# Cross-Domain Intention Detection in Discussion Forums

**Conference Paper** · December 2017

DOI: 10.1145/3155133.3155182



**3 authors**, including:

Ngo Xuan Bach
Posts and Telecommunications Institute of Technology


Tu Minh Phuong
Posts and Telecommunications Institute of Technology

# Cross-Domain Intention Detection in Discussion Forums

Ngo Xuan Bach*
Department of Computer Science,
Posts and Telecommunications
Institute of Technology, Vietnam
bachnx@ptit.edu.vn

Le Cong Linh
Department of Computer Science,
Posts and Telecommunications
Institute of Technology, Vietnam
linhlc@ptit.edu.vn

Tu Minh Phuong
Department of Computer Science,
Posts and Telecommunications
Institute of Technology, Vietnam
phuongtm@ptit.edu.vn

## ABSTRACT

This paper presents a method for cross-domain intention detection, which identifies posts expressing user intentions in discussion forums. The results of the task can be beneficial for intelligent business such as online advertising and marketing. Our method utilizes labeled data from several other domains to help the learning task in the target domain using a Naive Bayesian framework. Since the distributions of data vary from domain to domain, it is important to adjust the contributions of different data sources in the learning model to achieve high accuracy. Here, we propose to optimize the parameters of the Naive Bayes classifier using a stochastic gradient descent (SGD) algorithm. Experimental results show that our method outperforms several competitive baselines on a benchmark dataset, consisting of forum posts in four domains, namely Cellphone, Electronics, Camera, and TV.

## CCS CONCEPTS

• **Computing methodologies → Lexical semantics**; Semi-supervised learning settings;

## KEYWORDS

Cross-domain, discussion forums, intention detection, stochastic gradient descent

**ACM Reference Format:**
Ngo Xuan Bach, Le Cong Linh, and Tu Minh Phuong. 2017. Cross-Domain Intention Detection in Discussion Forums. In *SoICT '17: Eighth International Symposium on Information and Communication Technology, December 7–8, 2017, Nha Trang City, Viet Nam.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3155133.3155182

## 1 INTRODUCTION

Web 2.0 platforms such as blogs, wikis, social networks, and internet forums have facilitated the generation of a huge volume of user-generated content. Such social media data have become an important and interesting source, which attracts researchers in both natural language processing and data mining communities. Many research problems have been investigated on social media data, including part-of-speech tagging and named entity recognition [12, 24], sentiment analysis [3, 19, 21], recommender systems [1, 17, 18], social network analysis [11], and intention analysis [8, 13, 14, 16, 20].

In this paper, we consider the task of intention detection in discussion forums. The goal of the task is to identify discussion posts expressing certain user intentions such as buying intentions, which can be exploited in online advertising and marketing. Based on the user intentions, suitable ads will be displayed in social media platforms. We cast the task as a binary classification problem, which classifies discussion posts into intention posts (positive class) or non-intention posts (negative class).

Figure 1 shows examples of intention posts (Examples 1, 2) and non-intention posts (Example 3, 4) from the dataset of Chen et al. [8]. Sentences expressing user intentions are displayed in bold format. Although forum posts are quite long and complicated, only small part of posts indicates directly user intentions. Forum posts, moreover, usually contain abbreviations, personal names, and typos. Such characteristics make the classification very noisy.

To improve performance of classification problems, several methods have been proposed, including domain adaptation and transfer learning [4, 15], semi-supervised learning [5, 26], and weakly supervised learning [3], among others. While domain adaptation and transfer learning exploit labeled data from one or more domains to train the classification model in another domain, semi-supervised learning uses both labeled and unlabeled data to build the classification model. A typical scenario in semi-supervised learning is that we have a small set of labeled data and a large amount of unlabeled data. Weakly supervised learning is quite similar to semi-supervised learning, but we have a training set with noisy labels.

We propose in this paper a new method for intention detection, which leverages labeled data in multi-source domains to improve performance in the target domain. Specifically, our method uses stochastic gradient descent (SGD) to optimize the aggregation process of source and target data in a Naive Bayesian framework. To reduce noisy information in classification, we conduct feature selection based on information gain. We verify the effectiveness of the proposed method on a benchmark dataset, which contains four domains i.e., Cellphone, Electronics, Camera, and TV. Experimental results show that our method outperforms several competitive baselines.

*Also with FPT Software Research Lab.


*SoICT '17, December 7–8, 2017, Nha Trang City, Viet Nam*

ACM ISBN 978-1-4503-5328-1/17/12...$15.00
https://doi.org/10.1145/3155133.3155182

**Example 1**: Hello. ***The wife and I are looking to buy a blue-ray player***. We have come accros a nive bargain for the Panasonic DMP-BD110EB 3D Blue-Ray player. Our telly is 6 months old and is a Panasonic Viera PX-T42G20B which is not 3D. However, is it possible to view 3D movies on this tely through the blue-ray player? Any comments well appreciated. Cheers.

**Example 2:** Three years ago, I purchased some Klipsch Promedia 2.0 speakers after seeking advice from audiophiles like yourselves for the best 2.0 system under $150. And I loved them. I mean I really loved them! Unfortunately for my ears and for my budget, I am running into the problem that apparently everyone else but me knew about with these speakers, and the left speaker is now on the fritz. ***I am looking to replace them with another 2.0 system***. It's three years later, but I'm turning to you, the pros, again: (Nowadays) what's the best 2.0 speaker system in the 150-dollar ballpark?

**Example 3**: hi All We have a Samsung Plasma 50" which i bought from JL about 3 years ago (2 years Warranty left) we have noticed for few months now that there is screen burn of cartoon network logo in the corner but have ignored it thinking JL wont do anything about it but just had my relatives around who say JL should repair it under warranty so got me thinking! Would JL take this in for repair under warranty? anyone done this for screen burn? any advice? call them first or pop in to local JL? Thanks!

**Example 4**: Hi, I have a Pioneer 435XDE Plasma that I'm going to connect up to a new xbox 360 via a HDMI lead to the media box. All good so far, but what I wanted to know in advance (since it's all packed in boxes at mo due to house move) is will the sound go through the HDMI and come out of the Pioneer's TV speakers? I do have a Denon Amp and some Mission surround speakers but no room for those at the moment. I'm going to connect a couple of Mission speakers direct to the speaker outlets on the back of the TV rather than use the official Pio ones for now. I guess the xbox will just pass sound through the HDMI, and it will come out in stereo? I hope..... Thanks in advance for light people can shed onto this. Neil.

**Figure 1: Examples of intention and non-intention posts.**

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces our proposed method. Section 4 presents experiments on intention detection on four different domains. Finally, Section 5 concludes the paper and discusses future work.

## 2 RELATED WORK

A considerable amount of work has been done on intention detection and domain adaptation which we review in this section.

**Intention detection**. Because of its importance in advertisement and targeted marketing, the task of identifying intention from social media such as tweets, forum posts, has attracted substantial research interest in recent years. The most straightforward approach is to formulate the task as a text categorization problem, which can be solved with supervised learning methods. Hollerit et al. [14] classify tweets into containing and not containing intents by training support vector machines (SVMs) on words and part-of-speech n-grams extracted from tweets. Luong et al. [20] use maximum entropy with n-grams features to classify forum and Facebook posts, written in Vietnamese, as intention-containing and normal ones. While simple and straightforward, supervised learning approaches require laborious annotation of documents to create training data, which limit their applicability across various application domains. To alleviate this problem, techniques that make use of unlabeled data from the same domain or labeled data from other domain have been applied to identify and classify intents. Wang et al. [25] use both labeled and unlabeled data to detect intent-containing tweets and classify into six categories with a graph-based semi-supervised learning method. Informally, their method classify a tweet based on the proximities of its content to words that are specific to intent categories. Chen et al. [8] leverage labeled data from other domains to train a classifier for the target domain by using domain adaptation techniques. They propose a co-training like algorithm that alternates between two classifiers trained on source and target domains to boost the final performance. Ding et al. [10] also report on successful application of domain adaptation by using convolutional neural networks with shared middle layers, which were trained on labeled data from other domains. A different approach has been proposed by Li et al [16], in which they use Wikipedia as an external source of knowledge and map microblog posts to Wikipedia concepts to classify them, thus reducing the need of having large training sets.

**Domain adaptation and transfer learning**. Domain adaptation and closely related transfer learning has been long studied in settings when we have labeled data from one or more domains and want to use the data to train classifiers for another domain, but the distributions of features and/or labels are different from domain to domain [2, 4, 23]. Blitzer et al. [4] were among the first to apply domain adaptation to sentiment classification - a document categorization problem by nature. They train a classifier to predict the presence of domain dependent feature based on domain independent features, thus reducing the mismatch of features between different domains. Other approaches use different types of transformation to project source and/or target features into a new feature space so that they have similar distributions in the new space (an overview is given in [23]). Pan et al. [22] propose a so called transfer component analysis to optimize an objective function based on the Maximum Mean Discrepancy principle. They reported good transfer effect over several applications. Bach et al. [2] propose a method that finds a new feature space by combining canonical correlation analysis with word embeddings. They report improved performance on cross-domain sentiment classification task. Liu and colleagues use the name "lifelong learning" for a group of methods that utilize data from other domain to support supervised or unsupervised learning in the target domain [7]. Among the methods by that group is the method that uses topics learnt from other domain to improve topic modeling in the target domain [6]. Chen et al. [9] also propose a lifelong learning method that uses stochastic gradient descent to adjust the importance of features from multiple source and target domains to gain maximum positive effect of feature transfer while reducing the negative effect (if any). They report improved performance for sentiment classification using labeled data from multiple source domains.

For document categorization in a new domain, we may have some labeled data and a lot of unlabeled data. This is a typical semi-supervised learning setting. A popular semi-supervised learning method is co-training [5], which is used by Chen et al. [8] in their intention detection method to combine labeled and unlabeled data from different domains. Graph-based learning is another popular semi-supervised learning approach, which has also been applied to intent detection and classification as reported by Wang et al. [25]. In addition to semi-supervised, the setting when training data contain noisy labels, it is known as weakly supervised learning, an example of which is reported by Bach et al. [3], where they use noisy training data from ratings to augment labeled data to predict sentiment polarity of new review posts.

Our method here is based on domain adaptation of external data sources for improved intention detection in forum posts.

# 3 PROPOSED METHOD

In this section, we present our method for cross-domain intention detection in discussion forums. We consider the case that we have labeled data in both source domains and the target domain. The goal is to leverage labeled data in multi-source domains to improve the performance of intention detection in the target domain.

## 3.1 Method Overview

As illustrated in Figure 2, our method consists of three modules: data aggregation, optimization, and classification.

- **Data aggregation**: This module extracts knowledge from multi-source domains and stores it in a knowledge base.
- **Optimization**: This module utilizes knowledge from the knowledge base to optimize key parameters, which will be used in the classification model.
- **Classification**: This module uses the optimized parameters to build the Naive Bayesian classification model.

In the following, we describe these modules in detail. Classification with Naive Bayes will be presented before the optimization section for clarity and readability purposes.

## 3.2 Aggregation

For each source domain $\hat{s}$, we count the number of times a word $w$ appears in positive or negative class, $N^{\hat{s}}_{+,w}$ and $N^{\hat{s}}_{-,w}$. We then compute the number of occurrences of $w$ in the documents of the positive (and negative) class in all the source domains and store them in a knowledge base:

$$N^{KB}_{+,w} = \sum_{\hat{s}} N^{\hat{s}}_{+,w}$$

$$N^{KB}_{-,w} = \sum_{\hat{s}} N^{\hat{s}}_{-,w}$$

## 3.3 Classification with Naive Bayes

In classification step, the goal is to find a class label $c_j$ given a sample $d$. Here, $d$ is a document, i.e. a post in discussion forums, and $c_j$ is a label class indicating whether the post expresses an intention (+) or not (−). Naive Bayesian classification will find the class label that maximizes the conditional probability $P(c_j|d)$.

By using the Bayes' theorem and the independence assumption, we have:

$$P(c_j|d) = \frac{P(d|c_j)P(c_j)}{P(d)} \approx \frac{\prod_w P(w|c_j)P(c_j)}{P(d)} \tag{1}$$

here the product is computed over all word $w$ in $d$. Because the denominator of the Equation (1) is independent from the class label, it can be ignored in computation. Further more, $P(c_j)$ can be estimated based on the frequency of label $c_j$ in the training data, we focus on the key parameters $P(w|c_j)$, which are computed as follows:

$$P(w|c_j) = \frac{\lambda + N_{c_j,w}}{\lambda|V| + \sum_{v\in V} N_{c_j,v}}$$

where $N_{c_j,w}$ is the frequency of word $w$ in documents of class $c_j$. $|V|$ is the size of vocabulary $V$ and $\lambda$ ($0 \le \lambda \le 1$) is used for smoothing.

Recall that we consider the task in a cross-domain setting, $N_{c_j,w}$ will be counted in the whole dataset, consisting of labeled data in both source and target domains. A simple method to compute $N_{c_j,w}$ is to sum up the counts in multi-source domains, i.e. in the knowledge base, with the empirical counts in the target domain as follows:

$$N_{+,w} = N^{KB}_{+,w} + N^{t}_{+,w}$$

$$N_{-,w} = N^{KB}_{-,w} + N^{t}_{-,w}$$

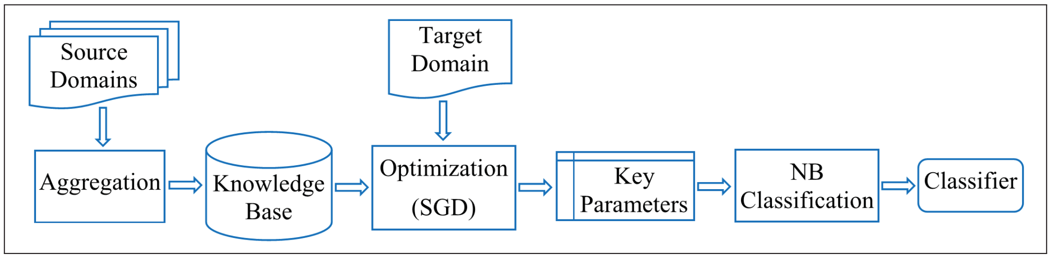


**Figure 2: A Method for cross-domain intention detection.**

where $t$ denotes target domain and $KB$ denotes source domains in the knowledge base. This method, however, has two weaknesses.

- Past domains contain much more data than the target domain. Merged results, therefore, may be dominated by the counts from the source domains.
- The method does not consider domain-dependent words. A word may be an indicator of intention (+) in the target domain but not (−) in source domains.

To deal with such problems, we introduce a method to revise these counts by optimizing two variables $X_{+,w}$ and $X_{-,w}$, the number of times that a word $w$ appears in the positive and negative class. In classification, we will use those virtual counts instead of empirical counts $N_{+,w}$ and $N_{-,w}$.

## 3.4 Optimization

Ideally, we expect $P(c_j|d_i) = 1$ if $c_j$ is the correct class label of document $d_i$ and $P(c_f|d_i) = 0$ for other labels. In optimization, for each positive document $d_i$, i.e. a post with an intention, we want the probability $P(+|d_i)$ as high as possible and the probability $P(-|d_i)$ as low as possible. On the other hand, for a negative document we want the probability $P(-|d_i)$ as high as possible and the probability $P(+|d_i)$ as low as possible.

We define the objective function for each positive document (the process for a negative document is similar) as follows:

$$Obj_{+,i} = log\frac{P(+|d_i)}{P(-|d_i)}$$

To maximize the objective function, we minimize the following loss function.

$$F_{+,i} = -Obj_{+,i} = logP(-|d_i) - logP(+|d_i) \quad (2)$$

We define $G_{+,w}$ for a word $w$ as follows:

$$G_{+,w} = logP(w \mid -) - logP(w \mid +)$$

We have:

$$\begin{aligned} F_{+,i} &= \log\big(P(-)\prod_{w\in d_i} P(w \mid -)^{n_{w,d_i}}\big) \\ &\quad - \log\big(P(+)\prod_{w\in d_i} P(w \mid +)^{n_{w,d_i}}\big) \\ &= \log P(-) - \log P(+) + \sum_{w\in d_i}(n_{w,d_i} \times G_{+,w}) \\ &= \log P(-) - \log P(+) + n_{u,d_i} \times G_{+,u} \\ &\quad + \sum_{w\in d_i, w\neq u}(n_{w,d_i} \times G_{+,w}) \end{aligned}$$

where $n_{w,d_i}$ denotes the number of times that word $w$ appears in document $d_i$, $P(+)$ is the prior probability of a document occurring in positive class ($P(-)$ is similar), $P(w \mid +)$ is the conditional probability of a word $w$ occurring in a document of positive class ($P(w \mid -)$ is similar). Replace

$$P(w \mid +) = \frac{\lambda + X_{+,w}}{\lambda|V| + \sum_{v\in V} X_{+,v}}$$

$$P(w \mid -) = \frac{\lambda + X_{-,w}}{\lambda|V| + \sum_{v\in V} X_{-,v}}$$

and $\lambda = 1$, we have:

$$G_{+,w} = \log\frac{1 + X_{-,w}}{|V| + \sum_{v\in V} X_{-,v}} - \log\frac{1 + X_{+,w}}{|V| + \sum_{v\in V} X_{+,v}}$$

Derivatives $\frac{\partial G_{+,u}}{\partial X_{+,u}}, \frac{\partial G_{+,u}}{\partial X_{-,u}}$ for a word $u$ and $\frac{\partial G_{+,w}}{\partial X_{+,u}}, \frac{\partial G_{+,w}}{\partial X_{-,u}}$ for a word $w$ (different to $u$) can be computed as follows:

$$\begin{aligned} \frac{\partial G_{+,u}}{\partial X_{+,u}} &= -\frac{(|V| + \sum_{v\in V} X_{+,v}) - (1 + X_{+,u})}{(|V| + \sum_{v\in V} X_{+,v})^2} \\ &\quad \times \frac{|V| + \sum_{v\in V} X_{+,v}}{1 + X_{+,u}} \\ &= -\frac{(|V| + \sum_{v\in V} X_{+,v}) - (1 + X_{+,u})}{(|V| + \sum_{v\in V} X_{+,v})(1 + X_{+,u})} \end{aligned}$$

$$\begin{aligned} \frac{\partial G_{+,w}}{\partial X_{+,u}} &= \frac{(1 + X_{+,w})}{(|V| + \sum_{v\in V} X_{+,v})^2} \times \frac{|V| + \sum_{v\in V} X_{+,v}}{1 + X_{+,w}} \\ &= \frac{1}{(|V| + \sum_{v\in V} X_{+,v})} \end{aligned}$$

$$\frac{\partial G_{+,u}}{\partial X_{-,u}} = \frac{(|V| + \sum_{v\in V} X_{-,v}) - (1 + X_{-,u})}{(|V| + \sum_{v\in V} X_{-,v})^2} \times \frac{|V| + \sum_{v\in V} X_{-,v}}{1 + X_{-,u}} = \frac{(|V| + \sum_{v\in V} X_{-,v}) - (1 + X_{-,u})}{(|V| + \sum_{v\in V} X_{-,v})(1 + X_{-,u})}$$

$$\frac{\partial G_{+,w}}{\partial X_{-,u}} = \frac{-(1 + X_{-,w})}{(|V| + \sum_{v\in V} X_{-,v})^2} \times \frac{|V| + \sum_{v\in V} X_{-,v}}{1 + X_{-,w}} = \frac{-1}{(|V| + \sum_{v\in V} X_{-,v})}$$

Therefore, derivatives of $F_{+,i}$ can be computed as follows:

$$\frac{\partial F_{+,i}}{\partial X_{+,u}} = n_{u,d_i} \times \frac{\partial G_{+,u}}{\partial X_{+,u}} + \sum_{w\in d_i, w\neq u} (n_{w,d_i} \times \frac{\partial G_{+,w}}{\partial X_{+,u}}) = -\frac{n_{u,d_i}}{1 + X_{+,u}} + \frac{n_{u,d_i}}{|V| + \sum_{v\in V} X_{+,v}} + \sum_{w\in d_i, w\neq u} \frac{n_{w,d_i}}{|V| + \sum_{v\in V} X_{+,v}}$$

$$\frac{\partial F_{+,i}}{\partial X_{-,u}} = n_{u,d_i} \times \frac{\partial G_{+,u}}{\partial X_{-,u}} + \sum_{w\in d_i, w\neq u} (n_{w,d_i} \times \frac{\partial G_{+,w}}{\partial X_{-,u}}) = \frac{n_{u,d_i}}{1 + X_{-,u}} - \frac{n_{u,d_i}}{|V| + \sum_{v\in V} X_{-,v}} - \sum_{w\in d_i, w\neq u} \frac{n_{w,d_i}}{|V| + \sum_{v\in V} X_{-,v}}$$

Stochastic gradient descent (SGD) is used to minimize the loss function (Equation (2)) by the following update rules, where $\gamma$ is the learning rate[1] and $l$ denotes the iteration:

$$X^l_{+,u} = X^{l-1}_{+,u} - \gamma \frac{\partial F_{+,i}}{\partial X_{+,u}}$$

$$X^l_{-,u} = X^{l-1}_{-,u} - \gamma \frac{\partial F_{+,i}}{\partial X_{-,u}}$$

$X^0_{+,w} = N_{+,w}$ and $X^0_{-,w} = N_{-,w}$ are served as the starting point for SGD.

### 3.5 Feature Selection

In intention-containing posts, only one or two sentences express the intention, while most sentences are not related directly to the intention. Therefore, feature selection is very important for the task to reduce noisy information. Similar to the work of Chen et al. [8], we select top ranked features according to their Information Gain (IG) values, which can be computed as follows:

$$IG(f) = -\sum_{i=1}^{m} P(c_i) \log P(c_i) + \sum_{f,\overline{f}} P(f) \sum_{i=1}^{m} P(c_i \mid f) \log P(c_i \mid f)$$

[1]In experiments, we set $\gamma = 0.01$.

**Table 1: Statistical information of the dataset**

| Domain | #intention | #non-intention | #total posts |
|---|---|---|---|
| Cellphone | 184 | 816 | 1000 |
| Electronics | 280 | 720 | 1000 |
| Camera | 282 | 718 | 1000 |
| TV | 263 | 737 | 1000 |

Because $-\sum_{i=1}^{m} P(c_i) log P(c_i)$ is independent of every feature $f$, we can rank features using the following function:

$$IG_2(f) = \sum_{f,\overline{f}} P(f) \sum_{i=1}^{m} P(c_i \mid f) \log P(c_i \mid f)$$

In our case, the set of labels contains positive (+) and negative (-) labels. Therefore:

$$IG_2(f) = P(f)P(+ \mid f) \log P(+ \mid f) + P(f)P(- \mid f) \log P(- \mid f) + P(\overline{f})P(+ \mid \overline{f}) \log P(+ \mid \overline{f}) + P(\overline{f})P(- \mid \overline{f}) \log P(- \mid \overline{f})$$

where $P(+|f)$ and $P(-|f)$ are probabilities that a post has positive (negative) label when it contains feature $f$. $P(+|\overline{f})$ and $P(-|\overline{f})$ are probabilities that a post has positive (negative) label when it does not contain feature $f$.

## 4 EXPERIMENTS

### 4.1 Experimental Setup

We conducted experiments using the dataset for intention detection introduced by Chen et al. [8]. The dataset consists of posts retrieved from different forum discussion sites, which belong to four domains: Cellphone[2], Electronics[3], Camera[4], and TV[5]. For each domain, there are 1000 labeled posts. The distributions of intention and non-intention posts for each domain are shown in Table 1.

In experiments, we considered one domain as the target domain and three others as source domains. We randomly divided the target domain data into 10 subsets and conducted 10-fold cross-validation. For example, if we consider Cellphone as the target domain, the training data consist of 9/10 data of Cellphone and all data of Electronics, Camera, and TV, while the remaining Cellphone data are used as the test set. The performance of the system was measured using precision, recall, and the $F_1$ score. Let $A$ and $B$ be the set of posts that the system predicted as intention posts and the set of actual intention posts respectively, precision, recall and $F_1$ score are defined as follows:

$$Precision = \frac{|A \cap B|}{|A|},$$

[2]Cellphone: http://www.howardforums.com/forums.php
[3]Electronics: http://www.avsforum.com/avs-vb/
[4]Camera: http://forum.digitalcamerareview.com/
[5]TV: http://www.avforums.com/forums/tvs/

$$Recall = \frac{|A \cap B|}{|B|},$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall}.$$

## 4.2 Methods to Compare

We compared the proposed method with several baselines and state-of-the-art method. Note that all methods used Naive Bayes (NB) as the classifier[6].

- **Baseline1**: This method used 9/10 data in the target domain as training data and the remaining data as test data. We did not use data from source domains in the training process. The purpose of this experiment is to investigate the task in an in-domain setting, without any help from labeled data in source domains.
- **Baseline2**: This model trained the classification system using three source domains and tested on the target domain. It did not use data of the target domain for training. The purpose of this experiment is to investigate the task in a cross-domain setting, without any labeled data in the target domain.
- **Co-Class**: Co-Class is a semi-supervised method for intention detection proposed by Chen et al. [8], which was reported to achieve state-of-the-art performance on intention detection. The method borrows the idea of Co-Training by using two classifiers in an alternating fashion. One classifier is trained on labeled data from source domains, and the other one is trained on unlabeled data in the target domain with labels predicted by the first classifier. The final classification is based on both classifiers using a bootstrapping technique. An important point is that both classifiers use the same feature set selected from the target data.
- **Combined**: This model used labeled data in three source domains and 9/10 data in the target domain to train a Naive Bayesian classification model. The purpose of this experiment is to investigate the performance of the detection system when we combine source and target domains without domain adaptation.
- **Our method**: Our method is similar to the Combined model but using an optimization technique for domain adaptation. In effect, the method combines source and target domain labeled data based on their contribution to final classification accuracy.

We summarize the experimented methods in Table 2. Note that the first three methods, i.e. Baseline1, Baseline2, and Co-Class were run with the same settings as described by Chen et al. [8]. For Combined and Our method, we selected 2500 features on the target domain and 1500 features on each source domain. We conducted 10-fold cross-validation with models that used part of data in the target domain in the training processes i.e. Baseline1, Combined, and Our method.

[6]As shown in Chen et al. [8], Naive Bayes is the suitable method for the task of intention detection in discussion forums.

**Table 3: Experimental results on Cellphone domain**

| Model | Precision(%) | Recall(%) | $F_1$(%) |
|---|---|---|---|
| Baseline1 | **69.04** | 56.21 | 61.97 |
| Baseline2 | 50.81 | 67.93 | 58.14 |
| Co-Class | 54.74 | 69.02 | 61.06 |
| Combined | 58.94 | **75.76** | 66.30 |
| Our method | 66.66 | 69.29 | **67.95** |

**Table 4: Experimental results on Electronics domain**

| Model | Precision(%) | Recall(%) | $F_1$(%) |
|---|---|---|---|
| Baseline1 | 65.02 | 80.36 | 71.88 |
| Baseline2 | 62.37 | **86.43** | 72.45 |
| Co-Class | 63.06 | 85.36 | 72.53 |
| Combined | 69.09 | 80.71 | 74.45 |
| Our method | **73.78** | 76.79 | **75.25** |

**Table 5: Experimental results on Camera domain**

| Model | Precision(%) | Recall(%) | $F_1$(%) |
|---|---|---|---|
| Baseline1 | 70.49 | **86.88** | 77.83 |
| Baseline2 | 86.81 | 72.34 | 78.92 |
| Co-Class | **87.34** | 73.40 | 79.77 |
| Combined | 84.56 | 75.88 | **79.99** |
| Our method | 84.56 | 75.88 | **79.99** |

**Table 6: Experimental results on TV domain**

| Model | Precision(%) | Recall(%) | $F_1$(%) |
|---|---|---|---|
| Baseline1 | **78.85** | 81.39 | 80.10 |
| Baseline2 | 64.03 | 89.35 | 74.60 |
| Co-Class | 73.57 | 87.83 | 80.07 |
| Combined | 73.76 | **90.85** | 81.42 |
| Our method | 78.21 | 88.97 | **83.24** |

## 4.3 Results

Tables 3, 4, 5, 6 summarize experimental results on four domains. In each table, we show Precision, Recall, and $F_1$ scores, averaged over 10 folds for one target domain. Note that $F_1$ score is the most important metrics as it balances Precision and Recall. As can be seen, there is no clear winner between Baseline1 and Baseline2. Each method achieved higher $F_1$ scores in two domains and lower scores in the others. These results suggest that using lots of training data from other domains may be more or less useful than using fewer training data from the same domain, depending on specific cases.

Co-class is comparable or slightly better than either of two baselines. Specifically, co-class achieved $F_1$ scores that are comparable with those of Baseline1 in Cellphone and TV and higher scores in Electronics and Camera, beating both

**Table 2: Methods to compare**

| Model | Training data | Test data | Exp method | Learning algorithm |
|---|---|---|---|---|
| Baseline1 | 9/10 target | 1/10 target | cross-validation | NB |
| Baseline2 | 3 sources | target | One time | NB |
| Co-Class | 3 sources | target | One time | NB, bootstrapping |
| Combined | 3 sources, 9/10 target | 1/10 target | cross-validation | NB |
| Our method | 3 sources, 9/10 target | 1/10 target | cross-validation | NB, optimization |

Baseline1 and Baseline2. A somewhat surprising observation is that a simple combination of labeled data from source and target domain as training data (Combined method) achieved better results than more sophisticated Co-class in all four cases. A possible explanation for the superiority of Combined over Co-class is that the former uses some labeled data of the target domain, and that data is important for classification. Combined method is also consistently better than both Baseline1 and Baseline2. These results render simple Combined as a very competitive method.

From four domains, Camera is the least sensitive to methods used. For this domain, the last three methods achieved nearly the same $F_1$ scores, while the worst performing method is behind by only 2%.

In all cases, except for insensitive Camera domain, our proposed method consistently outperformed the other experimented methods, achieving the most accurate results in terms of $F_1$ scores. The differences in $F_1$ score between our and the second best method, namely Combined, is almost 2% for Cellphone and TV and 0.8% for Electronics. All differences are statistically significant, according to a t-test with the threshold of 0.05. Since both Combined and our method use similar training and test data, we believe the improvement of the later method comes from optimization we have performed to calculate the posterior probability of the Naive Bayes classifier, which is our main contribution in this paper.

**Error Analysis**. We now analyze some cases in which our system made a mistake. We divide errors into two main types: False positives (a non-intent post was predicted as intent) and False negatives (an intent post was predicted as non-intent). For each type, we list several typical examples.

(1) **False positives**. These posts usually contain intention description words, but the meaning does not include purchase intention. Another case is that the post is an experience sharing. The author had bought the product when he posted. Here are some examples.
   - *Who is* ***looking to buy*** *a camera as a semi proffessional camera mix of SLR and normal digital camera and easy to learn I advice to* ***buy*** *Canon SX1 IS as HD viedo, 2.8 LCD Rotate, 20X zoom and other great options, really a nice camera. I been used before Nikon P80 but when I bought Canon SX1 IS there is no compare between both, with different option's I prefer who is* ***looking to buy*** *a camera try with this.* (CAMERA domain)
   - *Can the Pio 436 be ISF calibrated from new or do you really need to run it in? Also who do you contact to calibrate it and* ***how much does it cost?*** (TV domain)
   - *Just sharing my experience. My samsung 3D PS50C680G5KXXU has died after 6 months-sound, no picture, responds to remote, meaning the Y-SUS or Z-sus boards have blown. Dixons has asked to talk to Samsung, and Samsung to Dixons - Would obviously get repaired until 1 year is over. 1,000 3D TV for 1 year, with no hope after- AVOID, and* ***buy*** *a Sony or Panasonic - Most importantly,* ***buy*** *with extended warranty!!!* (TV domain)
   - *I have a LG Env Touch and it was activated last year. I'm* ***planning to change to*** *the Family Plan from the Individual Plan so I can add my wife. Will the change require the $9.99 monthly data plan for my phone?* (CELLPHONE domain)
   - *Are you* ***going to buy*** *the new grand theft auto? I rent most of my games and once i beat them I am done. I think this is how this game will be. What about online will it be worth it? I can keep the game rental for a month I think id rather do that.* (ELECTRONICS domain)

(2) **False negatives**. These posts are usually general descriptions or lack of important words indicating a buying action. Therefore, the purchase intention is unclear. Here are some examples.
   - *does anyone know of any that are specifically made for cleaning plasma's?? I am told the pc screen cleaning cloths should not be used - right or wrong?? your assistance appreciated thanks* (TV domain)
   - *HI folks, can some tell me if there is a better software to use than the one that comes with the camera Cannon SD1200 IS* (CAMERA domain)
   - *I am having trouble updating my new Samsung BDP-1500 player? Can someone talk me through what I do to get the latest firmware and then install it onto the player?? PLEASE HELP!!!* (ELECTRONICS domain)

# 5 CONCLUSION

We have presented in this paper a method for identifying intention posts in discussion forums. Our method leverages data from multi-source domains to improve the performance in the target domain. To combine data from source and target domains efficiently, we have proposed an SGD-based method to optimize key parameters of a Naive Bayesian classification model. The method has been shown to be effective for intention detection on a benchmark dataset consisting of four different domains.

In the future, we would like to investigate the method for other tasks in natural language processing and data mining. Empirical studies on Vietnamese language datasets are also an interesting research direction.

## ACKNOWLEDGMENTS

This work was supported in part by FPT Software.

## REFERENCES

[1] Ngo Xuan Bach, Nguyen Do Hai, and Tu Minh Phuong. 2016. Personalized recommendation of stories for commenting in forum-based social media. *Information Sciences* 352–353 (2016), 48–60.

[2] Ngo Xuan Bach, Vu Thanh Hai, and Tu Minh Phuong. 2016. Cross-domain sentiment classification with word embeddings and canonical correlation analysis. In *Proceedings of the 7th International Symposium on Information and Communication Technology (SoICT)*. 159–166.

[3] Ngo Xuan Bach and Tu Minh Phuong. 2015. Leveraging User Ratings for Resource-poor Sentiment Classification. In *Proceedings of the 19th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES)*. 322–331.

[4] J. Blitzer, M. Dredze, and F Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 440–447.

[5] A. Blum and T Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. In *Proceedings of the eleventh annual conference on Computational learning theory (COLT)*.

[6] Zhiyuan Chen and Bing Liu. 2014. Topic Modeling using Topics from Many Domains, Lifelong Learning and Big Data. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*.

[7] Zhiyuan Chen and Bing Liu. 2017. *Lifelong Machine Learning*. Morgan and Claypool.

[8] Zhiyuan Chen, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Identifying Intention Posts in Discussion Forums. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 1041–1050.

[9] Zhiyuan Chen, Nianzu Ma, and Bing Liu. 2015. Lifelong Learning for Sentiment Classification. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 750–756.

[10] Xiao Ding, Ting Liu, Junwen Duan, and Jian-Yun Nie. 2015. Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2389–2395.

[11] D. Easley and J Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.

[12] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N.A Smith. 2011. Part-of-speech tagging for twitter: annotation, features, and experiments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 42–47.

[13] Mohamed Hamrouna, Mohamed Salah Gouider, and Lamjed Ben Said. 2016. Large Scale Microblogging Intentions Analysis with Pattern Based Approach. In *Proceedings of International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES)*. 1249–1257.

[14] B. Hollerit, M. Kroll, and M. Strohmaier. 2013. Towards linking buyers and sellers: Detecting commercial intent on twitter. In *Proceedings of the World Wide Web Conference (WWW)*. 629–632.

[15] Jing Jiang. 2008. *A literature survey on domain adaptation of statistical classifiers.* Technical Report. University of Illinois Urbana-Champaign.

[16] Chen Xing Li, Ya Jun Du, Jia Liu, Hao Zheng, and Si Da Wang. 2016. A Novel Approach of Identifying User Intents in Microblog. In *Proceedings of International Conference on Intelligent Computing (ICIC)*. 391–400.

[17] L. Li, D. Wang, T. Li, D. Knox, and B Padmanabhan. 2011. Scene: A scalable two-stage personalized news recommendation system. In *Proceedings of the Thirty–fourth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 125–134.

[18] Q. Li, J. Wang, Y. Chen, and Z Lin. 2010. User comments for news recommendation in forum-based social media. *Information Sciences* 180, 24 (2010), 4929–4939.

[19] B Liu. 2012. *Sentiment Analysis and Opinion Mining: Synthesis lectures on human languages technologies*. Morgan and Claypool.

[20] Thai-Le Luong, Thi-Hanh Tran, Quoc-Tuan Truong, Thi-Minh-Ngoc Truong, Thi-Thu Phi, and Xuan-Hieu Phan. 2016. Learning to Filter User Explicit Intents in Online Vietnamese Social Media Texts. In *Proceedings of the Asian Conference on Intelligent Information and Database Systems (ACIIDS)*. 13–24.

[21] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. SemEval-2016 Task 4: Sentiment Analysis in Twitter. In *Proceedings of SemEval-2016*. 1–18.

[22] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q Yang. 2011. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks* 22, 2 (2011), 199–210.

[23] S. J. Pan and Q Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.

[24] Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named Entity Recognition in Tweets: An Experimental Study. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 1524–1534.

[25] J. Wang, Gao Cong, Wayne Xin Zhao, and Xiaoming Li. 2015. Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 339–345.

[26] Xiaojin Zhu. 2008. *Semi-Supervised Learning Literature Survey*. Technical Report. University of Wisconsin - Madison.