

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

12-2016

Domain identification for intention posts on online social media

Thai Le Luong

Quoc Tuan TRUONG


Singapore Management University, qttruong@smu.edu.sg

Hai-Trieu Dang

Xuan Hieu Phan

DOI: <https://doi.org/10.1145/3011077.3011134>

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

 Part of the [Numerical Analysis and Scientific Computing Commons](#), and the [Social Media Commons](#)

Citation

Luong, Thai Le; TRUONG, Quoc Tuan; Dang, Hai-Trieu; and Phan, Xuan Hieu. Domain identification for intention posts on online social media. (2016). *SoICT '16: Proceedings of the Seventh Symposium on Information and Communication Technology: Ho Chi Minh, Vietnam, December 8-9, 2016*. 52-57. Research Collection School Of Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research/3624

This Conference Proceeding Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Domain Identification for Intention Posts on Online Social Media

Thai-Le Luong
University of Transport and
Communications
Hanoi, Vietnam
luongthai80@utc.edu.vn

Quoc-Tuan Truong
SIS Research Center,
Singapore Management
University (SMU)
qttruong@smu.edu.sg

Hai-Trieu Dang
University of Engineering and
Technology, Vietnam National
University, Hanoi
trieudh_58@vnu.edu.vn

Xuan-Hieu Phan
University of Engineering and
Technology, Vietnam National
University, Hanoi
hieupx@vnu.edu.vn

ABSTRACT

Today, more and more Internet users are willing to share their feeling, activities, and even their intention about what they plan to do on online social media. We can easily see posts like “I plan to buy an apartment this year”, or “We are looking for a tour for 3 people to Nha Trang” on online forums or social networks. Recognizing those user intents on online social media is really useful for targeted advertising. However fully understanding user intents is a complicated and challenging process which includes three major stages: *user intent filtering*, *intent domain identification*, and *intent parsing and extraction*. In this paper, we propose the use of machine learning to classify intent-holding posts into one of several categories/domains. The proposed method has been evaluated on a medium-sized collections of posts in Vietnamese, and the empirical evaluation has shown promising results with an average accuracy of 88%.

CCS Concepts

•Information systems → Data mining; Web mining; Social tagging; •Computing methodologies → Information extraction;

Keywords

Intention mining; user intent identification; domain classification; social media text understanding; text classification

1. INTRODUCTION

Nowadays, many Internet users commonly share their feeling, daily activities, and even their intention on online social

media channels like Facebook and Twitter. For example, one user may post “I am going to buy a seven-seater car next week” or “We are looking for an apartment near the downtown center” on a discussion forum or on his/her own Facebook wall. Those posts are called “intention posts” because they carry user intents to do something in the near future. Intention posts or messages are obviously a valuable source of knowledge for enterprises. If enterprises know and understand exactly what online users are planning to do, they can easily locate a large number of potential customers relevant to their business domain. However, the most challenging question is how can we process, analyze and understand those intention posts automatically?

The process of analyzing and understanding intention posts on online social media consists of three major stages: *user intent filtering*, *intent domain identification*, and *intent parsing and extraction* [9]. *User intent filtering* means we need to crawl user posts and filter which are intention posts, i.e., posts that carry an intent. This step has been carried out in Luong et al. 2016 [9]. The second stage (*intent domain identification*) is to identify domain or category of an intention post, i.e., determining what a post is about (e.g., health, finance, food, job, traveling, etc.). The final stage (*intent parsing and extraction*) is to analyze each post (text) content in order to extract all concrete information about the intent, i.e., understanding all properties of that intent.

In the scope of this paper, we focus on solving the second stage (*intent domain identification*) that helps to determine what an intention post is about. We consider this problem as a classification task, that is, each intention post is classified into a most suitable domain/category. This classification task is actually a text categorization problem where the input texts are short and quite ambiguous. There are several challenges in this task. First, an intention post commonly contains several sentences and it is sometimes very hard to determine the real domain of a post. For example, a post like “I am going to buy a seven-seater car for traveling at weekend.” This intention is about “buying a car”, however it can also be classified into “tourism” because it contains the word “traveling”. The second challenge is that intention posts on online social media are very diverse. The number of specific domains is usually very large as users can share their

intention about anything. It is very hard to perform a classification task with large number of classes. Therefore, we only classify intention posts into one of 12 major domains like electronic device, fashion and accessory, finance, food service, furnishing and grocery, travel and hotel, property, job and education, transportation, health and beauty, sport and entertainment, and pet and tree.

We have conducted experiments with real data crawled automatically from four well-known discussion forums and social networks. We have built a medium-sized labeled data set of text posts in Vietnamese for evaluation. Classification models were trained using Support Vector Machines (SVMs) and Maximum Entropy (MaxEnt). We have achieved promising results with both classifiers.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the whole user intent identification process. Section 4 presents our main work, that is, building classification models to identify domains for intention posts. Experimental results and analysis will be presented in Section 5. Finally, conclusions will be given in Section 6.

2. RELATED WORK

Recently, there are more and more studies that aim to mine user intention from online social media data. There have been different approaches to this problem. In this section, we will present some studies that are more or less relevant to our work. To the best of our knowledge, there is no one studying intention mining for text documents until 2013. Most of them are for web search where they focused on intent identification for search queries. Rose and Levinson (2004), Jansen et al. (2004), Kathuria et al. (2010), they all tried to understand the user intent from web search queries by classifying the queries into three major categories: informational, navigational, or transactional [10, 7, 8]. Baeza-Yates et al. (2006) presented a framework for the identification of the user’s interests based on the analysis of query logs from web search engines. They first attempted to find the user goals and then mapping those queries into the categories: informational, not-informational, and ambiguous, and eighteen categories of topic to classify the queries into. Almost all categories are based on The Open Directory Project¹ [2]. Azin Ashkan et al. (2009) used the features of query based, content of search result pages and ad clickthrough to classify queries into two dimensions: {commercial, non-commercial} and {navigational, informational} [1].

The following studies are most relevant to our work. Chen et al. (2013) claimed that their solution is the first one that try to identify user intents in discussion forum posts. They proposed a new transfer learning method to classify the posts into two classes: intent posts (positive class) and non-intent posts (negative class) [4]. This work is most similar to our previous work that solves the first stage (user intent filtering) [9] in the user intent understanding process. But there is still a little difference between their work and ours: while they only consider purchase intents in four domains {cellphone, electronic, camera, tivi}, our work handles a lot of intent types, such as purchase, sell, hire, rent, borrow etc. and in a wide range of domains. Similarly, Gupta et al. (2014) attempted to identify only purchase intent from social post by categorizing the posts into two

¹Open Directory Project: <http://dmoz.org>

classes namely PI and non-PI. This has been done by extracting features at two different levels of text granularity, that are word and phrase based features and grammatical dependency based features [6]. More relevant to our work, Wang et al. (2015) attempted to mine user intents in Twitter by classifying tweets into six categories {food and drink, travel, career and education, goods and services, event and activities, and trifle} [11].

3. DOMAIN IDENTIFICATION AS A STAGE OF INTENT IDENTIFICATION PROCESS

As we proposed in our previous paper [9], the process of analyzing and understanding user intents includes three major stages as shown in Figure 1. They are:

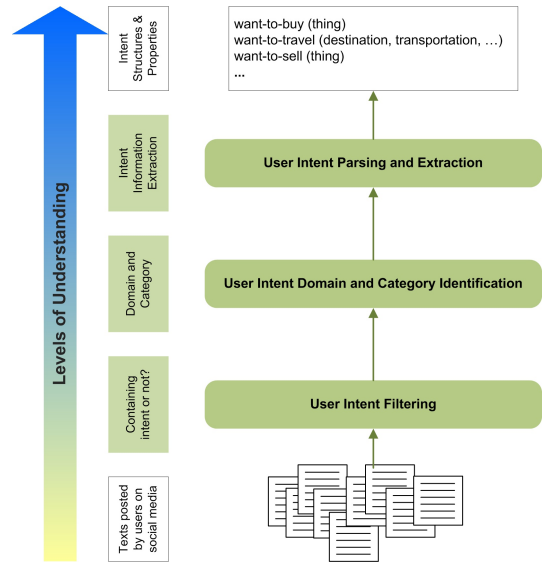


Figure 1: Process of mining/identifying user intent from (online social media) texts.

- Stage 1 – *User intent filtering*: this phase helps to filter text posts on online social media channels (blogs, forums, online social networks) to determine which posts contain user intents and which do not. Posts carrying user intents will be forwarded to the next stage below. This is actually a binary classification problem and has been solved in our previous work [9].
- Stage 2 – *Intent domain identification*: given a text post containing a user intent, this phase will analyze and identify the domain of the intent. This is the main problem we are aiming at to solve in this paper. In our work, an intent can be classified into one of the following categories: {electronic device, fashion and accessory, finance, food service, furnishing and grocery, travel and hotel, property, job and education, transportation, health and beauty, sport and entertainment, pet and tree}. This is actually a multi-class classification for short and ambiguous texts.
- Stage 3 – *Intent parsing and extraction*: given a text post containing an intent and its domain category, this phase will parse, analyze, and extract all concrete information (i.e., properties) of the intent. For example,

if an intent is about tourism, its properties may be {destination(s), transportation, time-period, number of people, etc.}.

Figure 2 shows a specific example of the user intent understanding process. The input is a text post on social media talking about the intent for a honeymoon trip of a married couple. *User intent filtering* module determined that this post holds an intent. In the next step, *intent domain identification* module determined its domain is travel/tourism. The post and its domain are then forwarded to the final phase, *User intent parsing and extraction*. At this step, the properties/constraints of the intent were parsed and extracted.

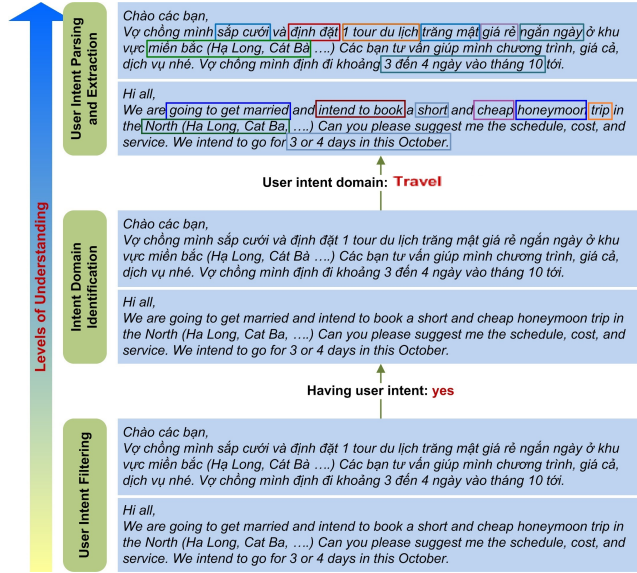


Figure 2: Example of the user intent mining process.

In our previous work, we aimed to solve the *user intent filtering* phase by proposing a classification model to filter the intent posts from online Vietnamese social media texts. In this paper, we focus to solve the second phase – *intent domain identification*, that determine the most suitable domain for each intent. We will propose the set of twelve intent domains. The classification models will be built with support vector machines (SVMs) and maximum entropy (MaxEnt).

4. INTENT DOMAIN IDENTIFICATION

4.1 The Set of Intent Domains

Building the set of intent domains turns out to be a difficult task. We had to discuss several times among data annotators to agree on a most suitable partitioning for intent posts. Each partition is considered as an intent domain. It means we want to make sure that if an intent post belongs to one domain, it cannot be assigned to any other domains. After carefully analyzing the set of data and referring to several reference web sites²³ in Vietnam, we decided to divide the intent posts into thirteen domains as shown in Table 1.

²<https://www.consumerbarometer.com/about>

³<https://www.chotot.com>

Since an intent post maybe appear in the middle of a long conversation that the clear intention was mentioned at the beginning, it is difficult to identify its domain if only based on the post. For example, a user may write “I’m going to buy the same one too” or “ship 1 kg for me at this weekend”. It is so difficult to understand the exact intent domain for these posts although we know that the posts carry purchase intents. Moreover, there are some posts simultaneously express more than one intent. For example, a post like “I want to buy a second-hand eating chair for my baby. By the way, I’m looking for an extra job to have more income” may be categorized in two different domains (furnishing & grocery and job & education). It will make the work more complicated. In the scope of this paper, we do not consider these sorts of posts. It means we only consider classifying posts that contain only one clear domain.

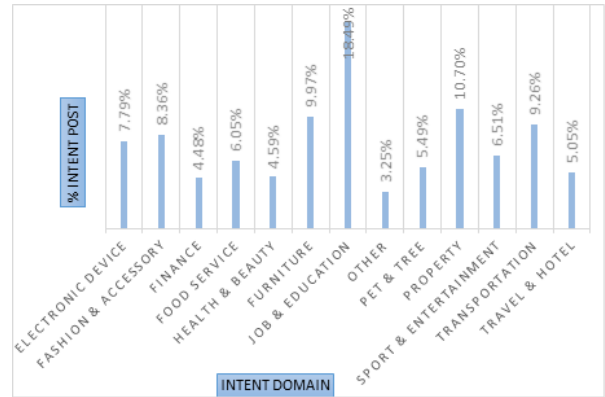


Figure 3: The statistic of intent posts from our data.

The chart in figure 3 shows the percentage of each intent domain. The data were crawled from several famous discussion forums⁴⁵⁶⁷ in Vietnam and from Facebook, this can be considered the distribution of intent domains for Vietnamese intent posts. As we can see, the domain job & education has the highest frequency, less frequent domains are property, furnishing & grocery, transportation and fashion & accessory.

4.2 Building Domain Classification Models

4.2.1 Maximum Entropy Classification (MaxEnt)

Classification based on the maximum entropy principle is to build a classification model with what have been known from data and assume nothing else about what are not known. This means MaxEnt model is the model having the highest entropy while satisfying all constraints observed from empirical data. Berger et al. (1996) [3] showed that MaxEnt model has the following mathematical form:

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp \left(\sum_{i=1}^n \lambda_i f_i(x, y) \right) \quad (1)$$

where x is the data object that needs to be classified and y is the output class label. $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ is the vector

⁴<http://www.webtretho.com/forum>

⁵<https://www.lamchame.com/forum>

⁶<http://sotaychame.com/dien-dan.html>

⁷<https://www.chotot.com>

Table 1: Intent domain descriptions and examples

Intent Domain	Descriptions / Examples	#(%)
Electronic Device	I want to liquidate the old refrigerator ... I have an old breast pumps want to sell ...	546 (7.79%)
Fashion & Accessory	I was presented a pair of leather shoes, but they do not fit me, so I want to sell them ... Is there any mum here know a nice fashion clothes store, please show me, I need to buy a new dress ...	586 8.36%
Finance	I urgently need to borrow a huge amount of money ... I'm looking for someone who can make capital contribution ...	314 (4.48%)
Food Service	This weekend, I have some nice bacon, who want to buy, please order with me... I'm looking for a restaurant to celebrate my son's birthday...	424 (6.05%)
Furnishing & Grocery	Is there any mom here want to liquidate a dinning chair for kid, I need one... I'm finding a brand new wardrobe...	699 (9.97%)
Health & Beauty	I'm going to buy a pressure cuff for my mother ... I really want to have a nose-lift performed ...	322 (4.59%)
Job & Education	I have a pressing need of finding a domestic helper ... I'm looking for an English class of communication for my 12-year-old child ...	1296 (18.49%)
Other	I need a smart accounting software ... I'm looking for a souvenir for my girl friend ...	228 (3.25%)
Pet & Tree	I need to sell my dog because I have no time to take care for him ...	385 (5.49%)
Property	I'm going to buy an apartment ...the price is about 1.5 million (Vietnam dong) ... For hire, shop premises with frontages on two streets ...	750 (10.70%)
Sport& Entertainment	I want to find a swimming class for my son ... I have a pair of tickets for Le Quyen liveshow this Saturday, want to resell ...	456 (6.51%)
Transportation	I'm looking for a new 7-seater car to replace my old one ... I have a redundant air ticket to Sai Gon, need to resell ...	649 (9.26%)
Travel & Hotel	I want to book a travel tour for 3 people to Nha Trang ...	354 (5.05%)

of weights associated with the features $F = (f_1, f_2, \dots, f_n)$, and $Z_\lambda(x) = \sum_{y \in \mathcal{L}} \exp(\sum_{i=1}^n \lambda_i f_i(x, y))$ is the normalizing factor to ensure that $p_\lambda(y|x)$ is a probabilistic distribution. Once trained, the MaxEnt model will be used to predict class labels for new data. Given a new object x , the predicted label is $y^* = \operatorname{argmax}_{y \in \mathcal{L}} p_\lambda(y|x)$.

4.2.2 Support Vector Machines (SVMs)

The idea behind binary SVMs [5] is to build a classification model based on the optimal separating hyperplane between the two classes by maximizing the margin between the two classes. In the Figure 4, the points lying on the boundary are called support vectors, and the middle of the margin is the optimal separating hyperplane. This means that the SVM algorithm can operate even in fairly large feature sets as the goal is to measure the margin of separation of the data rather than matches on features. Previous studies have shown that SVMs scale well and have good performance on large data sets.

Figure 5 below shows the basic idea behind Support Vector Machines when working with the nonlinear separable data. Here we see the original objects (left side of the figure) mapped, i.e., rearranged, using a mathematical function, known as kernel function. The process of rearranging the objects is known as mapping (transformation). Note that in this new space, the mapped objects (right side of the figure) is linearly separable and, thus, instead of constructing the complex curve (like the left), all we have to do is to find an optimal hyperplane in the new space.

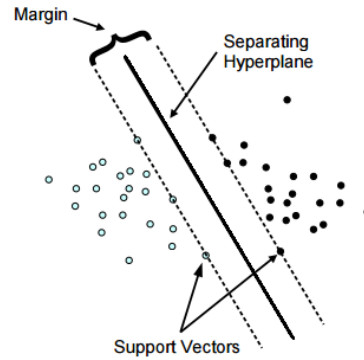


Figure 4: SVM Classification (linear separable case).

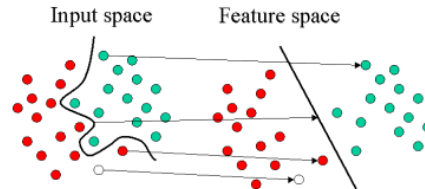


Figure 5: Transformation from nonlinear case to linear case.

4.3 Feature Templates

In order to build classification models with MaxEnt and

SVM, we need to define our feature templates. We used two types of features in our models. The first is n -grams and the second is dictionary look-up features. We used both 1-grams (word tokens themselves) and 2-grams (two consecutive word tokens). When combining two consecutive word tokens to form 2-grams, we did not join two consecutive tokens if there is a punctuation mark between them.

We also built a dictionary for look-up features. After training the models with n -grams features we selected top thirty words or phrases with highest weight features for each intent domain. From those chosen words or phrases, we filtered out the meaningless ones so that for each intent domain we only kept from ten to thirty key words or phrases to build the dictionary. By this way, this dictionary contains key words or key phrases used to express the thirteen intent domains most accurately. Figure 6 shows several key words or phrases having high weights for each domain.

Intent Domain	Key words/phrases
Electronic Device	ti:vi, mua:đt, máy, tủ:lạnh, điều:hòa, nokia, sam:sung, iphone, cũ (television, machine, refrigerator, nokia, samsung...)
Fashion & Accessory	size, mặc, đồng:hồ, da, shop, nữ, giày, dây, thời trang (size, wear, watch, leather, shop, woman, shoes,...)
Finance	vay, muốn:vay, cho:vay, mượn, cần:vay, ngân:hàng, vay:tiền, (borrow, want:to:borrow, loan,bank, borrow:money ...)
Food Service	ngon, kg, ăn, bánh, sữa, ship, đặt, g, nhà hàng (delicious, kg, eat, cake, milk, ship, oder,restaurant...)
Furnishing & Grocery	ghế, bím, sách, sat, tủ, ghế:ăn:dặm, cũ, giường (chair, diaper, book, wardrobe, cot, bed...)
Health & Beauty	nước:hoa, thuốc, cream, nước, khám, đo, giảm:cân, thẩm mỹ (perfume, cream, examine, lose:weight,...)
Job & Education	trường, tuyển, nhận, học, lớp, việc, giúp:việc, cho:con, anh, cô (school, recruit, au:pair, job, class, study, teacher...)
Other	tài liệu, hang, băng, đĩa, video, nick (product, document, cd, video, nick...)
Pet & Tree	chó, chim, đực, nuôi, trồng, con, cây (dog, bird, plant, feed, tree...)
Property	thuê, căn, nhà, tầng, đất, mua:nhà, chung cư, m2 (hire, apartment, floor, m2, house, plot...)
Sport & Entertainment	vợt, chơi, cầu, tập, iắc, yoga, tạ, bàn, patin, bàn:bida (play, racket, pole, fishing:rod, train, yoga, billard:table ...)
Transportation	xe, bay, vé máy bay, đi, chạy, bán:xe, mua:xe, máy:bay, (mean, flight, air:ticket, drive, selling:car, plane)
Travel & Hotel	voucher, tour, đêm, du, lịch, du:lịch, khách:sạn, nghỉ (voucher, tour, night, travel, hotel, vacation...)

Figure 6: Some high weighted look-up features for each intent domain.

5. EVALUATION

To evaluate the performance of intent domain classification, we have conducted careful experiments with SVMs and MaxEnt. The experimental results will be described as below.

5.1 Experimental Data

We have built a medium-sized collection of intent posts from famous discussion forums in Vietnam, such as Webtretho.com, Lamchame.com, Chotot.com, Sotaychame.com. We have also crawled intention posts from Facebook. After removing all irregular cases that we mentioned in Section 4.1, the data collection consists of 7009 intent posts. A group of students were asked to label each post into one of the thirteen domains based on a common annotation guideline and the agreement among them. Some examples for

each intent domain can be seen in the table 1. And Figure 3 also gives the statistic of the intent domains. The labeled data collection were then divided randomly into five parts. The experiments were then performed using 5-fold cross validation and the experimental results will be reported in the next subsection.

5.2 Experimental Results and Analysis

For all experiments, we use *precision*, *recall* and F_1 -score as the evaluation measures. Table 2 shows the experiment results of the best fold (the 5th fold). In this table, we can see the precision, recall and F_1 -score of both SVM and MaxEnt models for each intent domain. In this fold, the SVM model gave better results. We achieved the macro-averaged F_1 -measure of 87.38 and the micro-averaged F_1 -measure of 90.14 with the SVM model. This is a significantly high result because we only use n -grams and dictionary look-up features to build the classifying model.

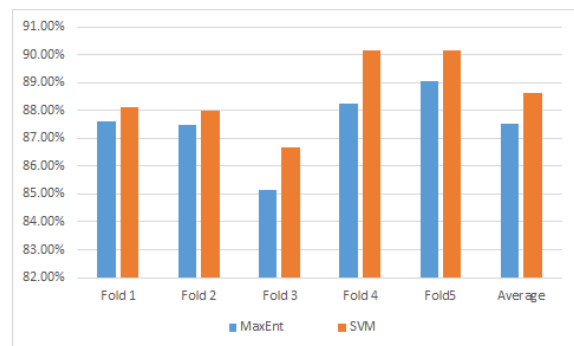


Figure 7: The accuracy of the 5-fold CV tests.

Figure 7 shows the accuracy (i.e., micro-averaged F_1 -score) of the five folds and the average over the five folds of both SVM and MaxEnt models. We can see that for every fold the SVM model always achieves better results than the MaxEnt model. For more details, we calculated the F_1 -score for each intent domain classification and the results are shown in Figure 8. We realized that in almost all intent domains, the F_1 -score values of the SVM models are higher than those of the MaxEnt model.

We can easily see that the domain other always has the lowest accuracy. This is understandable because of two reasons: (1) the number of intent posts belonging to the other class is smallest (accounts for only 3.25% of our total labeled data); (2) the other class contains miscellaneous intentions (as been mentioned in Table 1) that we cannot place them in any of the twelve intent domains. Thus it makes very difficult to find the dictionary look-up features for the other class. However, except the other class, we can see that the results are quite stable over the remaining twelve domains even though the number of intent posts for these domains are unequal. For example, job & education class has the number of intent posts be about three times as many as that of travel class, but as we can see in Table 2 that the F_1 -measure of these two class are almost the same. This shows that the classification models can work well on this data set.

6. CONCLUSIONS

In this paper, we have presented the problem of domain

Table 2: The precision, recall and F_1 -score of NE types of the SVM and MaxEnt best fold

Intent Domain	SVM-Prec	SVM-Rec	SVM- F_1	ME-Prec	ME-Rec	ME- F_1
Electronic Device	81.20	82.80	82.00	77.00	77.80	77.40
Fashion & Accessory	82.80	91.40	86.90	80.30	89.50	84.70
Finance	95.00	87.70	91.20	80.30	89.50	84.70
Food Service	96.10	90.20	93.10	96.80	93.80	95.30
Furnishing & Grocery	77.70	89.00	83.00	81.90	84.10	83.00
Health & Beauty	93.80	84.50	88.90	84.50	84.50	84.50
Job & Education	95.80	96.90	96.40	95.10	96.60	95.80
Other	70.00	42.40	52.80	56.30	54.50	55.40
Pet & Tree	89.60	92.00	90.80	90.40	88.00	89.20
Property	94.70	96.00	95.30	96.60	96.00	96.30
Sport& Entertainment	92.50	77.90	84.60	88.00	76.80	82.00
Transportation	94.40	97.50	95.90	90.60	96.70	93.50
Travel & Hotel	95.00	95.00	95.00	97.30	91.30	94.20
Average_{macro}	89.12	86.41	87.38	87.09	86.14	86.54
Average_{micro}	90.14	90.14	90.14	89.06	89.06	89.06

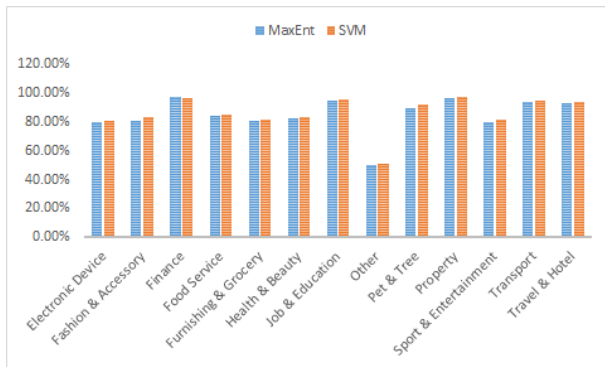


Figure 8: The MaxEnt F_1 -score and SVM F_1 -score of each intent domain.

identification for intention posts and proposed our solution to this problem. We considered this problem as a multi-class classification task. To evaluate, we crawled real posts from online social media, filtering posts containing user intents and performing domain annotation. By this way, we have built a medium-sized labeled dataset for conducting the experiments. In this work, we proposed a set of twelve intent domains for classification. We have built our classification models with SVMs and MaxEnt. The experimental results have shown that the SVM classifier performs a little better than MaxEnt. And both of the methods achieved significantly high results (about 88% of accuracy on average). In the future work, we will perform domain classification with richer features and at sentence level to reduce ambiguity.

7. ACKNOWLEDGMENTS

This work was supported by the project QG.16.34 from Vietnam National University, Hanoi (VNU).

8. REFERENCES

- [1] A. Ashkan, C. L. Clarke, E. Agichtein, and Q. Guo. Classifying and characterizing query intent. In *In Proceedings of The 31th European Conference on Information Retrieval (ECIR)*, pages 578–586, 2009.
- [2] R. Baeza-Yates, L. Calderon-Benavides, and C. Gonzalez-Caro. The intention behind web queries. In *String Processing and Information Retrieval*, pages 98–109, 2006.
- [3] A. Berger, S. A. D. Pietra, and V. J. D. Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996.
- [4] Z. Chen, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Identifying intention posts in discussion forums. In *In Proceedings of The The North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1041–1050, 2013.
- [5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- [6] V. Gupta, D. Kedia, D. Varshney, H. Jhamtani, and S. Karwa. Identifying purchase intent from social posts. In *Eighth International AAAI Conference on Weblogs and Social Media*, pages 180–186, 2014.
- [7] B. J. Jansen, D. L. Booth, and A. Spink. Determining the user intent of web search engine queries. In *In Proceedings of The World Wide Web Conference (WWW)*, pages 1149–1150, 2007.
- [8] A. Kathuria, B. J. Jansen, C. Hafernik, and A. Spink. Classifying the user intent of web queries using k-means clustering. *Internet Research*, 20(5):563–581, 2010.
- [9] T.-L. Luong, T.-H. Tran, Q.-T. Truong, T.-M.-N. Truong, T.-T. Phi, and X.-H. Phan. Learning to filter user explicit intents in online vietnamese social media texts. In *In Proceedings of The Asian Conference on Intelligent Information and Database Systems (ACIIDS)*, pages 13–24, 2016.
- [10] D. E. Rose and D. Levinson. Understanding user goals in web search. In *In Proceedings of The World Wide Web Conference (WWW)*, pages 13–19, 2004.
- [11] J. Wang, G. Cong, W. X. Zhao, and X. Li. Mining user intents in twitter: a semi-supervised approach to inferring intent categories for tweets. In *In Proceedings of The AAAI Conference on Artificial Intelligence*, pages 339–345, 2015.