from Detroit (balanced between job-related and not). Table 10 shows that C_3 and the business classifier generalize well to another region. This suggests the transferability of our humans-in-the-loop classification framework and of heuristic to separate individual from business accounts for tweets classified as job-related.

5 Understanding Job-Related Discourse

Using the job-related tweets – from both individual and business accounts – extracted by C_3 from the July 2013-June 2014 dataset (see Table 12), we conducted the following analyses.

C_3 Versus C_0

The fact that C_3 outperforms C_0 demonstrates our humans-in-the-loop framework is necessary and effective compared to an intuitive term-matching filter. We further examined the messages labeled as job-related by C_3, but not captured by C_0. More than 160,000 tweets fell into this *Difference* set, in which approximately 85,000 tweets are from individual accounts while the rest are from business accounts. Table 11 shows the top 3 most frequent uni-, bi-, and trigrams in the *Difference* dataset. These n-grams from the individual group suggest that people often talk about job-related topics while mentioning temporal

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

information or announcing their working schedules. We neglected such time-related phrases when defining C_0. In contrast, the frequencies of the listed n-grams in the business group are much higher than those in the individual group. This indicates that our definitions of inclusion terms in C_0 did not capture a considerable amount of posts involving broad job-related topics, which is also reflected in Table 9: our business classifier did not find business accounts from the job-related tweets extracted by C_0.

Hashtags

Individuals posted 11,935 unique hashtags and businesses only 414. The top 250 hashtags from each group are shown in Figure 2.

Individual users used an abbreviation for the name of the midsized city to mark their location, and *fml*[7] to express personal embarrassing stories. *Work* and *job* are self-explanatory. *Money*, *motivation* relates to jobs. *Tired*, *exhausted*, *fuck*, *insomnia*, *bored*, *struggle* express negative conditions. Likewise, *lovemyjob*, *happy*, *awesome*, *excited*, *yay*, *tgif*[8] convey positive affects experienced from jobs. Business accounts exhibit distinct patterns. Besides the hashtags queried (Table 7), we saw local place names, like *corning*, *rochester*, *batavia*, *pittsford*, and regional ones like *syracuse*, *ithaca. Customerservice*, *nursing*, *accounting*, *engineering*, *hospitality*, *construction* record occupations, while *kellyjobs*, *familydollar*, *cintasjobs*, *cfgjobs*, *searsjobs* point to business agents. Unlike individual users, businesses do not use hashtags reflecting affective expressions.

Linguistic Differences

We used the TweetNLP POS tagger (Gimpel et al., 2011). Figure 3 shows nine part-of-speech tag[9] frequencies for three subsets of tweets.

Business accounts use NNPs more than individuals, perhaps because they often advertise job openings at specific locations, like *New York*, *Sears*. Individuals use NNPs less frequently and in a more casual way, e.g., *Jojo*, *galactica*, *Valli*. Also, individuals use JJ, NN, NNS, PRP, PRP$, RB, UH, and VB more regularly than business accounts do. Not job-related tweets have similar patterns to job-related ones from individual accounts, suggesting that individual users exhibit analogous language habits regardless of topic.

Temporal Patterns

Our findings that individual users frequently used time-related n-grams (Table 11) prompted us to examine the temporal patterns of job discourse.

Figure 4a suggests that individuals talk about jobs the most in December and January (which also have the most tweets over other topics), and the least in the warmer months. July witnesses the busiest job-related tweeting from business and January the least. The user community is slightly less active in the warmer months, with fewer tweets then.

---

[7]An acronym for *Fuck My Life*.

[8]An acronym for *Thank God It's Friday* to express the joy one feels in knowing that the work week has officially ended and that one has two days off which to enjoy.

[9]JJ – Adjective; NN – Noun (singular or mass); NNS –Noun (plural); NNP – Proper noun (singular); PRP – Personal pronoun; PRP$ – Possessive pronoun; RB – Adverb; UH –Interjection; VB – Verb (base form) (Santorini, 1990).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 4b shows that job-related tweet volumes are higher on weekdays and lower on weekends, following the standard work week. Weekends see fewer business tweets than weekdays do. Sunday is the most – while Friday and Saturday are the least – active days from the not job-related perspective.

Figure 4c shows hourly trends. Job-related tweets from business accounts are most frequent during business hours, peaking at 11, and then taper off. Perhaps professionals are either getting their commercial tasks completed before lunch, or expecting others to check updates during lunch. Individuals post about jobs almost anytime awake and have a similar distribution to non-job-related tweets.

Measuring Affective Changes

We examined positive affect (PA) and negative affect (NA) to measure diurnal changes in public mood (Figures 5 and 6), using two recognized lexicons, in job-related tweets from individual accounts (left), job-related tweets from business accounts (middle), and not job-related tweets (right).

1. **Linguistic Inquiry and Word Count** We used LIWC's *positive emotion* and *negative emotion* to represent PA and NA respectively (Pennebaker et al., 2001) because it is common in behavioral health studies, and used as a standard comparison in referenced work. Figure 5 shows the mean daily trends of PA and NA.[10] Panels 5a and 5b reveal contrasting job-related affective patterns, compared to prior trends from enterprise-wide micro-blog usage (De Choudhury and Counts, 2013), i.e., public social media exhibit gradual increase in PA while internal enterprise network decrease after business. This perhaps confirms our suspicion that people talk about work on public social media differently than on work-based media.

2. **Word-Emotion Association Lexicon** We focused on the words from EmoLex's *positive* and *negative* categories, which represent sentiment polarities (Mohammad and Turney, 2013; Mohammad and Turney, 2010) and calculated the score for each tweet similarly as LIWC. The average daily positive and negative sentiment scores in Figure 6 display patterns analogous to Figure 5.

Labor Statistics

We explored associations between Twitter temporal patterns, affect, and official labor statistics (Figure 8). These monthly statistics[11] include: labor force, employment, unemployment, and unemployment rate. We collected one more year of Twitter data from the same area, and applied C_3 to extract the job-related posts from individual and business accounts (Table 12 summarizes the basic statistics), then defined the following monthwise statistics for our two-year dataset: count of overall/job-individual/job-business/others tweets;

[10]Non-equal y-axes help show peak/valley patterns here and in Figure 6, also motivated by lexicon's unequal sizes.
[11]Published by US Department of Labor, including: Local Area Unemployment Statistics; State and Metro Area Employment, Hours, and Earnings.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

percentage of job-individual/job-business/others tweets in overall tweets; average LIWC PA/NA scores of job-individual/job-business/others tweets[12].

Positive affect expressed in job-related discourse from both individual and business accounts correlate negatively with unemployment and unemployment rate. This is intuitive, as unemployment is generally believed to have a negative impact on individuals' lives. The counts of job-related tweets from individual and not job-related tweets are both positively correlated with unemployment and unemployment rate, suggesting that unemployment may lead to more activities in public social media. This correlation result shows that online textual disclosure themes and behaviors can reflect institutional survey data.

Inside vs. Outside City

We compared tweets occurring within the city boundary to those lying outside (Table 13). The percentages of job-related tweets from individual accounts, either in urban or rural areas, remain relatively even. The proportion of job-related tweets from business accounts decreased sharply from urban to rural locations. This may be because business districts are usually centered in urban areas and individual tweets reflect more complex geospatial distributions.

Job-Life Cycle Model

Based on hand inspection of a large number of job-related tweets and on models of the relationship between work and wellness found in behavioral studies (Archambault and Grudin, 2012; Schaufeli and Bakker, 2004), we tentatively propose a job-life model for job-related discourse from individual accounts (Figure 7). Each state in the model has three dimensions: the *point of view*, the *affect*, and the job-related *activity*, in terms of basic level of employment, expressed in the tweet.

We concatenated together all job-related tweets posted by each individual into a single document and performed latent Dirichlet allocation (LDA) (Blei et al., 2003) on this user-level corpus, using Gensim (Řehůřek and Sojka, 2010). We used 12 topics for the LDA based on the number of affect classes (three) times the number of job-related activities (four). See Table 14.

Topic 0 appears to be about getting ready to start a job, and topic 1 about leaving work permanently or temporarily. Topics 2, 5, 6, 8, and 11 suggest how key affect is for understanding job- related discourse: 2 and 6 lean towards dissatisfaction and 5 toward satisfaction. 11 looks like a mixture. Topic 7 connects to coworkers. Many topics point to the importance of time (including leisure time in topic 4).

6 Conclusion

We used crowdsourcing and local expertise to power a humans-in-the-loop classification framework that iteratively improves identification of public job-related tweets. We separated business accounts from individual in job-related discourse. We also analyzed identified

[12]IND: individual; BIZ: business; pct: %; avg: average.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

tweets integrating temporal, affective, geospatial, and statistical information. While jobs take up enormous amounts of most adults' time, job-related tweets are still rather infrequent. Examining affective changes reveals that PA and NA change independently; low NA appears to indicate the absence of negative feelings, not the presence of positive ones.

Our work is of social importance to working-age adults, especially for those who may struggle with job-related issues. Besides providing insights for discourse and its links to social science, our study could lead to practical applications, such as: aiding policy-makers with macro-level insights on job markets, connecting job-support resources to those in need, and facilitating the development of job recommendation systems.

This work has limitations. We did not study whether providing contextual information in our humans-in-the-loop framework would influence the model performance. This is left for future work. Additionally we recognize that the hash-tag inventory used to discover business accounts from job-related topics might need to change over time, to achieve robust performance in the future. As another point, due to Twitter demographics, we are less likely to observe working seniors.

Acknowledgments

We thank the anonymous reviewers for their helpful comments and suggestions. This work was supported in part by a GCCIS Kodak Endowed Chair Fund Health Information Technology Strategic Initiative Grant and NSF Award #SES-1111016.

References

Altman DG. Inter-Rater Agreement. Practical Statistics for Medical Research. 1991; 5:403–409.

Archambault, Anne; Grudin, Jonathan. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2012. A Longitudinal Study of Facebook, LinkedIn, & Twitter Use; p. 2741-2750.

Blei, David M.; Ng, Andrew Y.; Jordan, Michael I. Latent Dirichlet Allocation. the Journal of Machine Learning Research. 2003; 3:993–1022.

Brzozowski, Michael J. Proceedings of the ACM 2009 International Conference on Supporting Group Work. ACM; 2009. Watercooler: Exploring an Organization through Enterprise Social Media; p. 219-228.

Callison-Burch, Chris. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Vol. 1. Association for Computational Linguistics; 2009. Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk; p. 286-295.

Chang, Chih-Chung; Lin, Chih-Jen. LIB-SVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST). 2011; 2(3):27.

De Choudhury, Munmun; Counts, Scott. Proceedings of the 2013 Conference on Computer Supported Cooperative Work. ACM; 2013. Understanding Affect in the Workplace via Social Media; p. 303-316.

Evanini, Keelan; Higgins, Derrick; Zechner, Klaus. Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Association for Computational Linguistics; 2010. Using Amazon Mechanical Turk for Transcription of Non-Native Speech; p. 53-56.

Gallup. Majority of U.S. Employees not Engaged despite Gains in 2014. 2015a

Gallup. Only 35% of U.S. Managers are Engaged in their Jobs. 2015b

Geertzen, Jeroen. Inter-Rater Agreement with Multiple Raters and Variables. 2016 [17-February-2016] [Online;].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gimpel, Kevin; Schneider, Nathan; O'Connor, Brendan; Das, Dipanjan; Mills, Daniel; Eisenstein, Jacob; Heilman, Michael; Yogatama, Dani; Flanigan, Jeffrey; Smith, Noah A. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers. Vol. 2. Association for Computational Linguistics; 2011. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments; p. 42-47.

Hazards Magazine. Work Suicide. 2014

Kolari, Pranam; Finin, Tim; Lyons, Kelly; Yesha, Yelena; Yaacov, Yesha; Perelgut, Stephen; Hawkins, Jen. On the Structure, Properties and Utility of Internal Corporate Blogs. Growth. 2007; 45000:50000.

Li, Jiwei; Ritter, Alan; Cardie, Claire; Hovy, Eduard H. Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts. EMNLP. 2014:1997–2007.

Mohammad, Saif M.; Turney, Peter D. Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics; 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create An Emotion Lexicon; p. 26-34.

Mohammad, Saif M.; Turney, Peter D. Crowdsourcing a Word-Emotion Association Lexicon. 2013; 29(3):436–465.

Pennebaker, James W.; Francis, Martha E.; Booth, Roger J. Linguistic Inquiry and Word Count: LIWC 2001. Vol. 71. Mahway: Lawrence Erlbaum Associates; 2001. p. 2001

Řehůřek, Radim; Sojka, Petr. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Valletta, Malta: ELRA; 2010 May. Software Framework for Topic Modelling with Large Corpora; p. 45-50.

Santorini, Beatrice. Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision). 1990

Schaufeli, Wilmar B.; Bakker, Arnold B. Job Demands, Job Resources, and Their Relationship with Burnout and Engagement: A Multi-Sample Study. Journal of Organizational Behavior. 2004; 25(3):293–315.

Yardi, Sarita; Golder, Scott; Brzozowski, Mike. The Pulse of the Corporate Blogosphere. Conf Supplement of CSCW 2008. 2008:8–12. Citeseer.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

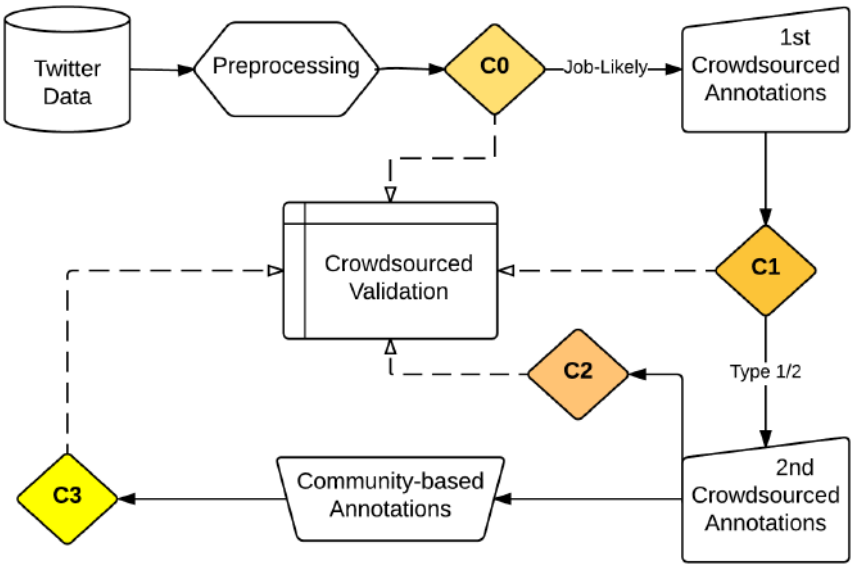


Figure 1.
Flowchart of our humans-in-the-loop framework, laid out in Section 3.

Figure 2.

Hashtags in job-related tweets: above – individual accounts; below – business accounts.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

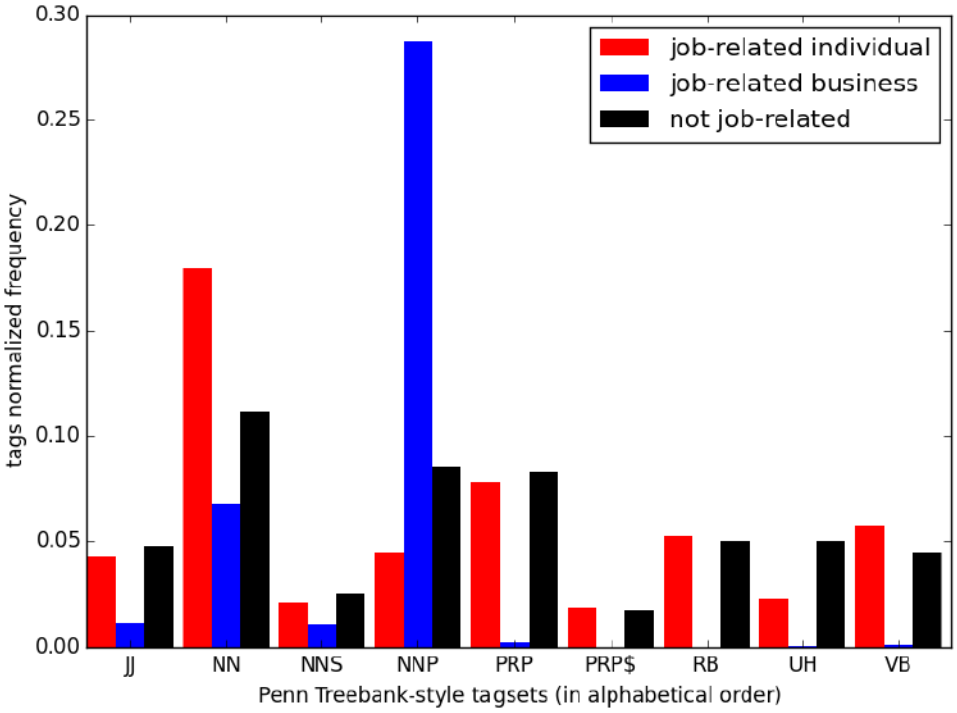
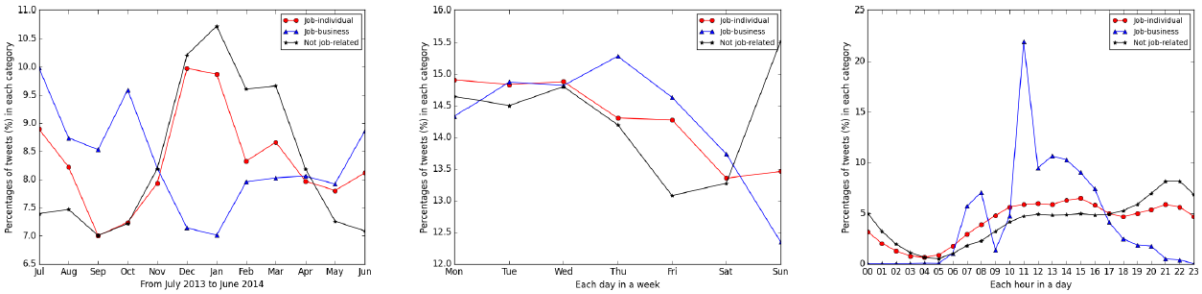


Figure 3.

POS tag comparisons (normalized, averaged) among three subsets of tweets: job-related tweets from individual accounts (red), job-related tweets from business accounts (blue) and not job-related tweets (black).

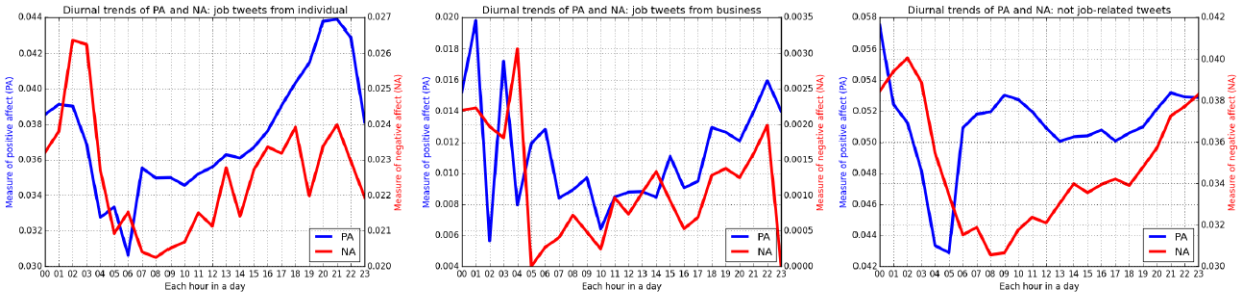


(a) In each month (b) On each day of week (c) In each hour

Figure 4.

Distributions of job-related tweets over time by job class. We converted timestamps from the Coordinated Universal Time standard (UTC) to local time zone with daylight saving time taken into account.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

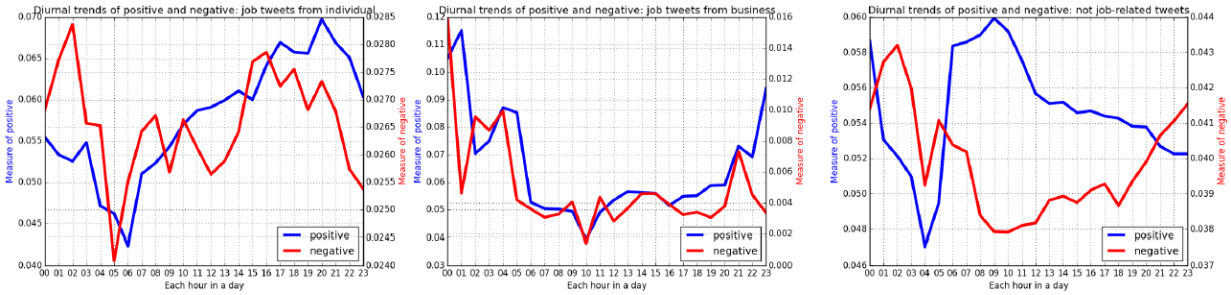


(a) Job-related tweets from individuals (b) Job-related tweets from businesses (c) Not job-related tweets

Figure 5.

Diurnal trends of positive and negative affect based on LIWC.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



(a) Job-related tweets from individuals (b) Job-related tweets from businesses (c) Not job-related tweets

Figure 6.

Diurnal trends of positive and negative affect based on EmoLex.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

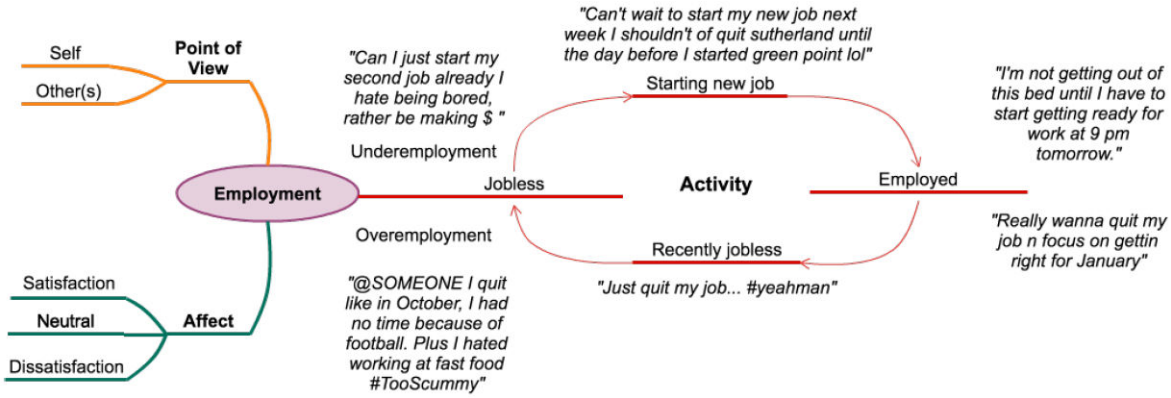


Figure 7.

The job-life model captures the point of view, affect, and job-related activity in tweets.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

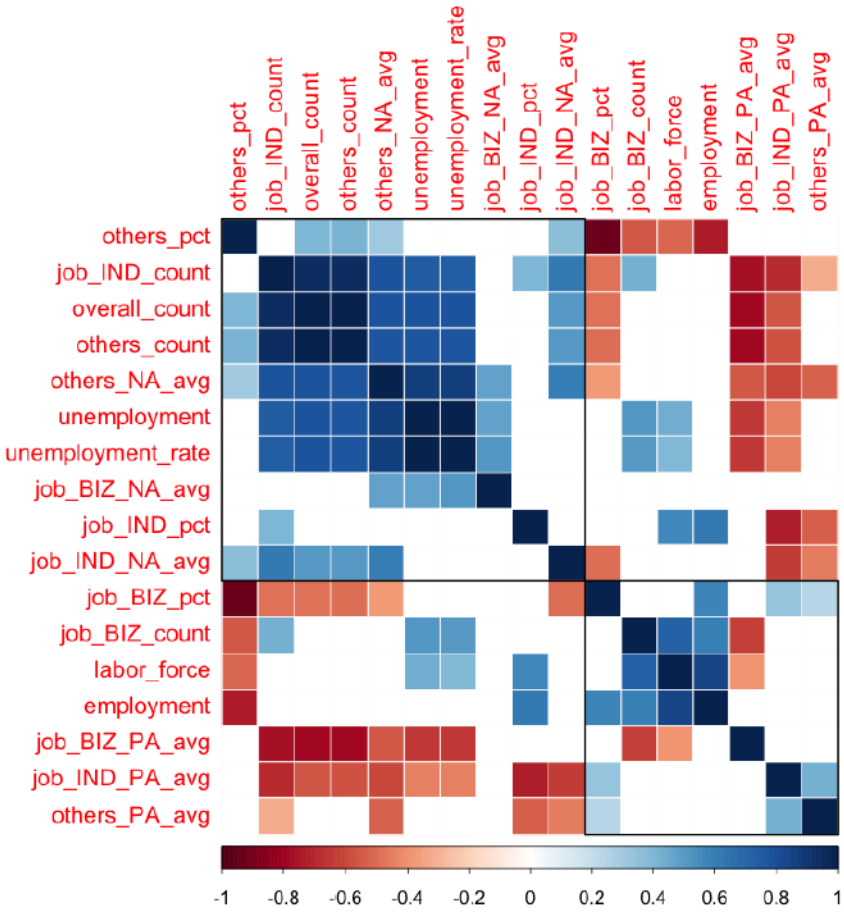


Figure 8.

Correlation matrix with Spearman used for test at level .05, with insignificant coefficients left blank. The matrix is ordered by a hierarchical clustering algorithm. Blue – positive correlation, red – negative correlation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

C_0 rules identifying *Job-Likely* tweets.

Include	job, jobless, manager, boss my/your/his/her/their/at work
Exclude	school, class, homework, student, course finals, good/nice/great job, boss ass[3]

[3]Describe something awesome in a sense of utter dominance, magical superiority, or being ridiculously good.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Summary of both annotation rounds.

AMTs	job-related			not job-related		
	3	4	5	3	4	5
Round 1	104	389	**1027**	78	116	**270**
Round 2	140	287	**721**	66	216	**2568**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Summary of tweet labels in Round 2 by confidence type (showing when 3/4/5 of 5 annotators agreed).

Round 2	job-related			not job-related		
	3	4	5	3	4	5
Type-1	129	280	713	50	149	1079
Type-2	11	7	8	16	67	1489

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Average ± stdev agreement from Round 1 and 2 are *Good*, *Very Good* (Altman, 1991)

AMTs	Fleiss' kappa	Krippendorf's alpha
Round 1	0.62 ± 0.14	0.62 ± 0.14
Round 2	0.81 ± 0.09	0.81 ± 0.08

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Inter-annotator agreement combinations with sample tweets. Y denotes job-related. Cases where the majority (not all) annotators agreed (3/4 out of 5) are underlined in bold.

Annotations	Sample Tweet
Y Y Y Y Y	Really bored….., no entertainment at work today
Y Y Y Y N	two more days of work then I finally get a day off.
Y Y Y N N	Leaving work at 430 and driving in this snow is going to be the death of me
Y Y N N N	Being a mommy is the hardest but most rewarding job a women can have #babyBliss #babybliss
Y N N N N	These refs need to DO THEIR FUCKING JOBS
N N N N N	One of the best Friday nights I've had in a while

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6

Top 15 features for both classes of C_3.

job-related	weights	not job-related	weights
work	2.026	did	-0.714
job	1.930	amazing	-0.613
manager	1.714	nut	-0.600
jobs	1.633	hard	-0.571
managers	1.190	constr	-0.470
working	0.827	phone	-0.403
bosses	0.500	doing	-0.403
lovemyjob	0.500	since	-0.373
shift	0.487	brdg	-0.363
worked	0.410	play	-0.348
paid	0.374	its	-0.337
worries	0.369	think	-0.330
boss	0.369	thru	-0.329
seriously	0.368	hand	-0.321
money	0.319	awesome	-0.319

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7

Counts of hashtags queried, and counts of their subsets with hashtags coupled with URL.

	hashtag only	hashtag + URL
#veteranjob	18,066	18,066
#job	79,362	79,327
#jobs	58,637	58,631
#tweetmyjobs	39,007	39,007
#hiring	148	147
#retail	17,037	17,035
#realestate	92	92
#hr	400	399

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 8

Crowdsourced validations of instances identified by 4 distinct models (1,600 total tweets).

Model	Class	P	R	$\hat{R}$	F1
C_0	Job	0.72	0.33	0.01	0.45
	notjob	0.68	0.92	1.00	0.78
C_1	Job	0.79	0.82	0.15	0.80
	notjob	0.88	0.86	0.99	0.87
C_2	Job	0.82	0.95	0.41	0.88
	notjob	0.97	0.86	0.99	0.91
C_3	Job	0.83	0.96	0.45	0.89
	notjob	0.97	0.87	0.99	0.92

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 9

Crowdsourced validations of individuals vs. businesses job-related tweets.

From	Class	P	R	F1
C_0	Individual	0.86	1.00	0.92
	business	0.00	0.00	0.00
	avg/total	*0.74*	*0.86*	*0.79*
C_1	Individual	1.00	0.97	0.98
	business	0.98	1.00	0.99
	avg/total	*0.99*	*0.99*	*0.99*
C_2	Individual	1.00	0.98	0.99
	business	0.98	1.00	0.99
	avg/total	*0.99*	*0.99*	*0.99*
C_3	Individual	1.00	0.99	0.99
	business	0.99	1.00	0.99
	avg/total	*0.99*	*0.99*	*0.99*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 10

Validations of C_3 and business classifier on Detroit data.

Model	Class	P	R	F1
C_3	Job	0.85	0.99	0.92
	notjob	0.99	0.87	0.93
Heuristic	Individual	1.00	0.96	0.98
	business	0.96	1.00	0.98
	avg/total	*0.98*	*0.98*	*0.98*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 11

Top 3 most frequent uni-, bi-, and trigrams with frequencies in the *Difference* set.

Individual	Business
Unigrams	
day, 6989	ny, 83907
today, 5370	#job, 75178
good, 4245	#jobs, 55105
Bigrams	
last night, 359	#jobs #tweetmyjobs, 32165
getting ready, 354	#rochester ny, 22101
first day, 296	#job #jobs, 16923
Trigrams	
working hour shift, 51	#job #jobs #tweetmyjobs, 12004
first day back, 48	ny #jobs #tweetmyjobs, 4955
separate leader follower, 44	ny #retail #job, 4704

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 12

Summary statistics of the two-year Twitter data classified by C_3.

Unique counts	job-related tweets from individual accounts		job-related tweets from business accounts		not job-related tweets	
	tweets	accounts	tweets	accounts	tweets	accounts
July 2013 - June 2014	114,302	17,183	79,721	292	6,912,306	84,718
July 2014 - June 2015	85,851	16,350	115,302	333	5,486,943	98,716
Total (unique counts)	200,153	28,161	195,023	431	12,399,249	136,703

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 13

Percent inside and outside city tweets.

%	job-related individual	job-related business	others
Inside	1.59	3.73	94.68
Outside	1.85	1.51	96.65
Combined	1.82	1.77	96.41

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 14

The top five words in each of the twelve topics discovered by LDA.

Topic index	Representative words
0	getting, ready, day, first, hopefully
1	last, finally, week, break, last day
2	fucking, hate, seriously, lol, really
3	come visit, some, talking, pissed
4	weekend, today, home, thank god
5	wish, love, better, money, working
6	shift, morning, leave, shit, bored
7	manager, guy, girl, watch, keep
8	feel, sure, supposed, help, miss
9	much, early, long, coffee, care
10	time, still, hour, interview, since
11	best, pay, bored, suck, proud