An Empirical Study on Learning based Methods for User Consumption Intention Classification

Mingzhou Yang, Daling Wang, Shi Feng and Yifei Zhang

Northeastern University, Shenyang, China
yangmingzhou11@163.com; wangdaling@cse.neu.edu.cn
fengshi@cse.neu.edu.cn; zhangyifei@cse.neu.edu.cn

Abstract. Recently, huge amount of text with user consumption intentions have been published on the social media platform, such as Twitter and Weibo, and classifying the intentions of users has great values for both scientific research and commercial applications. User consumption analysis in social media concerns about the text content representation and intention classification, whose solutions mainly focus on the traditional machine learning and the emerging deep learning techniques. In this paper, we conduct a comprehensive empirical study on the user intension classification problem with learning based techniques using different text representation methods. We compare different machine learning, deep learning methods and various conbinations of them in tweet text presentation and users' consumption intention classification. The experimental results show that LSTM models with pre-trained word vector representation can achieve the best classification performance.

Keywords: Consumption Intention, Intention Classification, Text Representation, Machine Learning, Deep Learning

1 Introduction

With the development of social media, more and more users like to express their views, ideas, and other contents on the platform. There is a large amount of information released by different users to express their needs to some kind of commodity or service, which is the so-called user consumption intention. These online texts with consumption intentions are of great value to both scientific research and commercial application. Therefore, consumption intention analysis in text data has drawn wide attentions in recent years.

One existing solution is regarding user consumption intension analysis as a classification problem [1, 2, 3, 4]. Based on this idea, the user generated texts (such as reviews, discussions, and posts published by users in social media) are used to train classifiers using either traditional machine learning techniques [1, 2, 3] or the emerging deep learning algorithms [4]. In addition, how to effectively represent the text for modeling user intensions is still a challenging task.

In this paper, we conduct an empirical study for learning based user consumption intention classification in user generated texts. We apply different language models to represent the texts, and use traditional machine learning and emerging deep learning techniques respectively for classifying the consumption intentions embedded in the texts. Our purpose is exploring the characteristics of machine learning and deep learning techniques in classifying users' intentions.

As a whole, our main contributions in this paper are as follows:

(1) We implement the representation of text using traditional language model such as TF-IDF and neural language model such as Word2Vec.

(2) We train different classifiers using traditional machine learning methods such as SVM and Naïve Bayes, and deep learning methods such as RNN and LSTM.

(3) We conduct comprehensive experiments by combining the different text presentation models and different classification methods, and discuss the characteristics of above models according to the experimental results.

The remainder of this paper is structured as follows: Section 2 describes the problem of consumption intention classification, and gives our implementation framework. In Section 3, we analyze a series of experiments conducted for exploring the characteristics of different text presentation and classification techniques. We introduce the related work in Section 4. Finally we conclude the paper and give the future work in Section 5.

2 Problem and Approach Description

2.1 Problem Description

In social media, user generated text include two types, i.e. with consumption intention and without consumption. For the former, the intention may be "*I want to buy a new dress*" or "*I'm going to see the sea in Dalian*". The above two sentences express clear consumption requirement. For the latter ones, such as "*The street is too dirty*" or "*I'm so glad you're all right!*", the text only express some emotions, but not intentions.

In previous literature, consumption intention is categorized into six classes including "Food & Drink", "Travel", "Career & Education", "Goods & Services", "Event & Activities", and "Trifle" [3]. The detailed explanations are as follows. We give examples of each category in Table 1.

Table 1. Categories of Intent Tweets

Category	Example
food	hungry...i need a salad......four more days to the BEYONCE CONCERT...
travel	I need a vacation really bad. I needa trip to Disneyland!
career	this makes me want to be a lawyer RT someuser new favorite line from an ...
goods	mhmmm, i wannna a new phone. Services ... i have to go to the hospital. ...
trifle	on my way to go swimming with the twoon @someuser; i love her so muchhhhh!
event	I'm so happy that I get to take a shower with myself. :D

- Food: Want to have some food or drink.
- Travel: Want to visit some specific points of interests or places.
- Career: Want to get a job, get a degree or do something for self-realization.
- Goods: Want to have some non-food/non-drink goods (e.g., car) or services (e.g., haircut).
- Trifle: Want to talks about daily routine, or some mood trifles.
- Event: Want to participate in some activities which do not belong to the aforementioned categories (e.g., concert).

Based on the class labels, models can be trained for user consumption intention classification.

2.2 Implementation Framework

In this paper, we also use the six intension classes and "Non-Intention" as labels. For the empirical study, we consider different text presentation and intention classification technique, so our framework is shown as Figure 1.

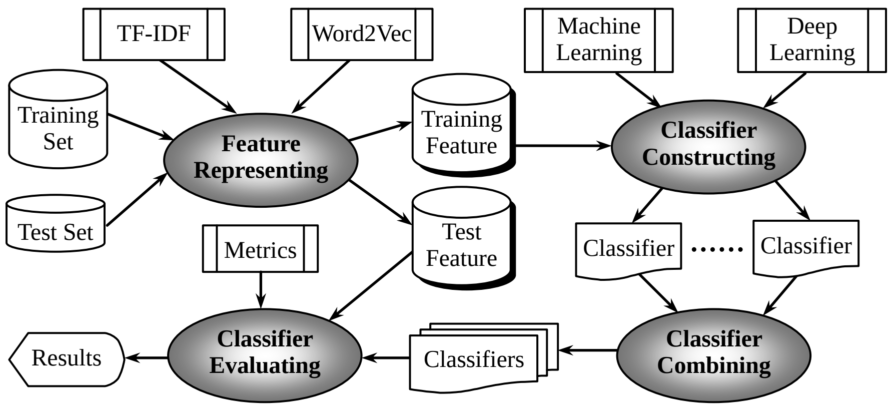

Fig. 1. Framework of Machine Learning and Deep Learning for Intention Classification

The Training and Test Set have been annotated with six consumption intentions and "Non-Intention". There are four operations in the framework.

- Feature Representing: For every text in Training Set and Test Set, TF-IDF or Word2Vec are used to represent it, and the results are Training Feature Set and Test Feature Set.
- Classifier Constructing: For above Training Feature Set, machine learning techniques such as SVM and Naïve Bayes, and deep learning techniques such as RNN and LSTM are applied to construct classifier. The results are some different classifiers.
- Classifier Combining: Based on different feature representing methods, different classifiers, and their different combinations, new classifiers are constructed. The results are some recombination classifiers.
- Classifier Evaluating: For Test Set, based on its feature representation and some evaluation metrics e.*g*. *F*-score, above new recombination classifiers are evaluated.

2.3 Text Feature Presentation and Classifier Construction

In feature representation, we consider TF-IDF [5] and Word2Vec [6] model respectively. In this paper, the text feature vector represented by TF-IDF is used to construct traditional machine learning classifiers, and the ones by Word2Vec is used to construct traditional machine learning and deep learning classifiers. Figure 2 shows the feature presentation and classifier construction for intention mining detail in Figure 1.

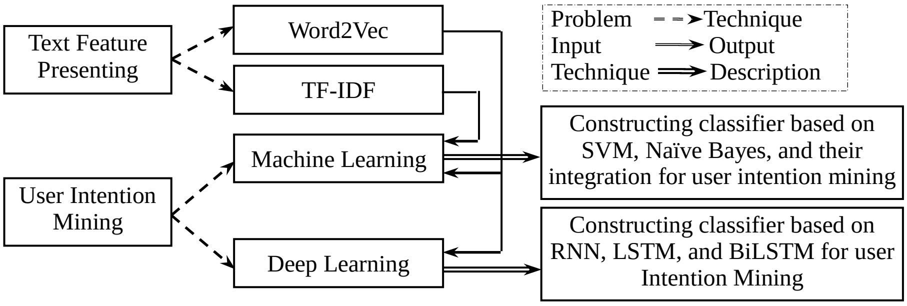

Fig. 2. Problem, Techniques Used, and Corresponding Descriptions

Based on TF-IDF, we utilize traditional machine learning techniques such as SVM and Naïve Bayes to construct different classifiers. The results include SVM classifier, Naïve Bayes classifier, and SVM+Naïve Bayes classifier. In addition, we utilize SMOTE method to balance dataset.

Based on Word2Vec, we utilize traditional machine learning techniques such as SVM and SVM+Naïve Bayes, and further utilize deep learning techniques such as RNN and LSTM to construct different classifiers. The results include SVM classifier, RNN classifier, SVM+Naïve Bayes classifier, LSTM classifier, and BiLSTM (bidirectional LSTM) classifier.

3 Experiments

3.1 Experiment Setup

We evaluate above classifiers by experiments. The dataset is from Twitter. With the process in [3], the resultant statistical information of the dataset is shown as Table 2. In the Table 2, we can see the dataset have the problem of imbalanced data. And the number of tweet in “Non-Intent” and “Trifle” class is nearly twice as much as that of other classes except “Event” class.

Table 2. Statistical Information of Experiment Data

Non-Intent	Food	Travel	Career	Goods	Trifle	Event	Total
531	245	187	159	251	436	321	2130

After reprocessing the dataset such as removing stop-words and balancing data classes, we use above classifiers and their combinations to classify consumption intentions. We apply *F*-score to evaluate the classification results. For a binary classification, the true class labels and the ones obtained by classifiers can constitute True Positive (*TP*), False Positive (*FP*), True Negative (*TN*), and False Negative (*FN*). In this paper, for an intention (e.g. Food), the corresponding confusion matrix is shown as Table 3.

Table 3. Confusion Matrix of Classification Results

Classified Intention / True Intention	Positive	Negative
Positive	True Positive (*TP*)	False Negative (*FN*)
Negative	False Positive (*FP*)	True Negative (*TN*)

For an intention class, here we take Food as an example, if the intention of a tweet is Food, and the tweet is classified as Food, the result is *TP*. If it is classified as other intention, the result is *FN*. If the intention of a tweet is not Food, but it is classified as Food, the result is *FP*. If it is classified as other intention, the result is *TN*. So do other intentions.

According to the confusion in Table 3, precision (*P*), recall (*R*), and *F*-score are defined as Formula (1).

$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN} \quad F-\text{score} = \frac{2 \times P \times R}{P+R} \tag{1}$$

3.2 Experiment Results and Analysis

Our experiments concern the following classification methods besides SVM^{TI} and NB^{TI} (The superscript TI means TF-IDF feature representation).

- SVM^{TI}+NB^{TI}: It is ensemble of SVM and Naïve Bayes classifiers, where TF-IDF is used to represent features.
- SVM^{200}, SVM^{400}: The SVM classifiers utilize 200, 400 dimensional word vectors as features respectively.
- $\text{SVM}^{\text{SMOTE}}$: The SVM classifiers use the SMOTE method to balance dataset.
- NB^{SMOTE}: The Naïve Bayes classifiers use the SMOTE method to balance dataset.
- SVM^{100}+NB^{TI}: It is the ensemble of SVM and Naïve Bayes classifier, and 100 dimensions word vector is the input of SVM.
- RNN^{100}, LSTM^{100}: The RNN and LSTM are used as classifier respectively, and 100 dimensional word vectors are used as their inputs.
- BiLSTM^{100}: It is bidirectional LSTM with 100 dimensional word vectors as its input.

We show the *F*-score results of different feature presentations, classifiers, and their combinations in Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7.

Figure 3 is the comparison of three traditional machine learning classifiers. We can see their ensemble model has the best results almost in all classes.

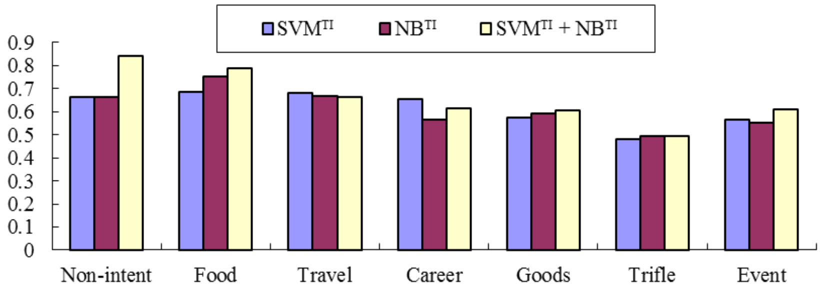

Fig. 3. Comparison of Traditional Machine Learning Methods

Figure 4 is the comparison of three deep learning classifiers. For the same Word2Vec dimension, both LSTM and BiLSTM classifiers have better results than RNN in most classes.

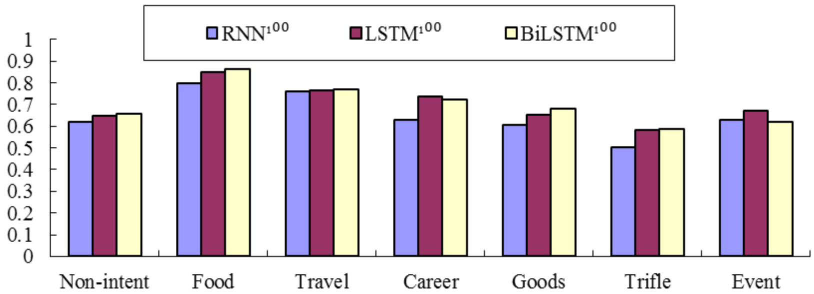

Fig. 4. Comparison of Adanced Deep Learning Methods

In Figure 5, for SVM, the word vector features show better results than traditional TF-IDF features in all the dimension settings.

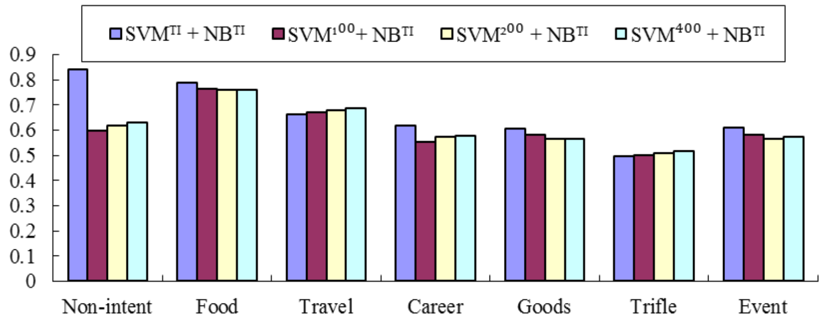

Fig. 5. Comparison of SVM with Different Feature Presentations

Moreover, when combining SVM and Naïve Bayes, the word vector based classifier seems better than TF-IDF. The phenomenon is shown in Figure 6.

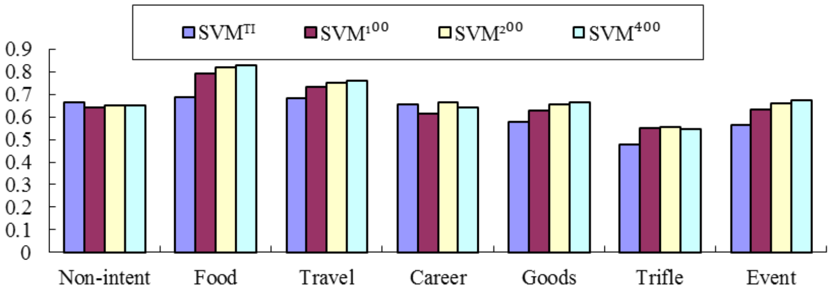

Fig. 6. Comparison of SVM+NB with Different Feature Presentations

We compared the traditional machine learning classifiers using SMOTE with the traditional machine learning classifiers without using it in Figure 7. We can see the classifiers using SMOTE method have the better results than the ones without using it. This observation demonstrates that balanced dataset can improve classification quality.

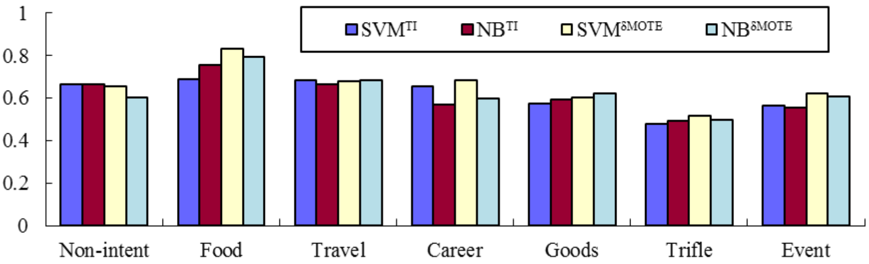

Fig. 7. Comparison of Traditional Machine Learning Methods using SMOTE Method

Form Figure 3 to Figure 7, in all results, we can see the "Trifle" class has a lower *F*-score almost in all figures, which may be due to some tweet in "Trifle" class is similar to tweet in "Event" class and "Goods" class.

Finally, we compare the average *F*-score of all intention classes, and the result is shown in Figure 8. As a whole, LSTM and BiLSTM have achieved the best results. Moreover, the SVM classifiers with word vector features also show the second best results. This observation further demonstrates that distributed word embedding can improve classification quality for both traditional classifiers and deep models.

In addition, for the same dataset, Wang et al. gave classification results [3] (Non-Intent 35.56%, Food 54.63%, Travel 58.64%, Career 45.73%, Goods 43.25%, Event 27.13%, Trifle 20.04%). Our results outperform Wang's method in all classes.

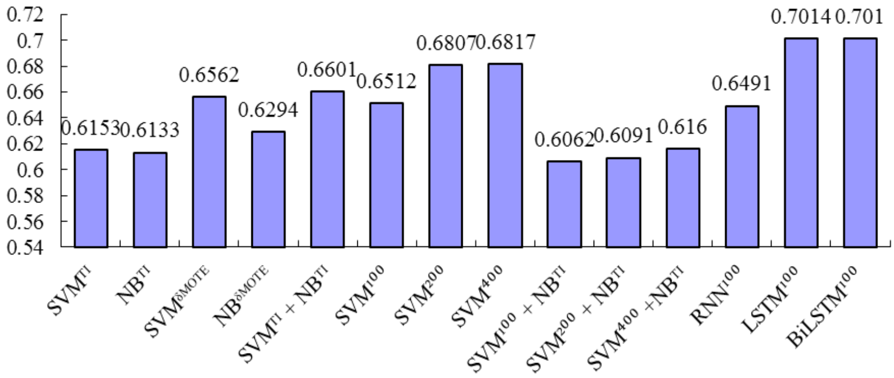

Fig. 8. Comparison of Average *F*-score for Different Classes of Intention

4 Related Work

Our work focuses on user consumption intention classification in social media text based on shallow and deep learning techniques. We do not consider extracting intention features from the text as [3]. In fact, in consumption intention analysis, there are many researches from different aspects. Besides text, the existing researches also consider other factors such as user behaviors [2, 7], user emotions [8], user interaction and participation [9], product features [10], website characteristics [11], and more factors [12]. Some literatures utilize other techniques such as innovation diffusion theory and the technology acceptance model [13] to detect user intensions.

Our purpose in this paper is exploring the characteristics of traditional machine learning and deep learning techniques in classifying users' intention for user generated text date, not proposing a new theory and technique. So we do not compare our work with above related work. But this work can be as our further research.

5 Conclusion and Future Work

For user generated text data in social media, we apply traditional and advanced language model to represent the text feature, and utilize learning based techniques to classify user consumption intention embedded in the text. We implement different classifiers using different text features, and evaluate their classification performance by experiments. Experiment results show that the LSTM models can achieve the best *F*-Score in model classes.

Obviously, the overall mechanism and detailed techniques used in this paper are not novel. However, how to utilize exist techniques to present tweet text feature, discover and class users' consumption intentions is very important for the special application. So our empirical study is significant, and indeed conveys valuable information on the problem.

In the further, we intend to adjust parameters for more deep learning models and import more factors as above related work for obtaining better improvements. Moreover, we also intend to propose new method for the purpose.

Acknowledgment

This project is supported by National Natural Science Foundation of China (61370074, 61402091)

References

1. Pérez-Vera S, Alfaro R, Allende-Cid H: Intent Classification of Social Media Texts with Machine Learning for Customer Service Improvement. HCI (13) 2017: 258-274
2. Shi F, Ghedira C: Intention-Based Online Consumer Classification for Recommendation and Personalization. HotWeb 2016: 36-41
3. Wang J, Cong G, Zhao W, Li X: Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets. AAAI 2015: 318-324
4. Ding X, Liu T, Duan J, Nie J: Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network. AAAI 2015: 2389-2395
5. Hiemstra D: A probabilistic justification for using tf x idf term weighting in information retrieval. Int. J. on Digital Libraries 3(2): 131-139 (2000)
6. Mikolov T, Chen K, Corrado G, Dean J: Efficient Estimation of Word Representations in Vector Space. CoRR abs/1301.3781 (2013)
7. Fortes N, Moreira A, Saraiva J: Determinants of Consumer Intention to Use Online Gambling Services: An Empirical Study of the Portuguese Market. IJEBR 12(4): 23-37 (2016)
8. Akman I, Mishra A: Factors influencing consumer intention in social commerce adoption. IT & People 30(2): 356-370 (2017)
9. Phang C, Zhang C, Sutanto J: The influence of user interaction and participation in social media on the consumption intention of niche products. Information & Management 50(8): 661-672 (2013)
10. Chen C: Young Female Consumers' Perceptions and Purchase Intentions towards Character Economy. HCI (9) 2017: 382-392
11. Chen Y, Wu C: Influence of Website Design on Consumer Emotion and Purchase Intention in Travel Websites. IJTHI 12(4): 15-29 (2016)
12. Shang S, Wu Y, Sie Y: Generating consumer resonance for purchase intention on social network sites. Computers in Human Behavior 69: 18-28 (2017)
13. Agag G, El-Masry A: Understanding consumer intention to participate in online travel community and effects on consumer intention to purchase travel online and WOM: An integration of innovation diffusion theory and TAM with trust. Computers in Human Behavior 60: 97-111 (2016)