Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network

Xiao Ding[†], Ting Liu[†*], Junwen Duan[†], Jian-Yun Nie[‡]
[†]Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{xding, tliu, jwduan}@ir.hit.edu.cn
[‡]University of Montreal, Canada
nie@iro.umontreal.ca

Abstract

Social media platforms are often used by people to express their needs and desires. Such data offer great opportunities to identify users' consumption intention from user-generated contents, so that better tailored products or services can be recommended. However, there have been few efforts on mining commercial intents from social media contents. In this paper, we investigate the use of social media data to identify consumption intentions for individuals. We develop a Consumption Intention Mining Model (**CIMM**) based on convolutional neural network (CNN), for identifying whether the user has a consumption intention. The task is domain-dependent, and learning CNN requires a large number of annotated instances, which can be available only in some domains. Hence, we investigate the possibility of transferring the CNN mid-level sentence representation learned from one domain to another by adding an adaptation layer. To demonstrate the effectiveness of CIMM, we conduct experiments on two domains. Our results show that CIMM offers a powerful paradigm for effectively identifying users' consumption intention based on their social media data. Moreover, our results also confirm that the CNN learned in one domain can be effectively transferred to another domain. This suggests that a great potential for our model to significantly increase effectiveness of product recommendations and targeted advertising.

Introduction

People often post their needs and desires on social media. Mining users' intents, especially commercial intents, from social media is of great interests to product/service providers, such as public company, government, or non-profit institutions, to help them better understand their potential customers and thus improve their offerings or advertising strategy to the general public.

This paper focuses on mining user consumption intention from social media. A post with consumption intention means that it explicitly or implicitly indicates that the user wants to purchase a specific product or service. For example, "*Please recommend me a smartphone of about $600.*" explicitly indicates that the user wants to buy a smartphone. The goal of this work is to automatically infer people's consumption intention from user-generated content and find the appropriate products to satisfy their needs.

Previously, researchers have developed numerous approaches to identify individual intentions. For example, the existing e-commerce recommender systems can recommend the right product to a user, based on what the user has bought or browsed, and what his/her friends have bought (Wang and Zhang 2013). However, most of them do not incorporate the linguistic information of users, and recommend the most *relevant* item rather than what the user *needs* the most. Analysing user sentiment and opinion can also be related to identifying user commercial intents (Wang et al. 2013). For example, "*I like the design of iphone 6!*" implies that the user may want to buy "*iphone 6*". However, mining consumption intention is orthogonal to sentiment analysis as well as opinion mining and thus provides a different perspective.

In this paper, we explore the linguistic relationships between users' consumption intention and their social media data in different domains. To this end, on the one hand, it is necessary to investigate and combine lexical and sentence level clues from diverse syntactic and semantic structures in a sentence. For example, to identify that the sentence "*My baby has infantile anorexia.*" contains user consumption intention, we should leverage the keywords (e.g. "*baby*", "*anorexia*") and the meaning of the entire sentence. On the other hand, we should endow our model with domain adaptation capacity because large-scale annotated data may be available only in some domains.

Inspired by Collobert et al. (2011), we exploit a CNN-based framework termed Consumption Intention Mining Model (**CIMM**) to extract lexical and sentence level features for identifying user consumption intention. CIMM has a convolutional layer that projects each word within a context window to a lexical contextual feature vector. Then, CIMM uses a *max pooling* layer to extract the most salient lexical features to form a fixed-length sentence level feature vector. The sentence level feature vector can be then fed to feedforward neural network layers, which perform affine transformations followed by non-linear functions to extract highly non-linear and effective features.

Moreover, our proposed CNN-based model has an advan-

*Corresponding Author

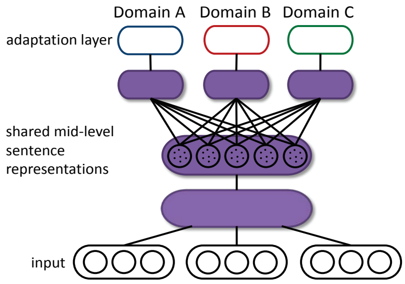

Figure 1: The framework of transfer learning based on CNN

tage for domain adaptation because it learns representations that capture underlying factors, a subset of which may be relevant for each particular domain (Bengio et al. 2013), as shown in Figure 1. Based on this advantage, we propose a domain adaptive CIMM, which transfers sentence representation learned with CIMM on a large-scale dataset in the source domain to the target domain with limited training data. Specifically, we design an approach that uses layers of CIMM trained in the source domain to compute mid-level sentence representation for sentences in the target domain.

Once we know the users' consumption intentions, the next step is to find the appropriate products to satisfy their needs. To address this problem, we propose to extract intention words from sentences with consumption intentions. Intention word refers to the word that can best indicate users' needs. According to the intention word, we can effectively recommend the target product, such as ("*pregnant*", "*maternity dress*"), and ("*anorexia*", "*activated probiotics*").

To the best of our knowledge, our work is the first attempt to investigate and model user consumption intention in a computational way, based on the social media linguistic data of hundreds of thousands of people. The major contributions of the work presented in this paper are as follows.

- Besides sketching potential economic perspectives, our work introduces a novel, an intentional dimension to characterize textual content on social media.
- We propose a new method that exploits domain adaptive CIMM for user consumption intention classification. We report the results that significantly outperform two baseline systems.
- We explore a novel approach based on deep neural network architecture, to extract intention words from the sentence that contains user consumption intention and propose an adaptive approach based on it.

Problem Statement

Consumption Intention. A post with consumption intention means that it explicitly or implicitly states the user may want to purchase some products or services. Our analysis shows that there are two categories of consumption intention - explicit consumption intention and implicit consumption intention. Consider the following two real tweets on Twitter.

Example 1 *I want to buy an air conditioner.*

Example 2 *My wife is pregnant.*

Example 1 explicitly states the user's consumption intention of buying an air conditioner. Example 2 does not show explicitly what product the user wants to buy. However, we can infer that the user may want to buy kids and baby products. This is an example of implicit consumption intention.

In this paper, we mainly focus on the implicit consumption intention for the following reasons.

- Most of user needs are implicitly expressed in Twitter and people may not be entirely consciously aware of (Hollerit et al. 2013). This is confirmed on our own data: we manually labelled 1,000 tweets from Sina Weibo (the most popular microblogging service in China) that contain user consumption intention, of which 625 tweets contain implicit consumption intention and 375 tweets contain explicit consumption intention.
- It is not difficult to recognize explicit consumption intentions. For example, we can use simple rule-based methods and syntactic patterns to identify more than 80% explicit consumption intention on our data. However, identifying implicit consumption intention and inferring the required product need deep semantic analysis techniques, which is explored in this paper.

Our task is as follows: given a sentence, we first determine whether the sentence involves a consumption intention (**Section 3**). If it does, we proceed to extract intention word ("*pregnant*" in Example 2) in it (**Section 4**). The approach of recommending products according to the intention word will be explored in future work.

Note that, for simplification, we mainly use "consumption intention" to mean "implicit consumption intention" in the remainder of this paper.

Consumption Intention Mining Model

In this paper, we formulate consumption intention mining as a classification problem and propose domain adaptive CIMM to solve it. The architecture of CIMM, is illustrated in Figure 2. Our model shares similar intuition with that of Collobert et al. (2011). The CIMM contains a word representation layer that transforms each word into a distributed input representation, a convolutional layer to extract local contextual features, a max pooling layer to form a global feature vector, two sigmoid layers to represent the high-level semantic feature vector of the input word sequence, and an adaptation layer to transfer mid-level sentence representation from the source domain to the target domain. In the following sections, we introduce each layer in detail.

Word Representation Layer

We learn the initial word representation from a large-scale microblog corpus based on the existing word embedding learning algorithm C&W model (Collobert et al. 2011). Given a sentence, we first break it down into several ngrams. Given an ngram "*has infantile anorexia*", C&W replaces the central word with a random word w^r in our dictionary $\mathcal{D}$ (it contains all the words in the training data) and derives a

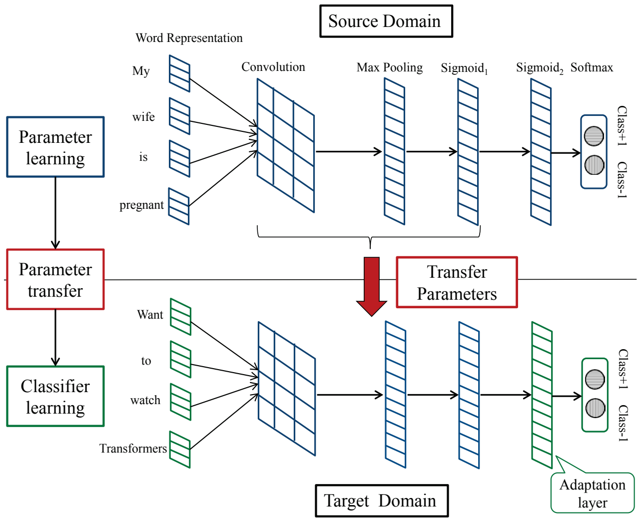

Figure 2: The architecture of CIMM

corrupted ngram "*has* w^r *anorexia*". The training objective function can be optimized by a hinge loss,

$$loss(x, x^r) = max(0, 1 - f(x) + f(x^r)) \tag{1}$$

where x is the original ngram, x^r is the corrupted ngram, $f(x)$ is the one-dimensional scalar representing the language model score of input ngram. At this end, for each word $w \in \mathcal{D}$, an internal d-dimensional feature vector representation is learned: $\mathbf{w} = (w_1, w_2, ..., w_d)$.

Convolutional Layer

The convolution operation can be viewed as feature extraction based on sliding window. It is designed to capture the contextual features for a word. Consider a word at the *i*-th position in a word sequence. The word representation feature vectors of all the context words within a window around w_i are first concatenated to form a context window vector, and then projected to a local contextual feature vector $\mathbf{V_i}$.

The contextual feature vectors extracted at the convolutional layer are local features, one for each word. We need to combine them to obtain a global feature vector with a fixed size. Hence, we use a *max pooling* layer on top of it, which forces the network to retain only the most useful local features produced by the convolutional layer.

Formally, given an input sentence $S \in \mathbb{R}^s$, and $\mathbf{S} = (\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_s})$ (in our word representation layer), the one-dimensional convolution is to take the dot product of the weight vector $\boldsymbol{M} \in \mathbb{R}^n$ with each n-gram (sliding window) in the sentence $\boldsymbol{S}$ to obtain a new sequence $\boldsymbol{Q}$:

$$\boldsymbol{Q}_j = \boldsymbol{M}^T \boldsymbol{S}_{j-n+1:j} \tag{2}$$

Sigmoid Layer

Sigmoid layer is used to extract highly non-linear features. In all sigmoid layers, we use the sigmoid activation function σ. Formally, let the values of the neurons of the output layer be y_{cls} $(cls \in \{+1, -1\})$, its input be net_{cls}, and $\mathbf{y_2}$ be the value vector of the neurons of the last sigmoid layer; then:

$$y_{cls} = f(net_{cls}) = \sigma(\boldsymbol{w_{cls}} \cdot \mathbf{y_2}) \tag{3}$$

where

$$\sigma = f(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

and $\boldsymbol{w_{cls}}$ is the weight vector between the neuron cls of the output layer and the neurons of the second sigmoid layer.
Softmax Layer. There are two output labels in this work, we design the dimension of top layer in CIMM as two and add a softmax layer upon the top layer. Softmax layer is suitable for this scenario because its outputs can be interpreted as conditional probabilities which are useful for intention word extraction, which will be describe in the following section.

Domain Adaptation Layer

The key idea of domain adaptive CIMM is that the internal of the convolutional neural network can act as generic extractor of mid-level sentence representation, which can be pre-trained on the source domain dataset and then re-used on other target domains. To address the problem of domain adaptation, we propose to re-weight the training cost function, which would amount to re-weighting its gradients during training procedure. For the source domain, we use the network architecture introduced above. For the target domain, in order to achieve the transfer, we remove the sigmoid layer $Sigmoid_2$ of the pre-trained network and add an $Adaptation$ layer formed by a fully connected layer that uses the output vector $\mathbf{y_1}$ of the layer $Sigmoid_1$ as input. Note that $\mathbf{y_1}$ is obtained with a complex non-linear function of potentially all input words and may capture mid-level sentence representations as well as their high-level configurations. The adaptation layer computes $\mathbf{y_2} = \sigma(\mathbf{w_2}\mathbf{y_1} + \mathbf{b_2})$, where $\mathbf{w_2}$, $\mathbf{b_2}$ are the trainable parameters.

The parameters of convolutional layer and $Sigmoid_1$ layer are first trained on the source domain, then transferred to the target domain and kept fixed. Only the $Adaptation$ layer is trained on the target domain training data, which requires less training data.

Training

The training algorithm repeats for several iterations over the training data, which is a set of sentences annotated with gold standard labels that indicate whether the sentence contains user consumption intention or not. In each iteration, the procedure is shown in Algorithm 1.

For each iteration, the algorithm calculates a *margin loss* based on two sentence-label pairs $(\overline{s}, \overline{l})$ and $(\widehat{s}, \widehat{l})$. The pair $(\overline{s}, \overline{l})$ denotes the sentence-label pair that has the highest model score among those that are *inconsistent* with the gold standard, while $(\widehat{s}, \widehat{l})$ denotes the one that has the highest model score among those that are *consistent* with the gold standard. If the loss is zero, the algorithm continues to process the next unlabeled sentence. Otherwise, the parameters are updated using back-propagation (BP).

The standard BP algorithm (Rumelhart, Hinton, and Williams 1985) cannot be applied directly, because the stan-

Algorithm 1: Training Process

Input: $(\boldsymbol{s}, \boldsymbol{l})$ a set of labeled sentences; the model $CIMM$
Output: updated model $CIMM'$

1 $\mathbf{S} \leftarrow [s_1, ..., s_m]$ //unlabeled sentence
2 $\mathbf{R} \leftarrow [(s_1, l_1), ..., (s_m, l_m)]$
3 **while** $\mathbf{S} \neq [\,]$ **do**
4 $\quad (\overline{s}, \overline{l}) \leftarrow argmax_{(s,l) \in (\mathbf{S} \times \mathbf{L}/\mathbf{R})} CIMM(s, l)$
5 $\quad (\widehat{s}, \widehat{l}) \leftarrow argmax_{(s,l) \in \mathbf{R}} CIMM(s, l)$
6 $\quad loss \leftarrow max(0, 1 + CIMM(\overline{s}, \overline{l}) - CIMM(\widehat{s}, \widehat{l}))$
7 $\quad$ **if** $loss\ > 0$ **then**
8 $\quad\quad \overline{e} \leftarrow BackPropErr(\langle \overline{s} \rangle, 1 + CIMM(\overline{s}, \overline{l}))$
9 $\quad\quad \widehat{e} \leftarrow BackPropErr(\langle \widehat{s} \rangle, -CIMM(\widehat{s}, \widehat{l}))$
10 $\quad\quad Update(\langle \overline{s} \rangle, \overline{e})$
11 $\quad\quad Update(\langle \widehat{s} \rangle, \widehat{e})$
12 $\quad$ **else**
13 $\quad\quad \mathbf{S} \leftarrow \mathbf{S}/\{\widehat{s}\}, \mathbf{R} \leftarrow \mathbf{R}/(\widehat{s}, \widehat{l})$
14 **return** $CIMM$

dard loss is calculated based on a *unique* input vector; while in our case, $\overline{s}$ and $\widehat{s}$ refer to different sentences, which means that the margin loss in line 6 of Algorithm 1 is calculated based on two different input vectors, denoted by $\langle \overline{s} \rangle$ and $\langle \widehat{s} \rangle$, respectively.

We solve this problem by decomposing the margin loss in line 6 into two parts:

- $1 + CIMM(\overline{s}, \overline{l})$, which is associated with $\langle \overline{s} \rangle$;
- $-CIMM(\widehat{s}, \widehat{l})$, which is associated with $\langle \widehat{s} \rangle$.

In this way, two separate back-propagation updates can be used to update the parameters.

Intention Word Extraction

Intuitively, the consumption intention classification probability of the sentence with intention word should be higher than that of the sentence without it. Based on this observation, we exploit the intention word extraction algorithm that is shown in Algorithm 2.

Given the sentence S with consumption intention, we obtain its classification probability P_s from CIMM. Then, we replace each word w_i of S with a random word ${w_i}^t$ in our dictionary $\mathcal{D}$ and generate a new sentence S'. We classify S' based on CIMM and obtain a new classification probability $P_{s'}$. If the word ${w_i}^t$ results in the minimum of $P_{s'}$, the word w_i is extracted as the intention word of S.

As we focus on mining consumption intention in a specific domain, a natural baseline method to compare with is domain-specific term extraction (Liu et al. 2005), which has proven to be effective. The method works as follows: we compute domain score for each word w_i in the sentence based on the function 5, and rank these words according to their domain scores $Score_{w_i}$. The word with the highest domain score is extracted as the intention word.

$$Score_{w_i} = \frac{NDI(w_i, D_i)}{NDI(w_i, \sum D_j)} \cdot e^{\frac{-1}{(\log NDI(w_i, D_i) + 1) + 1}} \qquad (5)$$

where

Algorithm 2: Intention Word Extraction

Input: trained CIMM; dictionary $\mathcal{D}$; a sentence S of m words $w_1, w_2, ..., w_m$ with consumption intention
Output: Intention word $w^{(I)}$ of S

$U \leftarrow [w_1, w_2, ..., w_m]$
$P_{min} \leftarrow P_s$ //P_s is the classification probability of S
while $U \neq [\,]$ **do**
$\quad$ replace w_i in S by w^t in $\mathcal{D}$, $S' = (w_1, ..., w^t, ..., w_m)$
$\quad$ classify S' based on CIMM, and output its probability $P_{s'}$
$\quad$ **if** $P_{min} > P_{s'}$ **then**
$\quad\quad P_{min} = P_{s'}$
$\quad\quad w^{(I)} = w_i$
$\quad U \leftarrow U/[w_i]$
return $w^{(I)}$

$$NDI(w, D_i) = -\sum_{j=1}^{n_i} P'(d_j|w) \log P'(d_j|w) \qquad (6)$$

$e^{\frac{-1}{(\log NDI(w_i, D_i)+1)+1}}$ is a penalty factor. D_i is a given domain (e.g., kids and baby domain), in which there are n_i documents expressed as $d_1, d_2, ..., d_{n_i}$.

$$P'(d_j|w) = \frac{P(d_j|w)/l_{ij}}{\sum_{j=1}^{n_i} P(d_j|w)/l_{ij}} \qquad (7)$$

where l_{ij} is the length of d_j in the domain D_i.

$$P(d_j|w) = \frac{count(w, d_j)}{count(w, D_i)} \qquad (8)$$

where $count(w, d_j)$ is the number of occurrences of the word w in the document d_j, and $count(w, D_i)$ is the number of occurrences of the word w in the domain D_i.

Intention word is useful for recommending products to the user. For example, once the intention word is identified, we could construct an external knowledge base that maps the intention word with a list of products. This aspect is out of the scope of this paper and we leave it to future research.

Experiments

Data Description

We collected a large-scale microblog corpus for training initial word embedding from Sina Weibo. The corpus contains 20 million posts, 76 million sentences and 1.3 billion words.

To our knowledge, there is no public corpus for evaluating the task of consumption intention mining. Hence, we constructed a manually annotated sub corpus. Notice that annotating the posts randomly selected from Sina Weibo has little chance of covering user consumption intention (Hollerit et al. 2013). We therefore mainly focus on the consumption intention of kids and baby domain and movie domain. For all sentences in our data, two annotators are asked to annotate whether it contains user consumption intention. If it does, two annotators proceed to label the intention word of it.

The agreement between our two annotators, measured using Cohen's Kappa Coefficient (Cohen 1968), is substantial (kappa = 0.8 for consumption intention classification, and kappa = 0.85 for intention word extraction). The annotated corpus contains 5000 positive and 5000 negative instances

Table 1: Experimental results of consumption intention classification

Domain	Method	Accuracy
kids and baby	CIMM	94.52%
	Word Embedding + SVM	85.77%
	Bag-of-words + SVM	73.56%
movie	domain adaptive CIMM	85.28%
	CIMM (without transfer)	73.05%
	Word Embedding + SVM	68.17%
	Bag-of-words + SVM	64.33%

for the kids and baby domain, and 5000 positive and 5000 negative instances for the movie domain (we only use 500 positive and 500 negative instances to simulate the situation where a limited number of annotated instances are available, thus transfer is necessary). The corpus is split into training, development and test dataset, 4/5 of which are used as the training data, 1/10 for development and 1/10 for test.

Consumption Intention Classification

In this paper, we formulate consumption intention mining as a binary classification problem, i.e., whether the sentence contains user consumption intention or not. We use *accuracy* as the evaluation metric.

Baseline Methods. We compare our approach with the following baseline methods.

1. Word embedding + SVM: Word embedding features can describe deep semantic information of the sentence. Support Vector Machines (SVM) is a state-of-the-art classification model (Hearst et al. 1998). This baseline method only use embedding without our CIMM framework.
2. Bag-of-words + SVM: The bag-of-words features and SVM are widely used for classification. We use it as another baseline. Linear kernel of SVM is used in this paper.

Results and Analysis. We first train CIMM on the source domain (kids and baby domain), and then apply the mid-level feature transfer scheme to the target domain (movie domain). Table 1 shows the accuracy of the baseline systems as well as the proposed CIMM on consumption intention classification. The following observations can be made:

(1) Word embedding is a better choice for representing the semantic information of the sentence. Given the same classification model (SVM), the word embedding based method achieves consistently better performance than the bag-of-words based method. This is likely due to the following two reasons. **First**, the bag-of-words representation treats each word as a one-hot vector. It has the same length as the size of the vocabulary, and only one dimension is one, with all others being zero. Due to its sparsity, it cannot capture complex linguistic phenomena of words. **Second**, word embedding uses a continuous real-vector for representing each word, i.e. a learned distributed feature vector which allows each training sentence to inform the model about an exponential number of semantically neighboring sentences. Therefore, word embedding features encode more semantic information than the bag-of-words features.

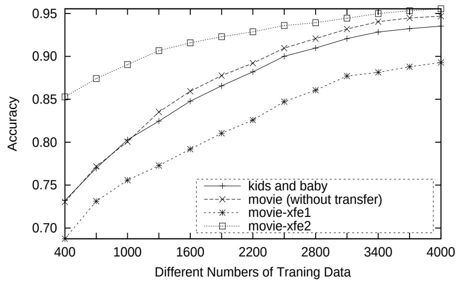

Figure 3: Experimental results of different numbers of training data on the source and target domain

(2) Our proposed CIMM achieves better performance than SVM model, because it can make effective use of the extracted local contextual features (via the convolutional layer) and global contextual features (via the max pooling), and learn non-linear relationships between features and the output (via the sigmoid layer).

(3) Domain adaptive CIMM can effectively learn mid-level sentence representation and transfer it from the source domain to the target domain. Table 1 shows that, for the movie domain, the performance drops largely by 12% if CIMM is only trained with the target domain data without transfer. Below, we analyze some factors that can influence the effectiveness of our method.

Experiments with Different Numbers of Training Data. To understand the learning curve of our proposed domain adaptive CIMM, we use 400, 700, 1000, ..., 4000 instances from the training data, and evaluate the following four tasks.

- **kids and baby**: We evaluate the performance of CIMM on the kids and baby domain with different numbers of training data.
- **movie (without transfer)**: We evaluate the performance of CIMM on the movie domain with different numbers of training data.
- **movie-xfe1**: We evaluate the performance of domain adaptive CIMM on the movie domain with different numbers of training data on the kids and baby domain, while the number of training data on the movie domain is fixed as 400 instances. By this experiment, we want to test the impact of more training data in the source domain.
- **movie-xfe2**: We evaluate the performance of domain adaptive CIMM on the movie domain with different numbers of training data, while the number of training data on the kids and baby domain is fixed as 4000 instances. This experiment is intended to test the impact of more training data in the target domain.

Figure 3 shows that the performance of CIMM increases with the amount of training data on the kids and baby domain (kids and baby curve) and the movie domain (movie curve), respectively. As the training data of the movie domain is limited (400 instances), it is not surprising that the performance of CIMM is quite poor. However, the trained CIMM on the kids and baby domain can provide mid-level sentence representations, and the adaptation layer of CIMM can use it and the limited training data to learn a new classi-

Table 2: Experimental results of intention word extraction

Method	Accuracy
Our method	71.51%
Baseline	46.12%

Table 3: Examples of intention words

Intention words		Translation	
厌食	催奶	anorexia	promote lactation
消毒	补钙	disinfect	calcium supplement
怀孕	学步	pregnancy	learn to walk
胎教	奶癣	antenatal training	infantile eczema
上映	抢票	release	ticket competition
首映	巨星	premiere	superstar

fier for the movie domain. This is shown by the comparison between movie and movie-xfe2 curve. In addition, experimental results on movie-xfe1 show that the performance of movie-xfe1 increases with the amount of training data on the kids and baby domain, because there is more knowledge transferred from this domain. Meanwhile, experimental results on movie-xfe2 show that the performance of movie-xfe2 increases with the amount of training data in the target domain. This is not surprising, because more training data in the target domain results in a better adaptation of CIMM.

Intention Word Extraction

Results and Analysis. As the task has not been investigated in previous work, there is no existing method to compare with. Hence, we use the approach of domain-specific term extraction introduced in Section 4 as the baseline method. Table 2 shows our results compared to the baseline method. As expected, our approach largely outperforms the baseline method. The experimental result confirms our observation that intention word can best indicate users' needs. The possible reasons are as follows:

(1) Domain-specific terms do not necessary imply user consumption intention. For example, "*baby*" is a domain-specific term in kids and baby domain, but it does not imply any specific intention.

(2) Our approach is consumption-intention-oriented. It is based on the output of consumption intention classification. We extract the word that can influence the most the classification result of user consumption intention. Hence, if a domain-specific word is not intention-related, it will not be extracted by our approach.

We extracted in total 245 intention words for the kids and baby domain and 78 intention words for the movie domain. Some of them are shown in Table 3.

Related Work

Mining users' intention has attracted much attention (Goldberg et al. 2009; Kröll and Strohmaier 2009; Fu and Liu 2013; Zhang and Pennacchiotti 2013; Zhang et al. 2014). Our work draws from the research area of psychology, marketing and natural language processing. Maslow (1943) suggests that human behaviors are motivated by expression and fulfilment of these deeply held needs, which people may not be entirely consciously aware of. Maslow theory has a significant impact on the research of human needs, and can serve as the theoretical basis of user consumption intention mining. However, Psychological research lacks seriously sound experimental data (Maslow 1954).

To address this issue, researchers focus on identifying online users' commercial intents which can fully use the prevalence of online data. Mining user intents (especially, commercial intents) from search queries has been an important research problem in the past (Jansen 2007; Ashkan and Clarke 2009a). Dai et al. (2006) first proposed to identify search queries that contain online commercial intention. Since queries do not carry much information, much research extends the queries by including information extracted from search logs (Strohmaier and Kröll 2012), click through behavior data (Ashkan and Clarke 2009b), and users' mouse movements behaviors data (Guo and Agichtein 2010).

However, the length of query is limited. In contrast, in microblog discussions, we have longer descriptions. Recently, microblogs have been one of the most popular social networking platform which provide the intents and opinions of diverse groups of people at low cost. Yang and Li (2013) investigated the use of social media data to identify fundamental needs for individuals through a crowd-sourced study. Wang et al. (2013) mainly focused on identifying trend-driven commercial intents from microblogs. Hollerit et al. (2013) proposed to identify tweet-level commercial intents and link buyers and sellers. Comparing with previous work, the main differences of this paper are as follows. **First**, all above studies focus on explicit commercial intents. To the best of our knowledge, this paper is the first to investigate the problem of mining user implicit consumption intention. **Second**, the problem of learning transfer between domains has not been addressed in the previous papers on commercial intent mining. We proposed a method for it and showed its effectiveness. **Third**, in addition to recognizing consumption intents, we further extract intention words. This is a new task that has not been investigated in the above studies.

In addition to the above work, recommendation systems also have been extensively studied for mining online user commercial intents (Resnick and Varian 1997; Bai 2011). The main difference of our work from recommendation algorithms is that we exploit the linguistic relationships between people's consumption intention and their social media content using CNN. This paper is orthogonal to the study of recommendation systems. As a result, our model can be combined with the recommendation algorithms.

Conclusions

In this paper, we make the first attempt to mine user implicit consumption intention from social media. We propose the domain adaptive CIMM, a complete framework to identify consumption intention and extract intention word based on the output of CIMM. Through experiments, we show that our model is able to infer user consumption intention in different domains and the proposed adaptation framework is effective. Our work is not designed to replace traditional methods of user needs detection proposed by psychologists.

Instead, we believe that these methods can complement each other to enable a better and more comprehensive understanding of how users express their needs, especially, consumption intention. This is not only useful for advancing the understanding of user needs in psychological science, but also essential to personalized recommendation and advertising.

Acknowledgments

We thank the anonymous reviewers for their constructive comments, and gratefully acknowledge the support of the National Basic Research Program (973 Program) of China via Grant 2014CB340503, the National Natural Science Foundation of China (NSFC) via Grant 61133012 and 61202277.

References

Ashkan, A., and Clarke, C. L. 2009a. Characterizing commercial intent. In *Proceedings of the 18th ACM conference on Information and knowledge management*, 67–76. ACM.

Ashkan, A., and Clarke, C. L. 2009b. Term-based commercial intent analysis. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 800–801. ACM.

Bai, X. 2011. Predicting consumer sentiments from online text. *Decision Support Systems* 50(4):732–742.

Bengio, Y.; Courville, A.; and Vincent, P. 2013. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(8):1798–1828.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* 70(4):213.

Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research* 12:2493–2537.

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

Fu, B., and Liu, T. 2013. Weakly-supervised consumption intent detection in microblogs. *Journal of Computational Information Systems* 6(9):2423–2431.

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

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

Hearst, M. A.; Dumais, S.; Osman, E.; Platt, J.; and Scholkopf, B. 1998. Support vector machines. *Intelligent Systems and their Applications, IEEE* 13(4):18–28.

Hollerit, B.; Kröll, M.; and Strohmaier, M. 2013. Towards linking buyers and sellers: detecting commercial intent on twitter. In *Proceedings of the 22nd international conference on World Wide Web companion*, 629–632. International World Wide Web Conferences Steering Committee.

Jansen, B. J. 2007. The comparative effectiveness of sponsored and nonsponsored links for web e-commerce queries. *ACM Transactions on the Web (TWEB)* 1(1):3.

Kröll, M., and Strohmaier, M. 2009. Analyzing human intentions in natural language text. In *Proceedings of the fifth international conference on Knowledge capture*, 197–198. ACM.

Liu, T.; Wang, X.; Guan, Y.; Xu, Z.-m.; et al. 2005. Domain-specific term extraction and its application in text classification. In *8th Joint Conference on Information Sciences*, 1481–1484.

Maslow, A. H. 1943. A theory of human motivation. *Psychological review* 50(4):370.

Maslow, A. H. 1954. Personality and motivation. *Harlow, England: Longman* 1:987.

Resnick, P., and Varian, H. R. 1997. Recommender systems. *Communications of the ACM* 40(3):56–58.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1985. Learning internal representations by error propagation. Technical report, DTIC Document.

Strohmaier, M., and Kröll, M. 2012. Acquiring knowledge about human goals from search query logs. *Information Processing & Management* 48(1):63–82.

Wang, J., and Zhang, Y. 2013. Opportunity model for e-commerce recommendation: right product; right time. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 303–312. ACM.

Wang, J.; Zhao, W. X.; Wei, H.; Yan, H.; and Li, X. 2013. Mining new business opportunities: Identifying trend related products by leveraging commercial intents from microblogs. In *Proceedings of EMNLP*, 1337–1347. Seattle, Washington, USA: Association for Computational Linguistics.

Wang, H.; Qian, G.; and Feng, X.-Q. 2013. Predicting consumer sentiments using online sequential extreme learning machine and intuitionistic fuzzy sets. *Neural Computing and Applications* 22(3-4):479–489.

Yang, H., and Li, Y. Identifying user needs from social media.

Zhang, Y., and Pennacchiotti, M. 2013. Predicting purchase behaviors from social media. In *Proceedings of WWW*, 1521–1532. International World Wide Web Conferences Steering Committee.

Zhang, F.; Yuan, N. J.; Lian, D.; and Xie, X. 2014. Mining novelty-seeking trait across heterogeneous domains. In *Proceedings of the 23rd international conference on World wide web*, 373–384. International World Wide Web Conferences Steering Committee.