Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Target-Dependent Churn Classification in Microblogs

Hadi Amiri and **Hal Daume III**
Computational Linguistics and Information Processing (CLIP) Lab
Institute for Advanced Computer Studies
University of Maryland
{hadi,hal}@umiacs.umd.edu

Abstract

We consider the problem of classifying micro-posts as churny or non-churny with respect to a given brand. Using Twitter data about three brands, we find that standard machine learning techniques clearly outperform keyword based approaches. However, the three machine learning techniques we employed (linear classification, support vector machines, and logistic regression) do not perform as well on churn classification as on other text classification problems. We investigate *demographic*, *content*, and *context* churn indicators in microblogs and examine factors that make this problem more challenging. Experimental results show an average F1 performance of 75% for target-dependent churn classification in microblogs.

Introduction

Understanding customer loyalty is an important part of any business. Banks, telecommunication companies, airlines, Internet service providers, pay TV companies, and insurance firms etc., utilize customer *churn* or *attrition* rates as one of their key business metrics. This metric is important as the churn rate of a business is a good indicator of customer response to services, pricing, and competitions. The ability to identify churny contents / behaviors can enable early intervention processes (as part of *retention campaigns*) and ultimately a reduction in customer churn. Retention campaigns often involve three major steps (Huang, Kechadi, and Buckley 2012): (a) *churn identification* that indicates if a customer is at the risk of canceling the company's service, (b) *evaluation* that measures if a customer worth retaining, and (c) *execution* that identifies the best retention technique for each potential churner.

Churn identification, the first step of retention campaigns, is crucial as the cost of retaining an existing customer is much less than acquiring a new one (Huang, Kechadi, and Buckley 2012). This problem has been extensively studied on Call-Record data (CRD) in the context of social graphs of telecommunication companies (Verbeke, Martens, and Baesens 2014; Karnstedt et al. 2010) and to a lesser extent on online gaming (Kawale, Pal, and Srivastava 2009) and chat and forum communities (Karnstedt et al. 2011; Oentaryo et al. 2012). In this work, we investigate churn identification in the context of microblog platforms such as Twitter and Facebook that provide publicly available user-generated contents. To the best of our knowledge, this is the first work that investigates churn identification in the microblog contexts. Table 1 shows some examples of churny micro-posts about different brands taken from our Twitter dataset. Identifying such churny contents will enable companies to provide personalized assistance to their potential churners, and competitors to hunt new customers.

We formally define the problem of target-dependent churn classification in microblogs as follows: Given a micro-post t_i posted by user u_j about brand b_k, determine if t_i is churny or non-churny with respect to b_k. Here, we assume that for each brand b_k there exists a list of competing brands b_i, $i = \{1 \dots n\}$, $i \neq k$. In this paper, we identify three major categories of churn indicators for target-dependent churn classification in microblogs. In particular, we investigate *demographic* churn indicators (obtained from users of micro-posts), *content* churn indicators (obtained from the textual content of micro-posts), and *context* churn indicators (obtained from threads containing the micro-posts). We examine factors that make this problem more challenging and investigate the performance of several state-of-the-art machine learning techniques on this problem. A challenging aspect of such classification task is that churny contents can be expressed in a subtle manner. For example the tweet "*debating if I should stay with Verizon*" contains no word that is obviously churny, but it represents a potential churn with respect to the brand. As such, churn classification requires a more

One of my main goals for 2013 is to leave *BrandName*. I cant take it anymore, the unlimited data isnt even worth it.
My days with *BrandName* are numbered.
BrandName: I will change carriers as soon as contract is up.
Really hate to leave *BrandName*-1 but... Hey *BrandName*-2?
BrandName your customer service is horrible. this loyal customer will be gone #awfulcustomerservice
Cant wait to leave *BrandName-1* for *BrandName-2*! One more bill!!
If I cant keep my unlimited data, then bye *BrandName*.

Table 1: Sample churny micro-posts (brand names are replaced with the placeholder *BrandName*).

solid understanding of the natural language.

We conduct experiments on a Twitter dataset created from a large number of tweets about three telecommunication brands. Experimental results show an average F1 performance of 75% for target-dependent churn classification in microblogs. Our dataset is available at www.umiacs.umd.edu/~hadi/chData.

Preliminary Analysis

In churn classification, one may suspect that there are certain words that can be used to express churny contents. Therefore, it might be enough to produce such a list of keywords and rely on them alone to classify micro-posts as churny or non-churny. To test this hypothesis, we manually created a list of 50 potential churny terms and phrases[1] based on a preliminary examination of our twitter dataset (see Section 4) and Wordnet synonyms. We then designed a rule-based classifier that labels micro-posts as churny or non-churny with respect to a target brand based on the presence or absence of such keywords (along with the brand name) in micro-posts respectively. Table shows the performance of the rule-based classifier on the positive (churn) class based on the standard precision, recall, and F1-score measures.

As the results show, the rule-based classifier has a very low precision and therefore low F1 performance[2]. Our closer look at the data reveals five major reasons that lead to the poor performance of the rule-based classifier:

First, there exist comparative micro-posts in which users compare two / several brands against each other and express their intention about leaving one brand for another. Such micro-posts are only churny with respect to one brand but not the others. However, the rule-based classifier labels the micro-posts regardless of any target brand. This greatly reduces the performance of the classifier.

Second, simple language constituents such as prepositions are important in accurate classification of micro-posts with respect to target brands. For example, prepositions like "*to*" and "*from*" can simply reverse the class of a micro-post with respect to a brand, compare phrases like "*switch to BrandName*" vs. "*switch from BrandName*". These important indicators are simply ignored by the rule-based classifier.

Third, negation has an important contextual effect on churn classification. It can simply reverse the label of a micro-posts. For example, the rule-based classifier will mistakenly label the micro-post "*BrandName is awesome, I'll never leave them*" as churny with respect to the brand.

Fourth, the presence of a churny keyword does not necessarily convey churny contents. For example, the micro-posts "*BrandName-1 to give BrandName-2 customers up to $450 to switch*" or " *I need a little BrandName's #help b4 leaving the states*" contain the terms "*switch*" and "*leaving*" respectively, but they are not churny with respect to the brands.

Finally, churny contents can be expressed in subtle ways as in "*debating if I should stay with BrandName*" or "*in 2 months, bye BrandName*" that contain no obvious churny keywords but clearly express churny contents with respect to the brand. Such micro-posts are simply missed by the rule-based classifier.

We conclude from these preliminary experiments that it is worthwhile to explore the major indicators for target-dependent churn classification in microblogs. We introduce these indicators in the subsequent Sections.

Brand	Churny	Non-churny	P^+	R^+	$F1^+$
Verizon	447	1543	36.5	74.7	48.9
T-Mobile	95	978	39.0	79.5	50.1
AT&T	402	1389	35.4	81.4	48.4

Table 2: Performance of the rule-based churn classifier for three brands. Churny and Non-churny columns show the number of churny and non-churny micro-posts for each brand respectively.

Churn Indicators in Microblogs

We introduce three categories of churn indicators for target-dependent churn classification in microblogs.

Demographic Churn Indicators

Demographic features contain user-specific information that can help classifiers to learn churn classification based on user profiles. These features are shown in Table 3 and are expected to provide useful signals to learn churn classification. For example, if a user is not active with respect to any competitors, then he is less likely to send churny contents about a target brand, or users who often share URLs are more likely news agencies or brand sale representatives who almost never send churny contents about brands. Here, we define an *active day* as a day in which the user sends at least one micro-post about the target brand. We also compute the average *friend* activity ratios for each user to account for the social influence on his behavior. Furthermore, we take into account information like number of friends, followers and etc that have been shown to be highly useful for churn classification on forum and call record data (Karnstedt et al. 2011).

Content Churn Indicators

Content Features represent semantic information extracted from the content of micro-posts. These features are shown in Table 4. N-gram features are usually employed by corpus-based techniques and have been shown to be useful for various text classification problems. In addition to the n-gram features, we find that the neighboring words of brand names in micro-posts often contain rich contextual information for churn classification. We capture the effect of the neighboring words of brand / competitors by considering $2k$ features representing the left and right k neighboring words of the brand, and $2k$ features representing the left and right k neighboring

[1] Sample set of churny keywords used: {*leave, leaving, switch, switching, numbered, cancel, canceling, discontinue, give up, call off, through with, get rid, end contract, change to, changing*, ...}.

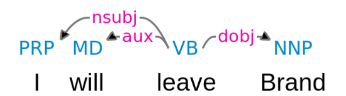
[2] Later experiments using the churny keywords as features of our classifiers did not yield to better results except with linear regression that yielded to a slightly better but still poor performance.

Description
Activity ratio: average No. of posts about brand/competitors per day ratio of active days about brand/competitor average time gap between posts about brand/competitor ratio of urls in posts about brand/competitor average No. of words in post about brand/competitor
Average of friends activity ratios
followers and friends
If user has bio information
If bio contains URL

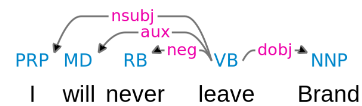
Table 3: Demographic indicators extracted from user / author profiles.

Description
Unigrams / Bigrams
Neighboring words of brand/competitors names
Syntactic and Comparative marker features
Sentiment features
Tense of tweet
News indicator features

Table 4: Content indicators for churn classification. These features will be extracted from the content of the target micro-posts.



(a) syntactic features: {*dobj-leave-Brand*, *nsubj-leave-i*, *aux-leave-will*}.



(b) syntactic features with negation effect: {*Neg-dobj-leave-Brand*, *Neg-nsubj-leave-i*, *Neg-aux-leave-will*}

Figure 1: Syntactic features extracted from micro-posts content by capturing the effect of negation. In Figure 1(b) the verb *leave* is negated by the negation modifier *never*.

words of all competitors (in the experiments, we set $k = 3$ as it leads to superior performance).

In addition, we rely on syntactic parse trees to identify expressions describing the target brand[3]. In particular, for any dependency relation rel_i, except the negation relation, between the brand name and any word w_j, we generate a feature as "rel_i-w_j-$Brand$" or "rel_i-$Brand$-w_j" depending on the role of brand as a governor or dependent in the dependency relation. We also extract such features from the dependency relations that contain a *verb* as their governor or dependent. This is because verbs, especially action verbs, capture most content information in sentences (Levin 1993). To tackle with negations, if any word included in the resultant syntactic features is modified by a negation word, we add a prefix "$Neg-$" to all its corresponding features. Figures 1(a) and 1(b) show the syntactic features extracted for the tweets "*I will leave Brand*" and "*I will never leave Brand*" with the transitive verb "*leave*" respectively.

Furthermore, we observed that the *prepositional*, *direct object*, and *nominal subject* dependency relations are the main syntactic relations useful to capture the effect of simple language constituents. For example, in the tweet "*I am leaving BrandName-1 for BrandName-2*", the prepositional dependency *prep_for(BrandName-1, BrandName-2)*[4] can be used to determine that the churn is in favor of *BrandName-2*, and the nominal subject *nsubj(leaving,I)* and the direct object *dobj(leaving, BrandName-1)* relations can be used to determine the user as the subject and *BrandName-1* as the object of the post. These features together can capture the effect of various language constituents in the micro-posts.

To deal with comparative micro-posts, we identify *comparative markers*. For this purpose, we utilize POS tag patterns (in particular comparative (JJR) and superlative (JJS) adjectives) in conjunction with the dependency relations obtained from the syntactic parse trees of micro-posts. As an example, in the micro-post "*TargetBrand has better coverage than CompBrand*", the POS pattern "*better/JJR */NN than/IN*" and the dependency relation "*prep_than(*, CompBrand)*" can be used to capture comparative effects.

Our analysis also shows that churny micro-posts may carry strong negative sentiment toward the target brands. To account for the sentiment effects, we consider two features indicating the number of positive or negative sentiment terms that co-occur with the target brand in micro-posts. A better approach to capture sentiment information is to obtain target-dependent sentiment score with respect to brands as studied in (Jiang et al. 2011). The approach utilizes the syntactic structure of micro-posts in conjunction with subjectivity clues and target dependent features to determine the sentiment with respect to a given target. We leave this analysis to the future work.

Furthermore, churny micro-posts are expected to have present or future tense. To capture this information, we make use of verb tenses (identified by their POS tag information) and consider three features representing the number of verbs with *past*, *present*, and *future* tenses available in micro-posts respectively. We also observed that news-oriented micro-posts about brands often carry the brand name as their subject and usually contain a URL. To account for such effects, we consider three binary features indicating if the target brand is the (a) subject or (b) object of the micro-post, or (c) if the tweet contains a URL respectively.

[3]We used the Twitter POS tagger developed in (Owoputi et al. 2013) to POS tag micro-posts and then used the resultant POS tagged micro-posts and Stanford parser to generate the parse trees.

[4]The prepositional modifiers serves to modify the meaning of verbs, adjectives, nouns, or even other prepositions, e.g. *prep(switching, from)*

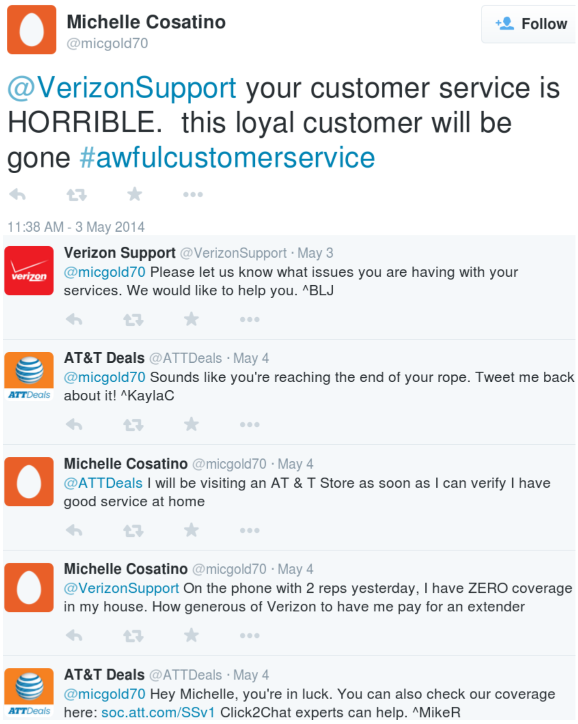


Figure 2: A thread containing a churny tweet with respect to Verizon crawled from Twitter.

Context Churn Indicators

Churny tweets may trigger discussions between users and their friends as well as *known accounts* of brands. The known-accounts of a brand are a few official accounts created on Twitter that either act as informers or provide various services to brand customers. These accounts are usually well-publicized and easy to find. In fact, brands participate in such discussions to retain their customers while competitors intervene to hunt new customers. Figure 2 shows a sample thread triggered by a churny tweet.

We expect thread information to be useful for churn classification as threads provide context information about target micro-posts. Table 5 shows churn indicators that can be extracted from threads. We extract content features (as shown in Table 4) for each micro-post in the thread. To distinguish the content generated by different parties in threads, we put the features into different namespaces by adding "*USer-*", "*FRiend-*", "*BRand-*", and "*COmp-*" prefixes to the content features extracted from user, friend, target brand, and competitor microposts respectively. Furthermore, *reciprocity* between user and brand posts (the time difference between the target micro-post, i.e. the classification input, and the first response from the brand) could be a good indicator of churn as we observed that brands respond to churny micro-posts more quickly than non-churny ones. Also, we compute reciprocity between user and competitors as the time difference between the target micro-post and the earliest response from any competitors.

Description
Content features of user/friends/brand/competitors posts in thread (as defined in Table 4)
posts from user/friends/brand/competitors in thread
posts in thread
Reciprocity between user and brand/competitors posts

Table 5: Context indicators for churn classification. These features will be extracted from threads containing the target micro-posts.

Churn Classification

We investigate the performance of three state-of-the-art classification approaches for churn classification in microblogs. Given input micro-posts (x_1, x_2,..., x_n) and their corresponding (churny, non-churny) labels (y_1, y_2,..., y_n), $y_i \in \{-1, +1\}$, the classification problem is to find a function $f(x)$ that minimizes:

$$L(x, y) = \sum_{i}^{n} l(y_i, f(x_i)) \tag{1}$$

where $l(.,.)$ represents the loss between the actual label and its corresponding prediction. Loss functions are usually functions that becomes close to zero when $f(x_i)$ agrees in sign with y_i. We consider three loss functions here, the linear loss (representing linear classification), the hinge loss (employed by the SVMs), and the logistic loss (representing logistic classifier) defined as follows respectively:

$$l(y, f(x)) = |y.f(x)|$$
$$l(y, f(x)) = max(0, 1 - y.f(x))$$
$$l(y, f(x)) = log(1 + exp(-y.f(x)))$$

We employ Vowpal Wabbit classification toolkit[5] with all parameters set to their default values to perform the classification experiments.

Data and Settings

We collect twitter data for three telecommunication brands, namely *Verizon*[6], *T-Mobile*[7], and *AT&T*[8]. To obtain a development dataset, we crawled tweets that contained the name of the above brands or were posted by the known-accounts (see Section "Context Churn Indicators") of the above brands using the streaming API of twitter for a period of six months. As the pre-processing step, we converted all the tweets into lower-case terms and removed those with less than three words.

The resultant dataset contains a highly imbalanced data distribution in terms of churny vs non-churny tweets. To create our gold-standard dataset, we first clustered the tweets into a chosen number of clusters and then, within each cluster, we further categorized tweets into sub-clusters using our

[5] http://hunch.net/ vw/
[6] https://twitter.com/VerizonSupport
[7] https://twitter.com/TMobileHelp
[8] https://twitter.com/ATTCustomerCare

		Verizon			T-Mobile			AT&T		
	Features	**classic**	**hinge**	**logistic**	**classic**	**hinge**	**logistic**	**classic**	**hinge**	**logistic**
(1)	Unigram	56.5	60.1	59.6	63.5	64.4	66.1	62.2	66.3	67.2
(2)	base:Uni+bigram	67.1	72.8	69.9	67.3	65.9	69.8	74.6	74.9	75.4
(3)	base+Content	**73.4*** (6.3)	75.4* (2.6)	73.8* (3.9)	68.3* (1.0)	68.1* (2.2)	70.6* (0.8)	75.4* (0.8)	77.7* (2.8)	78.5* (3.1)
(4)	base+Demog.	68.5 (1.4)	73.4 (0.6)	70.9 (1.0)	67.6 (0.3)	69.6* (3.7)	68.9 (-0.9)	74.8 (0.2)	76.5* (1.6)	75.7 (0.3)
(5)	base+Context	71.1* (4.0)	73.8* (1.0)	72.0* (2.1)	65.0* (-2.3)	**70.5*** (4.6)	70.0* (0.2)	77.4* (2.8)	77.0* (2.1)	77.8* (2.4)
(6)	All	72.8* (5.7)	**76.8*** (4.0)	**75.4*** (5.5)	**70.3*** (2.9)	70.2* (4.3)	**74.4*** (4.7)	**78.2*** (3.7)	**80.0*** (5.0)	**78.5*** (3.1)

Table 6: F1 Performance of target-dependent churn classification based on different indicators evaluated over three brands. Values in parenthesis show F1 improvement over the baseline, base:Uni+bigram.

manually created list of potential churny terms. We then randomly sampled data from all these clusters (including those that do not contain any churny keywords) for manual annotation. The first round of clustering helps to tackle data redundancy, while the second round helps to capture a fair distribution of churny tweets. We emphasize that learning churn classification is independent of this clustering step as each sub-cluster may contain both positive and negative examples (see "preliminary analysis" Section).

We asked three annotators to manually label the resultant tweets as churny or non-churny with respect to each target brand. Finally we obtained 447 and 1543 churny and non-churny tweets for Verizon, 95 and 978 churny and non-churny tweets for T-mobile, and 402 and 1389 churny and non-churny tweets for AT&T[9]. We weight the positive examples by the ratio of negative to positive examples to deal with imbalanced classification input.

Inter-annotator agreement is computed based on Fleiss' kappa measure (Fleiss 1971). This measure is computed over all the instances labeled by three annotators: the resultant κ value is 0.62 that indicates substantial agreement among annotators (Fleiss 1971). We also report Cohen's kappa value computed over 1073 instances related to T-mobile that were independently labeled by the first author of this paper and compared against the aggregation of the three annotators judgments over these instances. The resultant Cohen's kappa value is 0.93 that indicates substantial annotation agreement as well (note that the above two κ values are not directly comparable).

Results

In the experiments, we report the classification performance (F1-score) over the churn class. We performed all the experiments through 10-fold cross validation and used the two-tailed paired t-test $\rho < 0.01$ for significance testing. Throughout this Section, we use the asterisk mark (*) to indicate significant improvement over the baseline.

Since there is no previous work on churn classification in microblogs, we resort to n-gram models to determine a baseline for this task (note that ngram models lead to relatively good performance on various text classification problems). We experimented with ngrams (n=$\{1 \ldots 3\}$) and their combination both at the word and character levels. We found that the combination of unigrams and bigrams at the word level leads to the best performance. As such we considered this setting as the baseline for this task.

Table shows the F1 Performance based on different indicators evaluated over three brands. Row (2) shows that the combination of unigrams and bigrams greatly improves the performance over unigrams. Row (3) shows that the content indicators (see Table 4) significantly improve the performance over the baseline for all brands and through all the three classic, hinge, and logistic classifiers. Similarly row (5) shows that adding the context indicators (see Table 5) to the baseline features almost always significantly improve the performance except in the case of liner classifier on T-mobile data that results in a comparable performance. The results also show that, although adding demographic features (see Table 3) slightly improve the baseline, the difference is not often significant. We correlate this to the very complex patterns of user activities in social media that are often difficult to capture.

Table at row (6) shows that combining all the demographic, content, and context features together significantly improve the performance over the baseline. However, it does not lead to the best results in only two cases (see the performance of classic and hinge models at rows (3) and (5) for Verizon and T-Mobile respectively), as highlighted in Table . This motivated us to conduct an ablation study to investigate the effect of different indicators on target-specific churn classification in microblogs. In the subsequent Sections, we consider the hinge and logistic models for ablations analysis of content and context indicators and report the average F1 performance over the three brands.

Ablation Analysis of Content Chrun Indicators

Table shows the results of our ablation experiments with content indicators. "Content" shows the average performance when we only use content features. To conduct the ablation study we remove neighbor-

[9]Available at www.umiacs.umd.edu/~hadi/chData

Features	Hinge	Logistic
base:Uni+bigram	71.2	71.7
Content	73.7	74.3
Content−Nighbr	72.3 (−1.4)	73.3 (−1.0)
Content−SynComp	73.4 (−0.3)	73.4 (−0.9)
Content−STN	**75.0** (+1.3)	**74.6** (+0.3)

Table 7: Ablation analysis of content-specific indicators. F1 values are averaged over three brands.

Features	Hinge	Logistic
base:Uni+bigram	71.2	71.7
base+Context	73.8	73.3
base+Context−user	71.9 (−1.9)	72.3 (−1.0)
base+Context−friend	**75.0** (+1.2)	**73.4** (+0.1)
base+Context−competitor	74.4 (+0.6)	73.1 (−0.2)
base+Context−brand	74.5 (+0.7)	73.1 (−0.2)

Table 8: Ablation analysis of context-specific indicators. F1 values are averaged over three brands.

ing words (Content−Nighbr), syntactic and comparative (Content−SynComp), as well as sentiment, tense, and news indicator (Content−STN) features from the content features and report the average performance over the three brands.

As Table shows, neighboring words as well as syntactic and comparative features are important indicators as their removal reduces the average F1 performance. However, removing sentiment and tense features improves the performance but not significantly. Note that, the sentiment features may not accurately capture the sentiment with respect to the target brand mainly due to the negation effect. Also, as for the tense features, our observation shows that the majority of micro-posts have a present or future tense and as such, in contrast to our expectation, this feature may not be effective for churn classification. Table shows that removing the STN features from the combination of all features leads to a comparable performance with no significant change.

Ablation Analysis of Context Churn Indicators

In this experiments, we evaluate the effect of different parties contributing to threads (users, friends, brands, and competitors) on the churn classification performance. Table 8 shows the results of our ablation experiments.

As the results show, if we remove the context features that were obtained from users activities in the thread, i.e. (base+Context−user), the performance significantly reduces as compared to when we use all such features (base+Context). On the other hand, removing context features obtained from friends activities significantly improves the performance. This suggests that friends activities in threads might not be important for churn classification. We also observed the same results by removing such features from the combination of all features (see All−friend in Table).

We also analyzed the correlation of different churn indicators with class labels based on Information Gain Ratio using

Selected Features	Hinge	Logistic
base:Uni+bigram	71.2	71.7
All	75.7	**76.1**
All−STN	76.1 (+0.4)	75.8 (−0.3)
All−friend	**76.3** (+0.6)	**76.1** (0.0)
All−STN−friend	76.1 (+0.4)	75.8 (−0.3)

Table 9: Ablation analysis of selected churn indicators. F1 values are averaged over three brands.

the Verizon data. The top three most demographic indicatives are in order (1) *average friend active days about competitors*, (2) *average friend posts about competitors*, and (3) *average friend active days about brand*. The corresponding content indicatives are in order (1) *OBject=verizon* where verizon is the object of the micro-post, (2) "*L1=leaving*" where L1 is the first left neighboring word of brand as in *I am leaving Verizon*), and (3) "*poss_contract_my*" that indicates the possession modifier syntactic relation between the words "my" and "contract" as in "*my contract is over*". The corresponding context indicatives are all from user posts in threads and are in order (1) "*USer_leaving*", (2) "*USer_contract*", and (3) "*USer_dobj_leave_verizon*" that indicates direct object relation between the transitive verb "leave" and the brand as in "*I want to leave Verizon*").

Related Work

Despite the numerous applications of user retention and the availability of public user generated contents in microblogs, target-specific churn classification appears to be underexplored in microblogs. The majority of previous works extensively studied churn classification on Call-Record data (CRD) in the context of social graphs (Verbeke, Martens, and Baesens 2014; Huang, Kechadi, and Buckley 2012; Karnstedt et al. 2010) and to a lesser extent on online gaming (Kawale, Pal, and Srivastava 2009) and chat / forum communities (Karnstedt et al. 2011; Oentaryo et al. 2012; Patil et al. 2013). Research on CRD, utilized various indicators including user information such as age, gender, bill frequency, account balance, outstanding charges, and call details such as call duration, prices and fees etc, as well as historical information of bills and payments, and social relations based on incoming / outgoing calls. Churn classification in the game and forum domains also utilized various user-specific features such as age and gender as well as features obtained from social relations among people. Recently (Dave et al. 2013) presented a churn classifier for recommender systems. The approach utilizes user-specific features such as user ratings and the time that users spend on items to predict churn in recommender systems.

To the best of our knowledge, churn classification has been unexplored in microblogs. However, microblog platforms provide publicly available user-generated contents about different aspects of brands that can be effectively utilized to identify potential churners. In this work, we utilized such contents as well as other social signals for target-dependent churn classification in microblogs.

Conclusion and Future Work

In this paper we investigate the problem of target-dependent churn classification in microblogs. We study the most churn indicatives in microblogs and examine factors that make this problem more challenging. As our future work, we aim to further investigate this problem from both content representation and social influence perspectives.

References

Dave, K. S.; Vaingankar, V.; Kolar, S.; and Varma, V. 2013. Timespent based models for predicting user retention. In *Proceedings of WWW*.

Fleiss, J. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5):378–382.

Huang, B.; Kechadi, M. T.; and Buckley, B. 2012. Customer churn prediction in telecommunications. *Expert Syst. Appl.* 39(1):1414–1425.

Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; and Zhao, T. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, 151–160. Stroudsburg, PA, USA: Association for Computational Linguistics.

Karnstedt, M.; Hennessy, T.; Chan, J.; Basuchowdhuri, P.; Hayes, C.; and Strufe, T. 2010. Churn in social networks. In Furht, B., ed., *Handbook of Social Network Technologies and Applications*. Springer US.

Karnstedt, M.; Rowe, M.; Chan, J.; Alani, H.; and Hayes, C. 2011. The effect of user features on churn in social networks. In *Proceedings of Web Science*. ACM.

Kawale, J.; Pal, A.; and Srivastava, J. 2009. Churn prediction in mmorpgs: A social influence based approach. In *Proceedings of International Conference on Computational Science and Engineering*. IEEE Computer Society.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

Oentaryo, R. J.; Lim, E.-P.; Lo, D.; Zhu, F.; and Prasetyo, P. K. 2012. Collective churn prediction in social network. In *Proceedings International Conference on Advances in Social Networks Analysis and Mining*.

Owoputi, O.; Dyer, C.; Gimpel, K.; Schneider, N.; and Smith, N. A. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of NAACL*.

Patil, A.; Liu, J.; Shen, J.; Brdiczka, O.; Gao, J.; and Hanley, J. 2013. Modeling attrition in organizations from email communication. In *Social Computing (SocialCom)*.

Verbeke, W.; Martens, D.; and Baesens, B. 2014. Social network analysis for customer churn prediction. *Applied Soft Computing* 14,C(0):431–446.