Identifying Purchase Intent from Social Posts

Vineet Gupta
Adobe Research India Labs, Bangalore
vineetgupta10@gmail.com

Devesh Varshney and **Harsh Jhamtani**
Indian Institute of Technology, Roorkee
devaruec@iitr.ac.in and harshjhamtani@gmail.com

Deepam Kedia
Indian Institute of Technology, Kanpur
deepamkedia@gmail.com

Shweta Karwa
Indian Institute of Technology, Delhi
shwetakarwa@gmail.com

Abstract

In present times, social forums such as Quora[1] and Yahoo! Answers[2] constitute powerful media through which people discuss on a variety of topics and express their intentions and thoughts. Here they often reveal their potential intent to purchase - 'Purchase Intent' (PI). A purchase intent is defined as a text expression showing a desire to purchase a product or a service in future. Extracting posts having PI from a user's social posts gives huge opportunities towards web personalization, targeted marketing and improving community observing systems. In this paper, we explore the novel problem of detecting PIs from social posts and classifying them. We find that using linguistic features along with statistical features of PI expressions achieves a significant improvement in PI classification over 'bag-of-words' based features used in many present day social-media classification tasks. Our approach takes into consideration the specifics of social posts like limited contextual information, incorrect grammar, language ambiguities, etc. by extracting features at two different levels of text granularity - word and phrase based features and grammatical dependency based features. Apart from these, the patterns observed in PI posts help us to identify some specific features.

Introduction

With the advent of social media, online presence of people has increased significantly. People often augment their online information search with feedbacks from specialized 'Questions and Answers' (Q&A) sites like *Quora* - an emergent social-network based Q&A site (Paul, Hong, and Chi 2012). Here, users make posts on a range of topics and have discussions. Similarly, *Yahoo! Answers* (YA) is a large and diverse Q&A forum (Adamic et al. 2008), acting not only as a medium for sharing technical knowledge, but also as a place where one can seek advice and gather opinions. Some of these posts may reveal their potential purchase intent towards something (e.g. "What phone would you recommend me to buy?"). Identifying an intention of purchase from these posts is useful, and at the same time challenging. It is useful for enhancing community observing systems using social data (Miluzzo et al. 2008), for improving the knowledge gathering process of context-aware systems (Banerjee et al. 2009; Sakaki, Okazaki, and Matsuo 2010) and for targeted marketing solutions.

In concept, we define a Purchase Intention *(PI)* as a text expression signifying an intention to purchase or consume a product or a service. Any post not having such an expression belongs to *non-PI* category.

Specifically, our goal is to solve the two-class classification task - is the post a PI or a non-PI? Therefore, our focus is on extracting important features having significant discriminating power to enable an effective classification of PIs from user-posts. There are several characteristics of short-texts which pose a challenge to this task:

- *Informal language:* Language used in social media is often informal and lacks grammatical structure. Usage of acronyms and miss-spellings is also common. These factors not only make the semantic analysis of the text difficult, but also make the data high dimensional in nature.
- *Limited contextual information:* The short size of posts in social media (e.g. 140 characters limit on Twitter; our dataset had an average of 73 characters per post) makes it difficult for the traditional knowledge extraction algorithms to work effectively (Han, Pei, and Yin 2000). It is fairly difficult to gather contextual information from such short texts.
- *Language processing based ambiguities:* Multiple possibilities in syntactic parsing of texts, ambiguities in resolution of anaphoric expressions and other linguistic ambiguities make it difficult for natural language processing based feature extraction systems to work effectively (Zukerman and Litman 2001).

In this paper, we take into consideration these characteristics of short social-posts to design an efficient classification model for detecting PIs. This is done by extracting features at two different levels of text granularity - word and phrase based features and grammatical dependency based features. This technique is based on the fact that even though these short posts lack grammatical structure, but at a sub-sentence level, users tend to arrange words in correct order (Banerjee et al. 2012). Apart from these, the patterns observed in PI posts help us to identify some specific features.

[1]www.quora.com

[2]www.answers.yahoo.com

The rest of this paper is organized as follows - We review related work on user intent discovery and wish identification in Section 2. In Section 3, we explain our data collection and annotation process, followed by the extraction of various features. In Section 4, we present the experiments and their results. Finally, Section 5 concludes the paper and provides the scope of future work.

Related Work

There have been several works on identification of wishes from texts (Goldberg et al. 2009; Ramanand, Bhavsar, and Pedanekar 2010). Specifically, Ramanand et al. (Ramanand, Bhavsar, and Pedanekar 2010) consider the task of identifying 'buy' wishes from product reviews. These wishes include suggestions for a product or a desire to buy a product. They used linguistic rules to detect these two kinds of wishes. Although rule-based approaches for identifying the wishes are effective, but their coverage is not satisfactory and they can't be extended easily. PI detection task is close to the task of identifying wishes in product reviews. Here we don't use the rule based approach, but we present a machine learning approach with generic features extracted from the posts.

Works on sentiment analysis and mining consumer opinions have primarily focused on the problem of identifying positive or negative sentiments in posts containing brand specific comments (Pang and Lee 2008; Gupta 2011). Our classification task attempts to detect what people want to purchase, and not what they like or dislike. Further, we need to handle a wider range of expressions than just positive and negative sentiment indicative terms for capturing the intention of purchase.

Another related area is 'online commercial intent' detection in query logs (Dai et al. 2006). Here the aim is to detect intentions of purchasing, selling, auctioning, etc.. Most of the works in this direction have focused on estimation of click-through rate of an advertisement for a given query (Richardson, Dominowska, and Ragno 2007; Regelson and Fain 2006). Our task is directed towards social forums where click-through data is not available.

One more approach to predict user intention is by analyzing the user-interaction data (Guo and Agichtein 2010). Here, a method to detect whether a query has *research intent* or *purchase intent* based on user's behavior on a search engine result page *(SERP)* is presented. Here the major focus is to make use of the features like *SERP* content, clicks and mouse trajectory, but our task is focused on using the textual content of user posts on social platforms.

Proposed Approach

In this section, we describe the details of our approach to tackle the problem of PI detection. This includes the description of our data collection and annotation process. This is followed by our approach for feature selection and extraction.

Data collection and annotation

As there are no annotated Quora or Yahoo! answers (YA) corpora available publicly for detection of purchase intent, we created our own. We collected over 30 thousand publicly available query posts from Quora and over 12 thousand publicly available query posts from YA for our study and experiments. This was done using a web crawler developed by us which crawled the websites to collect the data. After removing the duplicates, we had a total of 28,267 posts from Quora and 11,354 posts from YA. Out of these 15,000 posts from Quora were randomly selected for training and testing the model and 7000 posts from YA were randomly selected for model validation on a different platform.

For the labeling procedure, we defined the following concepts related to *Purchase Intent* (PI):

Consumption indicative words (CI words):- *"These are the keywords which refer to one or more products or services which can be purchased or consumed. They are generally the central subject in a PI text"*. Such words are usually proper nouns or common nouns (e.g. cellphone, lunch), but not limited to these parts of speech.

Action indicative words (AI words):- *"These are the keywords that describe the action associated with a particular CI word."* These are typically the verbs in a sentence that characterize the activity performed (e.g. buy, eat). Here also, AI words are not limited to these parts of speech.

Purchase Intent (PI):- *"A text expression containing* ***one or more CI words*** *which provide the object of intent along with* ***one or more AI words*** *which further indicate an intent of consumption."*

This way, we give a generic definition of PIs and at the same time avoid ambiguous expressions like *"I am excited about trekking"*.

To generate the ground truth, we use Amazon Mechanical Turk (AMT) to recruit human annotators. AMT is a crowdsourcing internet marketplace that allows one to recruit workers to perform human intelligence tasks (HITs) for small payments. It has been widely used in survey and experimental research (Kittur, Chi, and Suh 2008; Mason and Watts 2010) and to get human annotations for data (Sheng, Provost, and Ipeirotis 2008; Snow et al. 2008). We provide the workers (annotators) with aforesaid definitions and a number of examples of PI and non-PI posts to help them in understanding the task. The annotators were asked to read the posts and annotate them as PI or non-PI posts. To ensure a reliable classification, we had each post annotated by 5 different annotators and we used the majority voting scheme to come up with final labels for the posts.

The mean Cohen's kappa coefficient of inter-annotator agreement between the sets of annotations was 0.712, indicating a strong agreement (Landis and Koch 1977). Using this process, we obtained 2,597 positives (PIs) and 12,403 negatives (Non-PIs) from Quora corpus. Similarly, from YA, 1,139 positives and 5,861 negatives were obtained. This further shows the presence of class imbalance in the data with approximately 17% posts belonging to PI class.

Feature extraction

Here, we describe our approach of selecting features for designing an efficient PI detection and classification model. We note that a PI post has two major components - presence of Consumption Intent and presence of corresponding Consumable Object. Hence, we start with extracting features related to these components and then extract other relevant features.

Purchase Action words (PA words) - In an English sentence, verbs describe the action performed by the subject. Thus, they contain information related to 'purchase action' (e.g. "I wish to ***buy*** a nice camera."). In this paper, we use the syntactic parse tree to analyze this component with the help of *PCFG* parser in *StanfordNLP*[3] tool (Klein and Manning 2003). The *Purchase Action* words (PA words) and *non-Purchase Action* words (non-PA words) are defined on the basis of their relative frequency of occurrence in PI class versus Non-PI class using a *pointwise mutual information (PMI)* based heuristic. This has been shown to be effective in extracting important features for classification tasks (Schneider 2005):

$$Score_{PI}(v) = \log_2 (\frac{n_{PI}(v)}{n(v) * N_{PI}}) \quad (1)$$

Here $n_{PI}(v)$ and $n(v)$ are presence counts of the verb v in posts belonging to PI class and in entire data respectively, and N_{PI} is the number of posts in PI class. Here, those words which are important for PI class are scored higher than those which are not. A similar $Score_{non-PI}$ was computed for each verb using $n_{non-PI}(v)$ and N_{non-PI} values. Those verbs whose $Score_{PI}$ values were above a threshold value were chosen as PA words; non-PA words were chosen using the same process in $Score_{non-PI}$ values. The threshold values for selecting PA and non-PA words were computed using *Otsu's thresholding method* (Otsu 1975). This is a clustering based thresholding method which calculates the optimum threshold separating two classes so that their intra-class variance is minimal. The number of PA and non-PA words thus obtained were 132 and 467 respectively. Table 1 shows top-10 PA and non-PA words, the ranking was done on the basis of their $Score_{PI}$ and $Score_{non-PI}$ values respectively. We can see that strong purchase indicative terms like *buy, hire* and *recommend* appear in the PA word list and terms like *facing, betray* and *reproduce* appear in non-PA word list.

For extracting features related to *Purchase Action* from a post *p*, we first parse the post and identify verbs in it. This gives us a set of verbs V_p in the post p. We extract two real-valued features - W_{PA} and W_{NPA} representing semantic closeness of V_p with PA word list and non-PA word list respectively. This was achieved with the help of *WordNet*[4]

[3]nlp.stanford.edu

[4]http://wordnet.princeton.edu

Group	Top-10 words
PA words	buy, recommend, hire, have, suggest, advise, want, need, purchase, wish
non-PA words	came, facing, disappear, joking, betray, share, fill, allow, shoot, reproduce

Table 1: Top-10 Purchase Action(PA) and non-Purchase Action(non-PA) words

(Miller et al. 1990). It groups English words into sets of synonyms called *synsets* in a 'is-a' hierarchy (for instance *car* 'is-a' *automobile*), thereby forming a *Wordnet* graph. The similarity between two *synsets* is measured based on the path length between them in this graph using the following approach:

WP-Similarity (Wu and Palmer 1994) :- Here, the depth of 'Lowest Common Subsumer' of both the *synsets* is found, which is then scaled by the sum of depths of each *synset*. Depth of a *synset* is simply its distance from a hypothetical root node in this graph. Here, a random node was chosen as root node. If nodes s_1 and s_2 are two *synsets* and node lcs is their lowest common subsumer, then:

$$\text{WP-similarity} = 2 * \frac{depth(lcs)}{(depth(s_1) + depth(s_2))} \quad (2)$$

We compute WP-similarity of synset of each verb in V_p with synset of each word in PA word list. W_{PA} is assigned the maximum of these similarity values with PA words. The maximum value is chosen so as to perfectly capture the presence of any verb depicting purchase action. The value of W_{NPA} is computed in a similar way.

Purchase Object categories (PO categories) - A purchase intent is always expressed towards an object (e.g. a product, a service) which can be purchased or consumed. Therefore, identifying whether the object of conversation in a post is consumable or not is important for classification of PIs. In a post, a consumable object is described using a noun phrase (NP) (e.g. "What ***phone*** would you recommend me to buy?"). Extracting NPs just from the annotated data gives a narrow set of objects. Hence, to capture a more generic feature, we find the categories to which these objects belong. This was found to give better results than using just the NPs. We determine the *categories* of the NPs obtained by syntactic parsing of text, to find *Purchase Object and non-Purchase Object* categories. For this, we take the help of *Freebase* (Google 2013; Bollacker et al. 2008) which is a collaboratively created structured database of general human knowledge. *Freebase* associates objects present in its database with topics, and these topics serve as object categories for our study (e.g.: *iPad* has *'Consumer product'* and *'Computer'* as its categories).

We use a similar approach that we used for generating PA and non-PA word lists to obtain Purchase Object categories (PO categories) and non-Purchase categories (non-PO categories). Table 2 shows top-3 PO and non-PO categories, and some examples of words associated with them in the data.

Group	Top Categories	Words
PO categories	Consumer product	iPad, sunglasses
	Food	pizza, dinner
	Hospitality business	holiday, hotel, vacation
non-PO categories	Unit of time	hour, year
	Religion	Pope, Islam
	Military Conflict	war, combat

Table 2: Top-3 Purchase Object(PO) and non-Purchase Object(non-PO) categories and words from labeled data associated with them

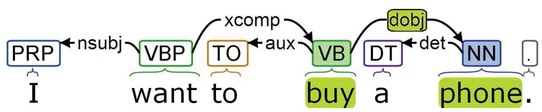

Figure 1: An illustration of dependency based ADO feature

For extracting features related to *Purchase Object* from a post p, we first parse the post and identify the NPs in it. The categories to which each of these NPs belong are identified using *Freebase*. This gives us a set of categories C_p. We extract two binary features - C_{PO} and C_{NPO}. If any category in C_p is a PO category, $C_{PO} = 1$, otherwise $C_{PO} = 0$. Similarly, if any category in C_p is a non-PO category, $C_{NPO} = 1$, otherwise $C_{NPO} = 0$.

After having discovered the PA and non-PA words and the PO and non-PO categories and having extracted the relevant features, we consider the features which relate them using linguistic dependencies.

PA word has dependent object of PO category (ADO) - In a PI post, a purchase action is targeted towards a consumable object. This is reflected in the dependency structure of the text and we extract this information by the help of *dependency parser* provided by *StanfordNLP* (De Marneffe et al. 2006). In a PI post, the consumable object is usually the directly dependent object of the purchase action verb.

We extract a binary feature ADO to capture this aspect. If there is a PA word in the text and it has a dependent object belonging to a PO category, $ADO = 1$, otherwise $ADO = 0$. For instance, in the post *"I want to buy a phone."*, the verb "buy" which belongs to the PA word list has "phone" which belongs to a PO category as its directly dependent object, so here $ADO = 1$. This is illustrated in Figure 1. This feature not only takes into account presence of PA word and PO category word in a text, but also the relationship between them.

Purchase supportive word dependencies (PSD) - Purchase supportive words are the keywords that provide knowledge about a particular *Purchase Object* or *Purchase Action*. They are usually the names of locations or organizations related with the *Purchase Object* or *Purchase Action* and this relation is captured in preposition based dependencies in the text (e.g. "Can you please recommend

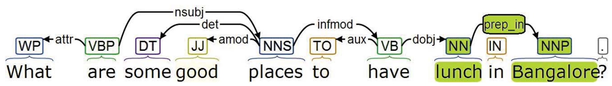

Figure 2: An illustration of dependency based PSD feature

some good place to have lunch ***in Bangalore***?"). We extract this information with the help of *dependency parser* provided by *StanfordNLP* (De Marneffe et al. 2006). We use *Named Entity Recognition* tool of *Natural Language Toolkit (NLTK)* (Bird, Klein, and Loper 2009) which recognizes names of organizations and locations in the text. We extract two binary features related to PSD - ORG and LOC. If there is an object belonging to PO category that has a prepositionally dependent organization then $ORG = 1$ else $ORG = 0$. Similarly, if there is an object belonging to PO category that has a prepositionally dependent location then $LOC = 1$ else $LOC = 0$. This is illustrated in Figure 2.

Delta TFIDF - Delta TFIDF is a technique to efficiently weigh word scores for classification tasks. It has been shown to be more effective in binary classification of class-imbalanced data using unigrams, bigrams and trigrams (Martineau et al. 2009). In our data, only 17% of the annotated posts were labeled as PI. Hence we used Delta TFIDF features for unigrams, bigrams and trigrams in the annotated data. For any term t in post p, the Delta TFIDF score $V_{t,p}$ is computed as follows:

$$V_{t,p} = n_{t,p} * \log_2(\frac{N_t}{P_t}) \tag{3}$$

Here $n_{t,p}$ is the frequency count of term t in post p. P_t and N_t are the number of posts in the PI labeled data set with term t and the non-PI labeled data set with term t respectively.

Sentiment Score - Sentiment refers to the polarity of a given text at the document, sentence or feature level (Pang and Lee 2008). It was observed that the posts having purchase intent often contained positive sentiment words like *good, nice,* etc. (e.g. "What are some ***good*** places to have lunch in Palo Alto?"). We used *AlchemyAPI*[5] (Orchestr 2009) to extract the sentiment for each post in our corpora. The sentiment value given by *AlchemyAPI* is in range $[-1, +1]$, -1 being the maximum value of negative sentiment and $+1$ being the maximum value of positive sentiment. Figure 3 shows the sentiment distribution density of annotated posts from PI and non-PI classes. The posts from PI class have a mean sentiment value of 0.124 and a median value of 0.126, while the posts from non-PI class have a mean sentiment value of 0.026 and a median value of 0.000. Thus, we used the sentiment score provided by *AlchemyAPI* as a real value feature.

[5]www.alchemyapi.com

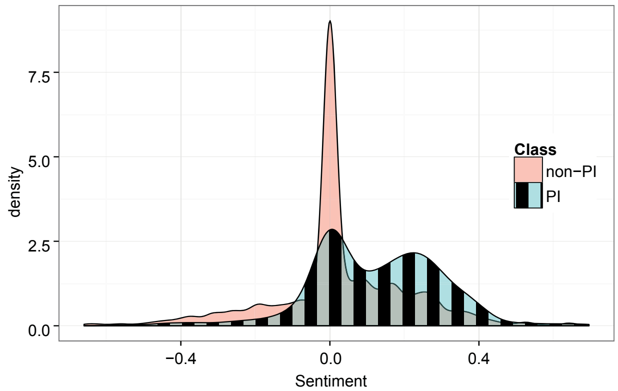

Figure 3: Distribution density plot of sentiments in PI and non-PI classes

Experiments and Results

For our experiments, we use SVM classifier provided by *LIBSVM*[6] toolkit with linear kernel to detect whether the post belongs to PI class or non-PI class (Chang and Lin 2011). As we described before, there is a class imbalance in the data with only 17% posts belonging to PI class. Thus, to better evaluate the performance of the classifier we use the area under the curve (AUC) of a receiver operating characteristic (ROC) curve as a measure of accuracy. The ROC curve has "True positive rate (sensitivity)" along y-axis and "False positive rate" along x-axis. Larger the area under the ROC curve, better is the classifier performance. We use all 15,000 annotated Quora posts, out of which 2,597 are positives (PIs) and 12,403 are negatives (Non-PIs), and all 7,000 annotated Yahoo! Answers posts, out of which 1,139 are positives and 5,861 are negatives for our experiments. We use a 10-fold cross validation process for performance evaluation for Quora dataset. For Yahoo! Answers dataset, which serves as a validation set, we use the model trained on Quora dataset for performance evaluation.

We perform three experiments using different sets of features and evaluate the incremental performance improvement on Quora dataset. Figure 4 shows the ROC curve for these three feature combinations. We observe that:

- Delta TFIDF features are important as they have a significant discriminating power of distinguishing PI posts from non-PI posts. They give an AUC of 0.79 in the ROC curve.
- Including Purchase Actions (PA) based features and Purchase Object Category (PO) based features ($W_{PA} + W_{NPA} + C_{PO} + C_{NPO}$), which are extracted from words and phrases present in the text, gives a significant performance improvement. The AUC improved from 0.79 to 0.86.
- Adding dependency based features and other features (i.e. ADO + Sentiment Score + PSD) gives further improvement in performance. The AUC improved from 0.86 to 0.93.

[6]www.csie.ntu.edu.tw/˜cjlin/libsvm

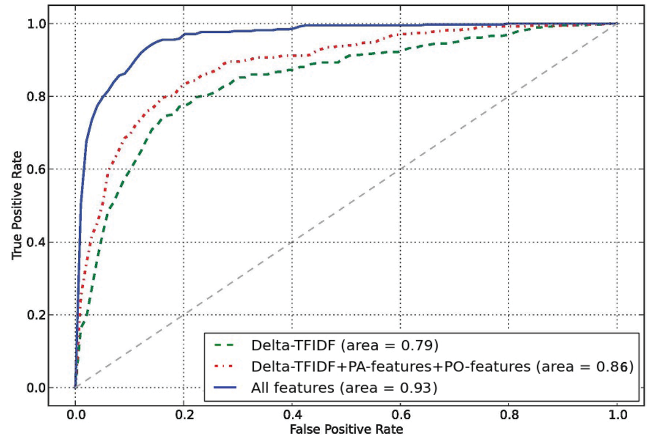

Figure 4: The receiver operating characteristic (ROC) curve of PI classifier for different feature combinations on Quora dataset

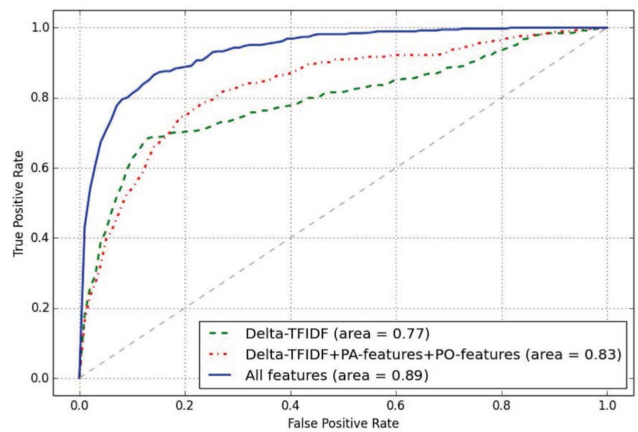

Figure 5: The receiver operating characteristic (ROC) curve of PI classifier for different feature combinations on Yahoo! Answers dataset

Similar experiments were performed on Yahoo! Answers dataset using the model trained on Quora dataset. Figure 5 shows the ROC curve for these three feature combinations. Following results were obtained:

- Delta TFIDF features have a significant discriminating power of distinguishing PI posts from non-PI posts. They give an AUC of 0.77 in the ROC curve.
- Including Purchase Actions (PA) based features and Purchase Object Category (PO) based features ($W_{PA} + W_{NPA} + C_{PO} + C_{NPO}$), gives a significant performance improvement. The AUC improved from 0.77 to 0.83.
- Adding dependency based features and other features (i.e. ADO + Sentiment Score + PSD) gives further improvement in performance. The AUC improved from 0.83 to 0.89.

The above results show that using Purchase Intent is captured well by using the features described in this paper. Best performance on both Quora dataset (AUC=0.93) and Yahoo! Answers dataset (AUC=0.89) is achieved by using all of these features. It can be seen that the model built using Quora dataset gives decent results on Yahoo! Answers dataset, demonstrating that the feature selection is generic and applicable to diverse domains.

Conclusion

This paper studies a novel problem of identifying purchase intent (PI) in user text-posts. We carry out a deep analysis of the structure and content of posts showing purchase intent and present a feature extraction method that captures them effectively. We then train a classifier using these features to classify each post into PI or non-PI category. We believe that our work can provide important insights to applications focusing on exploiting free-text intentions from social media.

Our future research would focus on understanding and using more specific characteristics of social media like *friend network* to make better utilization of available information.

References

Adamic, L. A.; Zhang, J.; Bakshy, E.; and Ackerman, M. S. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web*, 665–674. ACM.

Banerjee, N.; Chakraborty, D.; Dasgupta, K.; Mittal, S.; Nagar, S.; et al. 2009. Ru-in?-exploiting rich presence and converged communications for next-generation activity-oriented social networking. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*, 222–231. IEEE.

Banerjee, N.; Chakraborty, D.; Joshi, A.; Mittal, S.; Rai, A.; and Ravindran, B. 2012. Towards analyzing micro-blogs for detection and classification of real-time intentions. In *ICWSM*.

Bird, S.; Klein, E.; and Loper, E. 2009. *Natural language processing with Python*. O'reilly.

Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 1247–1250. ACM.

Chang, C.-C., and Lin, C.-J. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3):27.

Dai, H. K.; Zhao, L.; Nie, Z.; Wen, J.-R.; Wang, L.; and Li, Y. 2006. Detecting online commercial intention (oci). In *Proceedings of the 15th international conference on World Wide Web*, 829–837. ACM.

De Marneffe, M.-C.; MacCartney, B.; Manning, C. D.; et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, 449–454.

Goldberg, A. B.; Fillmore, N.; Andrzejewski, D.; Xu, Z.; Gibson, B.; and Zhu, X. 2009. May all your wishes come true: a study of wishes and how to recognize them. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, 263–271. Stroudsburg, PA, USA: Association for Computational Linguistics.

Google. 2013. Freebase data dumps. https://developers.google.com/freebase/data.

Guo, Q., and Agichtein, E. 2010. Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 130–137. ACM.

Gupta, N. K. 2011. Extracting descriptions of problems with product and services from twitter data. In *Proceedings of the 3rd Workshop on Social Web Search and Mining (SWSM2011) in conjunction with SIGIR 2011*. The 3rd Workshop on Social Web Search and Mining (SWSM2011) in conjunction with SIGIR 2011, July 24.

Han, J.; Pei, J.; and Yin, Y. 2000. Mining frequent patterns without candidate generation. In *ACM SIGMOD Record*, volume 29, 1–12. ACM.

Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 453–456. ACM.

Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, 423–430. Association for Computational Linguistics.

Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *biometrics* 159–174.

Martineau, J.; Finin, T.; Joshi, A.; and Patel, S. 2009. Improving binary classification on text problems using differential word features. In *Proceedings of the 18th ACM conference on Information and knowledge management*, 2019–2024. ACM.

Mason, W., and Watts, D. J. 2010. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter* 11(2):100–108.

Miller, G. A.; Beckwith, R.; Fellbaum, C.; Gross, D.; and Miller, K. J. 1990. Introduction to wordnet: An on-line lexical database*. *International journal of lexicography* 3(4):235–244.

Miluzzo, E.; Lane, N. D.; Fodor, K.; Peterson, R.; Lu, H.; Musolesi, M.; Eisenman, S. B.; Zheng, X.; and Campbell, A. T. 2008. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, 337–350. ACM.

Orchestr, L. 2009. Alchemyapi.

Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica* 11(285-296):23–27.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2(1-2):1–135.

Paul, S. A.; Hong, L.; and Chi, E. H. 2012. Who is authoritative? understanding reputation mechanisms in quora. *arXiv preprint arXiv:1204.3724*.

Ramanand, J.; Bhavsar, K.; and Pedanekar, N. 2010. Wishful thinking: finding suggestions and'buy'wishes from product reviews. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 54–61. Association for Computational Linguistics.

Regelson, M., and Fain, D. 2006. Predicting click-through rate using keyword clusters. In *Proceedings of the Second Workshop on Sponsored Search Auctions*, volume 9623.

Richardson, M.; Dominowska, E.; and Ragno, R. 2007. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, 521–530. ACM.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, 851–860. ACM.

Schneider, K.-M. 2005. Weighted average pointwise mutual information for feature selection in text categorization. In *Knowledge Discovery in Databases: PKDD 2005*. Springer. 252–263.

Sheng, V. S.; Provost, F.; and Ipeirotis, P. G. 2008. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 614–622. ACM.

Snow, R.; O'Connor, B.; Jurafsky, D.; and Ng, A. Y. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*, 254–263. Association for Computational Linguistics.

Wu, Z., and Palmer, M. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, 133–138. Association for Computational Linguistics.

Zukerman, I., and Litman, D. 2001. Natural language processing and user modeling: Synergies and limitations. *User Modeling and User-Adapted Interaction* 11(1-2):129–158.