# An Unsupervised Cross-Lingual Topic Model Framework for Sentiment Classification

Zheng Lin, Xiaolong Jin, Xueke Xu, Yuanzhuo Wang, Xueqi Cheng, Weiping Wang, and Dan Meng

***Abstract*—Sentiment classification aims to determine the sentiment polarity expressed in a text. In online customer reviews, the sentiment polarities of words are usually dependent on the corresponding aspects. For instance, in mobile phone reviews, we may expect the *long* battery time but not enjoy the *long* response time of the operating system. Therefore, it is necessary and appealing to consider aspects when conducting sentiment classification. Probabilistic topic models that jointly detect aspects and sentiments have gained much success recently. However, most of the existing models are designed to work well in a language with rich resources. Directly applying those models on poor-quality corpora often leads to poor results. Consequently, a potential solution is to use the cross-lingual topic model to improve the sentiment classification for a target language by leveraging data and knowledge from a source language. However, the existing cross-lingual topic models are not suitable for sentiment classification because sentiment factors are not considered therein. To solve these problems, we propose for the first time a novel cross-lingual topic model framework which can be easily combined with the state-of-the-art aspect/sentiment models. Extensive experiments in different domains and multiple languages demonstrate that our model can significantly improve the accuracy of sentiment classification in the target language.**

***Index Terms*—Cross-language, sentiment classification, topic model.**

## I. Introduction

In REVIEWS on products or services, customers usually talk about aspects of a thing (e.g., the rooms or the location of a hotel) rather than the thing itself as a whole and users may be more interested in some particular aspects when making a purchasing or booking decision. In addition, it is well recognized that the polarities of opinion words vary significantly from aspect to aspect. Therefore, it is necessary and appealing to consider aspects in sentiment classification.

Probabilistic topic models [1] that jointly detect aspects and sentiments have been widely investigated in recent years, because they provide an unsupervised way for sentiment classification, such as the Joint Sentiment/Topic (JST) model [2] and the Aspect and Sentiment Unification Model (ASUM) [3]. However, most of the existing probabilistic topic models are suitable for a specific language and cannot be readily applied to other languages, as they usually require external resources, including sentiment lexica and rich corpora, which may not be publicly available for other languages. In this paper, we intend to improve sentiment classification of a target language by leveraging data and knowledge available in a source language by virtue of a cross-lingual mechanism.

Some pilot studies on cross-lingual sentiment classification depend on machine translation [4], while existing machine translation systems are not powerful enough to provide accurate translations due to a variety of reasons. For instance, a machine translation system usually generates only one best result, which may not be suitable for the case at hand. Moreover, although there have been some studies conducted for cross-lingual sentiment classification, most of them are designed on document level instead of aspect level (e.g., [5]). The cross-lingual topic model provides a potential solution to help the aspect-level sentiment classification in a target language by transferring knowledge from a source language. However, existing cross-lingual topic models cannot be applied in sentiment classification, because they do not take sentiment into account.

To address the above problems, in this paper we propose an unsupervised cross-lingual topic model framework. Note that most of the existing cross-lingual sentiment classification models are supervised or semi-supervised, because they require the data in the source language to be labeled. However, our model is unsupervised, as the corpora in both source and target languages are unlabeled. We then incorporate two up-to-date aspect/sentiment models, JST and ASUM, into the proposed framework and obtain two cross-lingual topic models, called Cross-Lingual JST (CLJST) and Cross-Lingual ASUM (CLASUM), for sentiment classification at aspect level. Compared to the existing models, the prominent advantage of the proposed models is that they do not require parallel corpora, machine translation systems, and labeled sentiment texts. In general, the contributions of this study can be summarized as follows:

1) We propose a novel cross-lingual topic model framework for bridging two different languages. With this cross-lingual framework, the existing monolingual

Manuscript received June 23, 2015; revised November 02, 2015 and December 11, 2015; accepted December 12, 2015. Date of publication December 23, 2015; date of current version February 16, 2016. This work was supported in part by the 973 Program of China under Grants 2012CB316303 and 2014CB340401, in part by the National Natural Science Foundation of China under Grants 61502478, 61572469, 61572473, 61232010, and 61173008, in part by the National High-Tech Research and Development Program of China under Grant 2013AA013204, and in part by National HeGaoJi Key Project 2013ZX01039-002-001-001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Min Zhang.

Z. Lin, W. Wang, and D. Meng are with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China (e-mail: linzheng@iie.ac.cn; wangweiping@iie.ac.cn; mengdan@iie.ac.cn).

X. Jin, X. Xu, Y. Wang, and X. Cheng are with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100864, China (e-mail: jinxiaolong@ict.ac.cn; xuxueke@ict.ac.cn; wangyuanzhuo@ict.ac.cn; cxq@ict.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2015.2512041

aspect/sentiment models can be easily transferred to the multilingual scenario.

2) We specifically present two unsupervised cross-lingual joint aspect/sentiment models, namely, CLJST and CLASUM, by integrating the state-of-the-art models JST and ASUM into the above cross-lingual topic model framework, which can improve sentiment classification for a target language by exploiting correspondences with the source language.
3) Through extensive experiments on real datasets in different domains and different languages and comparison with existing state-of-the-art models, we examine the effectiveness and validity of the proposed cross-lingual topic model framework and the CLJST and CLASUM models. Experimental results demonstrate that the proposed models can be successfully applied to practical sentiment classification applications and improve the accuracy of sentiment classification in target languages.

The rest of this paper is organized as follows. Section II introduces the related work of this study; Section III presents the proposed cross-lingual topic model framework and two cross-lingual aspect/sentiment models CLJST and CLASUM; Section IV presents experimental results. Particularly, through comparison with existing models, we validate the effectiveness of the proposed cross-lingual framework and models. Finally, Section V concludes the paper.

## II. Related Work

In this section, we first give an overview of sentiment classification and then present related studies on monolingual aspect-level sentiment classification and cross-lingual topic model.

### A. Sentiment Classification

Sentiment classification has been investigated in various domains and different languages. According to the training mode, existing methods can be categorized into three types, namely, supervised, unsupervised, and semi-supervised. Supervised methods usually regard polarity identification as a classification task and use a labeled corpus to train a sentiment classifier. Unsupervised methods directly learn a sentiment classification model from an unlabeled corpus. Semi-supervised methods make use of both labeled and unlabeled data for training, typically a small amount of labeled data with a large amount of unlabeled data. As compared to unsupervised and semi-supervised methods, supervised methods may achieve high accuracy on sentiment classification. However, the performance of supervised methods is significantly affected by the quality and quantity of annotated training data, while annotating data is costly and time-consuming.

In recent years, there emerge some new research hotspots, including concept-level [6], [7], short-text [8], [9] and cross-lingual sentiment classification [10]. For instance, Poria et al. [11] introduced a paradigm to concept-level sentiment classification that merges linguistics, common-sense computing, and machine learning for improving the accuracy of polarity detection. Kiritchenko et al. [12] described the state-of-the-art system that detects the sentiment of short informal textual messages. Severyn et al. [13] improved twitter sentiment classification using deep convolutional neural networks. Li et al. [14] presented a data quality controlling approach to select high-quality samples from the source language for cross-lingual sentiment classification. Chen et al. [15] proposed a knowledge validation model in transfer learning in order to reduce noisy data caused by machine translation errors or inevitable mistakes made by the source language sentiment classifier. However, all these cross-lingual methods are supervised or semi-supervised. For some language pairs, a labeled corpus in the source language may be unavailable. Therefore, in this paper we intend to develop an unsupervised cross-lingual method for sentiment classification.

### B. Aspect-Level Sentiment Classification

Sentiment classification can be carried out at different levels, including document level, sentence level and aspect level.

Brody and Elhadad [16] first detected aspects using Local Latent Dirichlet Allocation (Local-LDA) and then identified aspect-sensitive polarities of adjectives using polarity propagation based on an aspect-specific polarity graph. However, they simply selected adjectives as aspect-specific opinion words, which cannot cover all sentiment words. Lu et al. [17] proposed an optimization framework to combine different signals for determining aspect-aware sentiment polarities. However, the aspects are predefined with manually selected keywords, and the sentiment words are extracted beforehand. Unlike the above studies, our model can detect aspects and aspect-specific sentiments in multiple languages in a unified framework.

There also have been unified models developed by incorporating sentiment into classic topic models to joint detect aspects and sentiments. The Joint Sentiment/Topic (JST) model [2] is the first LDA based model considering topics and sentiments simultaneously. The Aspect and Sentiment Unification Model (ASUM) [3] follows a similar generative process to JST except that a sentiment-topic pair is selected for a single sentence, rather than for a word as JST. Kim et al. [18] proposed a hierarchical aspect sentiment model (HASM) to discover a hierarchical structure of aspect-based sentiments from unlabeled online reviews. Mukherjee et al. [19] proposed a Joint Author Sentiment Topic Model (JAST) that takes author preferences and writing style into account. Tan et al. [20] proposed an LDA based model, Foreground and Background LDA (FB-LDA), to distill foreground topics and filter out longstanding background topics, which can give potential interpretations of the sentiment variations. In general, all of the above models are designed on a monolingual setting.

### C. Cross-Lingual Topic Model

The key to a cross-lingual topic model is to find a bridge for connecting different languages. Existing models usually connect two languages via parallel or comparable corpora [21], [22]. Topic models on unaligned text in multiple languages would allow for applications on a broader class of corpora. Boyd-Graber and Blei [23] developed the MUltilingual TOpic (MUTO) model to exploit matching across languages at term level to detect multilingual latent topics from unaligned text. The knowledge of matching can be derived from different sources (e.g., bilingual dictionaries). However, it does not consider sentiment factors, and thus cannot help cross-lingual sentiment classification. In [24], a topic model based method was proposed to group aspects from different languages into aspect categories, but this model cannot capture the aspect-aware sentiments because aspects and sentiments are not modeled in a unified way. Boyd-Graber and Resnik [25] proposed a holistic model for multilingual sentiment classification based on LDA, which is supervised, while the models proposed in this paper are unsupervised. Lin et al. [26] proposed a cross-lingual joint aspect/sentiment model (CLJAS) for sentiment classification which is unsupervised, but parameter-adjusting is an onerous task since CLJAS has too many parameters. Furthermore, CLJAS aims to detect the aspect-specific opinion words, whereas CLJST and CLASUM aim to detect sentiment-coupled aspects.

## III. The Proposed Models

In this section, we first present the Cross-Lingual LDA (CLLDA) model, which only detects topics without considering sentiments. Next, the two existing aspect/sentiment models, JST and ASUM, are incorporated into CLLDA to capture important aspects that are closely coupled with sentiments in both languages.

### A. Cross-Lingual LDA

Under the classic LDA, the procedure of generating each word in a document can be divided into two steps: First, choosing a distribution over a mixture of $K$ topics; Next, picking up a topic randomly from the topic distribution and drawing a word from this topic according to its word probability distribution. CLLDA extends LDA in order to model two different languages together. Its key idea is two-fold: First, for all reviews from the same domain, we assume that they share the same topic distribution despite they are in different languages. In this way, the words in the reviews of different languages but from the same domain tend to be assigned with the same topics. Second, a bilingual dictionary is adopted to connect the topics in two languages. Through translation, we can exploit word co-occurrences cross the boundary of different languages to learn semantically aligned topics. The graphical representation of CLLDA is shown in Fig. 1 and its formal generative process is presented in Algorithm 1. The notation used for all models in this paper is explained in Table I, while the notation specified for CLLDA is presented in Table II.

In CLLDA, each topic $z$ corresponds to a pair of multinomial distributions over words in the source and target languages (i.e., $\phi^{sc,z}$ and $\phi^{tg,z}$), respectively. Unlike the existing topic models where the words in documents are directly generated from multinomial distributions over words, we introduce a per-word intermediate word variable $y$ for generating an observed word $w$. We also employ a per-word binary switch, $x$, to choose between $\phi^{sc,z}$ and $\phi^{tg,z}$ when drawing $y$. Taking a document in the target language for example, if $x = sc$, $y$ will be drawn from a distribution over the words of the source language and $w$ will be generated by a translation from $y$; if $x = tg$, $y$ will be drawn from a distribution over the words of the target language and $w$ directly equals to $y$.

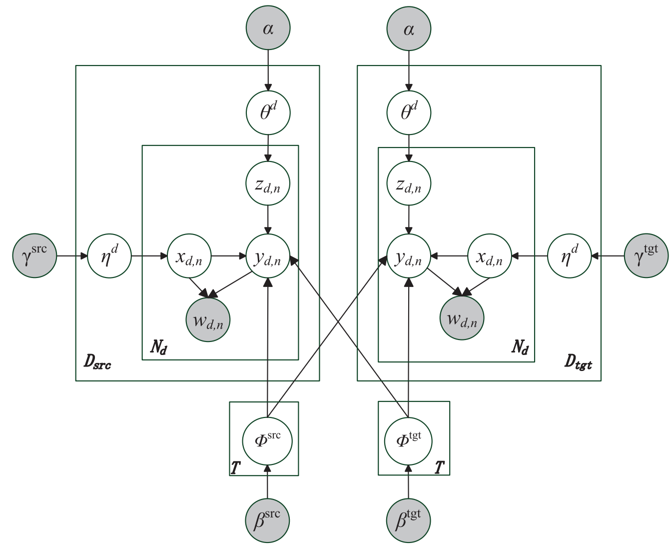



Fig. 1. The graphical representation of the CLLDA model.

---

**Algorithm 1.** The generative process of the CLLDA model.

---

For each topic $z$:
1. Draw a multinomial distribution of the source language: $\phi^{sc,z} \sim Dir(\beta^{sc})$;
2. Draw a multinomial distribution of the target language: $\phi^{tg,z} \sim Dir(\beta^{tg})$;

For each document $d$ in the corpus of the source language:
1. Draw a multinomial distribution, $\theta^d \sim Dir(\alpha)$;
2. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{sc})$;
3. For each word $w_{d,n}$ in document $d$:
   (1) Draw a topic $z_{d,n} \sim \theta^d$;
   (2) Draw a language label $x_{d,n} \sim \eta^d$;
      a. If $x_{d,n} = sc$, draw an intermediate word in the source language $y_{d,n} \sim \phi^{sc,z_{d,n}}$; generate a word directly: $w_{d,n} = y_{d,n}$
      b. If $x_{d,n} = tg$, draw an intermediate word in the target language $y_{d,n} \sim \phi^{tg,z_{d,n}}$; generate a word by translating from $y_{d,n}$ according to the translation probability, $\tau^{sc,w_{d,n}}_{tg,y_{d,n}}$;

For each document $d$ in the corpus of the target language:
1. Draw a multinomial distribution, $\theta^d \sim Dir(\alpha)$;
2. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{tg})$;
3. For each word $w_{d,n}$ in document $d$:
   (1) Draw a topic $z_{d,n} \sim \theta^d$;
   (2) Draw a language label $x_{d,n} \sim \eta^d$;
      a. If $x_{d,n} = sc$, draw an intermediate word in the source language $y_{d,n} \sim \phi^{sc,z_{d,n}}$; generate a word by translating from $y_{d,n}$ according to the translation probability, $\tau^{tg,w_{d,n}}_{sc,y_{d,n}}$;
      b. If $x_{d,n} = tg$, draw an intermediate word in the target language $y_{d,n} \sim \phi^{tg,z_{d,n}}$; generate a word directly: $w_{d,n} = y_{d,n}$.

---

Hyperparameter $\gamma$ in CLLDA indicates the prior knowledge on choosing between the source and target languages. Specifically, asymmetric hyperparameters $\gamma^{sc/tg}_{sc/tg}$ enable knowledge transfer from the auxiliary reviews in the source language

TABLE I
THE NOTATION FOR ALL MODELS IN THE PAPER

| | |
|---|---|
| *D/M/N/T* | the number of reviews/sentences/words/aspects |
| $V_{sc/tg}$ | the vocabulary size of the source/target language |
| $w$ | a word |
| $z$ | a topic/aspect |
| $x$ | the language label, sc (source) vs. tg (target) |
| $l$ | the sentiment label, ps (positive) vs. ng (negative) |
| $y$ | an intermediate variable indicating a word in either the source or target language |
| $\phi$ | the multinomial distribution over words |
| $\theta$ | the multinomial distribution over aspects |
| $\pi$ | the binomial distribution over sentiment labels |
| $\eta$ | the binomial distribution over language labels |
| $\tau_{tg,y}^{sc,w}/\tau_{sc,y}^{tg,w}$ | the probability of the translation from a target (source) word $y$ to the given source (target) word $w$ |
| $\alpha$ | the Dirichlet prior vector for $\theta$ |
| $\beta$ | the Dirichlet prior vector for $\phi$ |
| $\gamma$ | the Beta prior vector for $\eta$ |
| $\mu$ | the Beta prior vector for $\pi$ |
| **w** | the list of words of the entire corpora including the source and target languages |

TABLE II
THE NOTATION SPECIFIED FOR THE CLLDA MODEL

| | |
|---|---|
| $\mathbf{z}/\mathbf{z}_{\neg(d,n)}$ | the topic assignments to all words (excluding the $n$th word in review $d$) |
| $\mathbf{x}/\mathbf{x}_{\neg(d,n)}$ | the language label assignments to all words (excluding the $n$th word in review $d$) |
| $\mathbf{y}/\mathbf{y}_{\neg(d,n)}$ | the intermediate variable assignments to all words (excluding the $n$th word in review $d$) |
| $c^d(c_k^d)$ | the number of words (assigned topic $k$) in review $d$ |
| $c_{sc/tg}^d$ | the times that any word is assigned language label $sc/tg$ in review $d$ |
| $c^{sc/tg,k}$ | the times that any word is assigned language label $sc/tg$ and topic $k$ |
| $c_t^{sc/tg,k}$ | the times that any word is assigned language label $sc/tg$ and topic $k$ when the corresponding intermediate variable is assigned $t$ |

to the reviews in the target language. In more detail, for reviews in the source language, if we set $\gamma_{sc}^{sc} > \gamma_{tg}^{sc}$, they can contribute more to topic modeling of reviews in the target language; for reviews in the target language, if we set $\gamma_{sc}^{tg} > \gamma_{tg}^{tg}$, they can adopt more knowledge from the source language.

In CLLDA, Gibbs sampling is employed to estimate the latent variables and distributions. In order to obtain the distributions of $\phi$, $\theta$ and $\eta$, we first estimate the posterior distribution over $z$, $x$, $y$, i.e., the assignment of words to topics, language labels, and intermediate words. Specifically, $z_{d,n}$, $x_{d,n}$ and $y_{d,n}$ for the $n$th word in document $d$ will be jointly sampled from a distribution given the current values of all the other variables.

For the source language: If $x_{d,n} = sc$,

$$P\big(z_{d,n}=k, y_{d,n}=w_{d,n}, x_{d,n}=sc|\mathbf{w}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}\big) \propto \frac{c_{sc}^d + \gamma_{sc}^{sc}}{c^d + \gamma_{tg}^{sc} + \gamma_{sc}^{sc}} \frac{c_k^d + \alpha_k}{c^d + \sum_{k'} \alpha_{k'}} \frac{c_{w_{d,n}}^{sc,k} + \beta_{w_{d,n}}^{sc}}{c^{sc,k} + \sum_{w' \in V_{sc}} \beta_{w'}^{sc}}. \quad (1)$$

If $x_{d,n} = tg$ and the translation word of $y_{d,n}$ is $t$,

$$P\big(z_{d,n} = k, y_{d,n} = t, x_{d,n} = tg|\mathbf{w}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}\big) \propto \frac{c_{tg}^d + \gamma_{tg}^{sc}}{c^d + \gamma_{sc}^{sc} + \gamma_{tg}^{sc}} \frac{c_k^d + \alpha_k}{c^d + \sum_{k'} \alpha_{k'}} \tau_{tg,t}^{sc,w_{d,n}} \frac{c_t^{tg,k} + \beta_t^{tg}}{c^{tg,k} + \sum_{w' \in V_{tg}} \beta_{w'}^{tg}}. \quad (2)$$

For the target language: If $x_{d,n} = tg$,

$$P\big(z_{d,n}=k, y_{d,n}=w_{d,n}, x_{d,n}=tg|\mathbf{w}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}\big) \propto \frac{c_{tg}^d + \gamma_{tg}^{tg}}{c^d + \gamma_{tg}^{tg} + \gamma_{sc}^{tg}} \frac{c_k^d + \alpha_k}{c^d + \sum_{k'} \alpha_{k'}} \frac{c_{w_{d,n}}^{tg,k} + \beta_{w_{d,n}}^{tg}}{c^{tg,k} + \sum_{w' \in V_{tg}} \beta_{w'}^{tg}}. \quad (3)$$

If $x_{d,n} = sc$ and the translation word of $y_{d,n}$ is $t$,

$$P\big(z_{d,n} = k, y_{d,n} = t, x_{d,n} = sc|\mathbf{w}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}\big) \propto \frac{c_{sc}^d + \gamma_{sc}^{tg}}{c^d + \gamma_{sc}^{tg} + \gamma_{tg}^{tg}} \frac{c_k^d + \alpha_k}{c^d + \sum_{k'} \alpha_{k'}} \tau_{sc,t}^{tg,w_{d,n}} \frac{c_t^{sc,k} + \beta_t^{sc}}{c^{sc,k} + \sum_{w' \in V_{sc}} \beta_{w'}^{sc}}. \quad (4)$$

Consequently, the approximate probability of word $w$ in topic $k$ for language $x$ is

$$\phi_w^{x,k} = \frac{c_w^{x,k} + \beta_w^x}{c^{x,k} + \sum_{w' \in V_x} \beta_{w'}^x}, \quad x \in \{sc, tg\}. \quad (5)$$

And, finally, the approximate probability of topic $k$ in review $d$ can be calculated as

$$\theta_k^d = \frac{c_k^d + \alpha_k}{c^d + \sum_{k'} \alpha_{k'}}. \quad (6)$$

## B. Cross-Lingual JST

The Joint Sentiment/Topic model (JST) is a monolingual model that can simultaneously detect sentiments and topics from texts. In order to accommodate JST to the multilingual scenario, next we propose a CLJST model by integrating JST into the CLLDA framework. LDA has three hierarchical layers, where topics are associated with documents and words are associated with topics. JST extends LDA by adding sentiment layer between document layer and topic layer. Hence, CLJST is a four-layer model, where sentiment labels are associated with documents, topics are associated with sentiment labels, and words are associated with both sentiment labels and topics, respectively. The graphical model of CLJST is represented in Fig. 2. Algorithm 2 presents the generative process of the CLJST model and the specific notation for CLJST is explained in Table III.

In CLJST, the latent variables $\phi$, $\theta$, $\pi$ and $\eta$ are inferred by Gibbs sampling. At each transition step of the Markov chain, the sentiment and aspect of the $n$th word in review $d$ are chosen according to the following conditional probabilities.

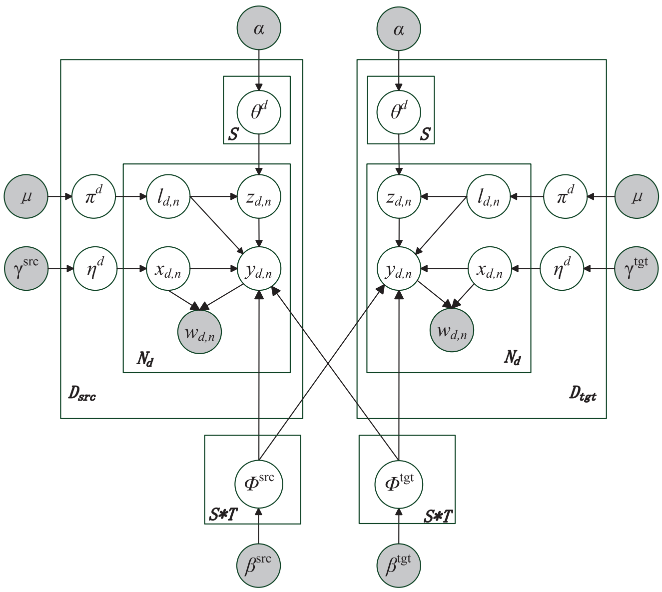



Fig. 2. The graphical representation of CLJST.

TABLE III
THE NOTATION SPECIFIED FOR THE CLJST MODEL

| | |
|---|---|
| $\mathbf{l}/\mathbf{l}_{\neg(d,n)}$ | the sentiment label assignments to all words (excluding the $n$th word in review $d$) |
| $c^d(c^d_l)$ | the number of words (assigned sentiment label $l$) in review $d$ |
| $c^d_{l,k}$ | the number of words assigned sentiment label $l$ and topic $k$ in review $d$ |
| $c^{sc/tg,ps/ng,k}$ | the times that any word is assigned language label $sc/tg$, sentiment label $ps/ng$ and topic $k$ |
| $c^{sc/tg,ps/ng,k}_t$ | the ntimes that any word is assigned language label $sc/tg$, sentiment label $ps/ng$ and topic $k$ when the corresponding intermediate variable is assigned $t$ |

For the source language: If $x_{d,n} = sc$,

$$P(l_{d,n} = l, z_{d,n} = k, y_{d,n} = w_{d,n}, x_{d,n} = sc | \mathbf{w}, \mathbf{l}_{\neg(d,n)}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}) \propto \frac{c^d_{sc} + \gamma^{sc}_{sc}}{c^d + \gamma^{sc}_{tg} + \gamma^{sc}_{sc}} \frac{c^d_l + \mu_l}{c^d + \sum\limits_{l' \in \{ps,ng\}} \mu'_l} \times \frac{c^d_{l,k} + \alpha_k}{c^d_l + \sum\limits_{k'} \alpha_{k'}} \frac{c^{sc,l,k}_{w_{d,n}} + \beta^{sc}_{w_{d,n}}}{c^{sc,l,k} + \sum\limits_{w' \in V_{sc}} \beta^{sc}_{w'}}. \tag{7}$$

And, if $x_{d,n} = tg$ and the translation word of $y_{d,n}$ is $t$,

$$P(l_{d,n} = l, z_{d,n} = k, y_{d,n} = t, x_{d,n} = tg | \mathbf{w}, \mathbf{l}_{\neg(d,n)}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}) \propto \frac{c^d_{tg} + \gamma^{sc}_{tg}}{c^d + \gamma^{sc}_{sc} + \gamma^{sc}_{tg}} \frac{c^d_l + \mu_l}{c^d + \sum\limits_{l' \in \{ps,ng\}} \mu'_l} \times \frac{c^d_{l,k} + \alpha_k}{c^d_l + \sum\limits_{k'} \alpha_{k'}} \tau^{sc,w_{d,n}}_{tg,t} \frac{c^{tg,l,k}_t + \beta^{tg}_t}{c^{tg,l,k} + \sum\limits_{w' \in V_{tg}} \beta^{tg}_{w'}}. \tag{8}$$

---

**Algorithm 2.** The generative process of CLJST.

---

For each pair of topic $z$ and sentiment label $l$:
1. Draw a multinomial distribution of the source language: $\phi^{sc,l,z} \sim Dir(\beta^{sc})$;
2. Draw a multinomial distribution of the target language: $\phi^{tg,l,z} \sim Dir(\beta^{tg})$;

For each document $d$ in the corpus of the source language:
1. Draw a distribution $\pi^d \sim Dir(\mu)$;
2. For each sentiment label $l$, draw a distribution $\theta^{d,l} \sim Dir(\alpha)$;
3. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{sc})$;
4. For each word $w_{d,n}$ in document $d$:
   (1) Draw a sentiment label $l_{d,n} \sim \pi^d$;
   (2) Draw a topic $z_{d,n} \sim \theta^{d,l_{d,n}}$;
   (3) Draw a language label $x_{d,n} \sim \eta^d$;
   a. If $x_{d,n} = sc$, draw an intermediate word in the source language $y_{d,n} \sim \phi^{sc,l_{d,n},z_{d,n}}$; generate a word directly: $w_{d,n} = y_{d,n}$;
   b. If $x_{d,n} = tg$, draw an intermediate word in the target language $y_{d,n} \sim \phi^{tg,l_{d,n},z_{d,n}}$; generate a word by translating from $y_{d,n}$ according to the translation probability, $\tau^{sc,w_{d,n}}_{tg,y_{d,n}}$;

For each document $d$ in the corpus of the target language:
1. Draw a distribution $\pi^d \sim Dir(\mu)$;
2. For each sentiment label $l$, draw a distribution $\theta^{d,l} \sim Dir(\alpha)$;
3. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{tg})$;
4. For each word $w_{d,n}$ in document $d$:
   (1) Draw a sentiment label $l_{d,n} \sim \pi^d$;
   (2) Draw a topic $z_{d,n} \sim \theta^{d,l_{d,n}}$;
   (3) Draw a language label $x_{d,n} \sim \eta^d$;
   a. If $x_{d,n} = sc$, draw an intermediate word in the source language $y_{d,n} \sim \phi^{sc,l_{d,n},z_{d,n}}$; generate a word by translating from $y_{d,n}$ according to the translation probability, $\tau^{tg,w_{d,n}}_{sc,y_{d,n}}$;
   b. If $x_{d,n} = tg$, draw an intermediate word in the target language $y_{d,n} \sim \phi^{tg,l_{d,n},z_{d,n}}$; generate a word directly: $w_{d,n} = y_{d,n}$.

---

For the target language: If $x_{d,n} = tg$,

$$P(l_{d,n} = l, z_{d,n} = k, y_{d,n} = w_{d,n}, x_{d,n} = tg | \mathbf{w}, \mathbf{l}_{\neg(d,n)}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)}) \propto \frac{c^d_{tg} + \gamma^{tg}_{tg}}{c^d + \gamma^{tg}_{tg} + \gamma^{tg}_{sc}} \frac{c^d_l + \mu_l}{c^d + \sum\limits_{l' \in \{ps,ng\}} \mu'_l} \times \frac{c^d_{l,k} + \alpha_k}{c^d_l + \sum\limits_{k'} \alpha_{k'}} \frac{c^{tg,l,k}_{w_{d,n}} + \beta^{tg}_{w_{d,n}}}{c^{tg,l,k} + \sum\limits_{w' \in V_{tg}} \beta^{tg}_{w'}}. \tag{9}$$

And, if $x_{d,n} = sc$ and the translation word of $y_{d,n}$ is $t$,

$$P(l_{d,n} = l, z_{d,n} = k, y_{d,n} = t, x_{d,n} = sc | \mathbf{w}, \mathbf{l}_{\neg(d,n)}, \mathbf{z}_{\neg(d,n)}, \mathbf{x}_{\neg(d,n)}, \mathbf{y}_{\neg(d,n)})$$

$$\propto \frac{c^d_{sc}+\gamma^{tg}_{sc}}{c^d+\gamma^{tg}_{sc}+\gamma^{tg}_{tg}} \frac{c^d_l+\mu_l}{c^d+\sum\limits_{l'\in\{ps,ng\}}\mu'_l} \times \frac{c^d_{l,k}+\alpha_k}{c^d_l+\sum\limits_{k'}\alpha_{k'}} \tau^{tg,w_{d,n}}_{sc,t} \frac{c^{sc,l,k}_t+\beta^{sc}_t}{c^{sc,l,k}+\sum\limits_{w'\in V_{sc}}\beta^{sc}_{w'}}. \qquad (10)$$

Therefore, the approximate probability of sentiment $l$ in review $d$ is

$$\pi^d_l = \frac{c^d_l+\mu_l}{c^d+\sum\limits_{l'\in\{ps,ng\}}\mu'_l}. \qquad (11)$$

And, the approximate probability of topic $k$ for sentiment $l$ in review $d$ is

$$\theta^{d,l}_k = \frac{c^d_{l,k}+\alpha_k}{c^d_l+\sum\limits_{k'}\alpha_{k'}}. \qquad (12)$$

In the end, the approximate probability of word $w$ in sentiment $l$ coupled with topic $k$ for language $x$ can be obtained as

$$\phi^{x,l,z}_w = \frac{c^{x,l,k}_w+\beta^x_w}{c^{x,l,k}+\sum\limits_{w'\in V_x}\beta^x_{w'}}, \qquad (13)$$

where $l \in \{ng, ps\}$ and $x \in \{sc, tg\}$.

## B. Cross-Lingual ASUM

The Aspect and Sentiment Unification Model (ASUM) adopts an approach similar to the JST model, where a sentiment is integrated with a topic. But, unlike JST, ASUM imposes a constraint that all words in a sentence are generated from the same topic. Although this is not always true, the constraint holds up well in practice especially for sentiment reviews. Furthermore, both JST and ASUM make use of a small set of sentiment seed words, and ASUM integrates the seed words into the generative process. In a similar way of CLJST, we embed ASUM into the cross-lingual topic model framework to form a new model, CLASUM. The graphical representation of the CLASUM model is shown in Fig. 3. Formally, the generative process of CLASUM is illustrated in Algorithm 3 and the specific notation for CLASUM is explained in Table IV.

The latent variables $\phi$, $\theta$, $\pi$ and $\eta$ are also inferred by Gibbs sampling. At each transition step of the Markov chain, the sentiment and aspect of the $m$th sentence in review $d$ are chosen according to the following conditional probabilities.

We first draw sentiment label $l$ and aspect $z$ for each sentence as follows:

$$P\big(l_{d,m}=l, z_{d,m}=k|\mathbf{w}, \mathbf{l}_{\neg(d,m)}, \mathbf{z}_{\neg(d,m)}, \mathbf{x}, \mathbf{y}\big) \propto \frac{c^d_l+\mu_l}{c^d+\sum\limits_{l'\in\{ps,ng\}}\mu'_l} \frac{c^d_{l,k}+\alpha_k}{c^d_l+\sum\limits_{k'}\alpha_{k'}} \times \prod_{x\in\{sc,tg\}} \left( \frac{\left(\dfrac{\Gamma\left(\sum_{w=1}^{V_x}\left(c^{x,l,k}_w+\beta^x_w\right)\right)}{\Gamma\left(\sum_{w=1}^{V_x}\left(c^{x,l,k}_w+\beta^x_w\right)+n^{d,m,x}\right)}\right)\times}{\prod_{w=1}^{V_x}\left(\dfrac{\Gamma\left(c^{x,l,k}_w+\beta^x_w+n^{d,m,x}_w\right)}{\Gamma\left(c^{x,l,k}_w+\beta^x_w\right)}\right)} \right). \qquad (14)$$

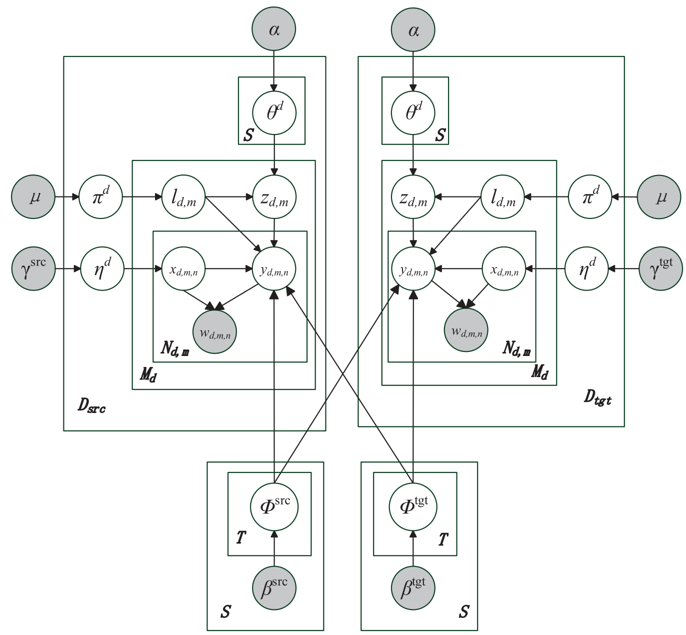



Fig. 3. The graphical representation of the CLASUM model.

TABLE IV
THE NOTATION SPECIFIED FOR THE CLASUM MODEL

| | |
|---|---|
| $\mathbf{z}/\mathbf{z}_{\neg(d,m)}$ | the topic assignments to all sentences (excluding the $m$th sentence in review $d$) |
| $\mathbf{x}/\mathbf{x}_{\neg(d,m,n)}$ | the language label assignments to all words (excluding the $n$th word of the $m$th sentence in review $d$) |
| $\mathbf{y}/\mathbf{y}_{\neg(d,m,n)}$ | the intermediate variable assignments to all words (excluding the $n$th word of the $m$th sentence in review $d$) |
| $\mathbf{l}/\mathbf{l}_{\neg(d,m)}$ | the sentiment label assignments to all sentences (excluding the $m$th sentence in review $d$) |
| $c^d(c^d_l)$ | the number of sentences (assigned sentiment label $l$) in review $d$ |
| $c^d_{l,k}$ | the number of sentences assigned sentiment label $l$ and topic $k$ in review $d$ |
| $n^{d,m,sc/tg}_{(w)}$ | the times that any word (word $w$) of the $m$th sentence assigned the language label $sc/tg$ in review $d$ |

Next, we draw $y_{d,m,n}$ and $x_{d,m,n}$ for the source language documents. Specifically, if $x_{d,m,n} = sc$,

$$P\big(y_{d,m,n}=w_{d,m,n}, x_{d,m,n}=sc|\mathbf{w}, \mathbf{l}, \mathbf{z}, \mathbf{x}_{\neg(d,m,n)}, \mathbf{y}_{\neg(d,m,n)}\big) \propto \frac{c^d_{sc}+\gamma^{sc}_{sc}}{c^d+\gamma^{sc}_{tg}+\gamma^{sc}_{sc}} \frac{c^{sc,l_{d,m},k_{d,m}}_{w_{d,m,n}}+\beta^{sc}_{w_{d,m,n}}}{c^{sc,l_{d,m},k_{d,m}}+\sum\limits_{w'\in V_{sc}}\beta^{sc}_{w'}}. \qquad (15)$$

If $x_{d,m,n} = tg$ and the translation word of $y_{d,m,n}$ is $t$,

$$P\big(y_{d,m,n}=t, x_{d,m,n}=tg|\mathbf{w}, \mathbf{l}, \mathbf{z}, \mathbf{x}_{\neg(d,m,n)}, \mathbf{y}_{\neg(d,m,n)}\big) \propto \frac{c^d_{tg}+\gamma^{sc}_{tg}}{c^d+\gamma^{sc}_{sc}+\gamma^{sc}_{tg}} \tau^{sc,w_{d,m,n}}_{tg,t} \frac{c^{tg,l_{d,m},k_{d,m}}_t+\beta^{tg}_t}{c^{tg,l_{d,m},k_{d,m}}+\sum\limits_{w'\in V_{tg}}\beta^{tg}_{w'}}. \qquad (16)$$

**Algorithm 3.** The generative process of the CLASUM model.

For each pair of topic $z$ and sentiment label $l$:
1. Draw a multinomial distribution of the source language: $\phi^{sc,l,z} \sim Dir(\beta^{sc})$;
2. Draw a multinomial distribution of the target language: $\phi^{tg,l,z} \sim Dir(\beta^{tg})$;

For each document $d$ in the corpus of the source language:
1. Draw a distribution $\pi^d \sim Dir(\mu)$;
2. For each sentiment label $l$, draw a distribution $\theta^{d,l} \sim Dir(\alpha)$;
3. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{sc})$;
4. For each sentence $s_{d,m}$ in document $d$:
   (1) Draw a sentiment label $l_{d,m} \sim \pi^d$;
   (2) Draw a topic $z_{d,m} \sim \theta^{d,l_{d,m}}$;
   (3) For each word $w_{d,m,n}$ in the sentence $s_{d,m}$:
      a. Draw a language label $x_{d,m,n} \sim \eta^d$;
      b. If $x_{d,m,n} = sc$, draw an intermediate word in the source language $y_{d,m,n} \sim \phi^{sc,l_{d,m},z_{d,m}}$; generate a word directly: $w_{d,m,n} = y_{d,m,n}$;
      c. If $x_{d,m,n} = tg$, draw an intermediate word in the target language $y_{d,m,n} \sim \phi^{tg,l_{d,m},z_{d,m}}$; generate a word by translating from $y_{d,m,n}$ according to the translation probability, $\tau_{tg,y_{d,m,n}}^{sc,w_{d,m,n}}$;

For each document $d$ in the corpus of the target language:
1. Draw a distribution $\pi^d \sim Dir(\mu)$;
2. For each sentiment label $l$, draw a distribution $\theta^{d,l} \sim Dir(\alpha)$;
3. Draw a binomial distribution, $\eta^d \sim Beta(\gamma^{tg})$;
4. For each sentence $s_{d,m}$ in document $d$:
   (1) Draw a sentiment label $l_{d,m} \sim \pi^d$;
   (2) Draw a topic $z_{d,m} \sim \theta^{d,l_{d,m}}$;
   (3) For each word $w_{d,m,n}$ in the sentence $s_{d,m}$:
      a. Draw a language label $x_{d,m,n} \sim \eta^d$;
      b. If $x_{d,m,n} = sc$, draw an intermediate word in the source language $y_{d,m,n} \sim \phi^{sc,l_{d,m},z_{d,m}}$; generate a word by translating from $y_{d,m,n}$ according to the translation probability, $\tau_{sc,y_{d,m,n}}^{tg,w_{d,m,n}}$;
      c. If $x_{d,m,n} = tg$, draw an intermediate word in the target language $y_{d,m,n} \sim \phi^{tg,l_{d,m},z_{d,m}}$; generate a word directly: $w_{d,m,n} = y_{d,m,n}$.

For the target language documents we draw $y_{d,m,n}$ and $x_{d,m,n}$ as follows: If $x_{d,m,n} = tg$,

$$P\big(y_{d,m,n} = w_{d,m,n}, x_{d,m,n} = tg | \mathbf{w}, \mathbf{l}, \mathbf{z}, \mathbf{x}_{\neg(d,m,n)}, \mathbf{y}_{\neg(d,m,n)}\big) \propto \frac{c_{tg}^d + \gamma_{tg}^{tg}}{c^d + \gamma_{tg}^{tg} + \gamma_{sc}^{tg}} \frac{c_{w_{d,m,n}}^{tg,l_{d,m},k_{d,m}} + \beta_{w_{d,m,n}}^{tg}}{c^{tg,l_{d,m},k_{d,m}} + \sum_{w' \in V_{tg}} \beta_{w'}^{tg}}. \tag{17}$$

If $x_{d,m,n} = sc$ and the translation word of $y_{d,m,n}$ is $t$,

$$P\big(y_{d,m,n} = t, x_{d,m,n} = sc | \mathbf{w}, \mathbf{l}, \mathbf{z}, \mathbf{x}_{\neg(d,m,n)}, \mathbf{y}_{\neg(d,m,n)}\big) \propto \frac{c_{sc}^d + \gamma_{sc}^{tg}}{c^d + \gamma_{sc}^{tg} + \gamma_{tg}^{tg}} \tau_{sc,t}^{tg,w_{d,m,n}} \frac{c_t^{sc,l_{d,m},k_{d,m}} + \beta_t^{sc}}{c^{sc,l_{d,m},k_{d,m}} + \sum_{w' \in V_{sc}} \beta_{w'}^{sc}}. \tag{18}$$

Then, the approximate probability of sentiment $l$ in review $d$ can be calculated as

$$\pi_l^d = \frac{c_l^d + \mu_l}{c^d + \sum_{l' \in \{ps,ng\}} \mu_{l'}}. \tag{19}$$

This probability is important to sentiment classification. If $\pi_{ps}^d > \pi_{ng}^d$, the review is classified as a positive one, otherwise a negative one. The approximate probability of aspect $k$ for sentiment $l$ in review $d$ is

$$\theta_k^{d,l} = \frac{c_{l,k}^d + \alpha_k}{c_l^d + \sum_{k'} \alpha_{k'}}. \tag{20}$$

Finally, the approximate probability of word $w$ in sentiment $l$ coupled with topic $k$ for language $x$ is

$$\phi_w^{x,l,z} = \frac{c_w^{x,l,k} + \beta_w^x}{c^{x,l,k} + \sum_{w' \in V_x} \beta_{w'}^x}, \tag{21}$$

where $l \in \{ng, ps\}$ and $x \in \{sc, tg\}$.

## IV. Experimental Validation

In this section, we validate the proposed models via experiments and comparisons with existing models. Specifically, we first investigate the performance of the CLLDA model on discovering aspects and validate the effectiveness of its cross-lingual mechanism. Next, we examine the performance of the CLJST and CLASUM models by applying them to practical sentiment classification tasks. Finally, we investigate the impacts of parameters of the models on their performance.

### A. Experimental Settings

*1) Datasets and Lexica:* The real datasets used in the subsequent experiments contain hotel reviews and product reviews collected from well-known websites and in different languages (including English, Chinese, French, German, Spanish, Dutch, and Italian). In all subsequent cross-lingual experiments, English is taken as the source language and the others as target languages. The product reviews cover four domains, namely, electronics, kitchen, network, and health. In more detail, for English, the hotel dataset contains 12000 reviews collected from Booking.com[1], and the product dataset in each domain has 2000 reviews collected from Amazon[2, 3]. For Chinese, the hotel dataset also contains 12000 reviews, obtained from an open source website[4], and the product dataset contains 2000 reviews for each domain, collected from Jingdong[5], one of the most popular e-shopping websites in China. The datasets in other languages are collected from Booking.com, each of which contains 4000 hotel reviews. In the case of supervised sentiment classification, each dataset contains 1000 labeled reviews

[1] http://www.booking.com
[2] http://www.seas.upenn.edu/mdredze/datasets/sentiment/
[3] http://llt.cbs.polyu.edu.hk/~lss/ACL2010_Data_SSLi.zip
[4] http://nlp.csai.tsinghua.edu.cn/
[5] http://www.jd.com

TABLE V
THE LIST OF SEED WORDS FOR DIFFERENT TARGET LANGUAGES

| | Positive | Negatvie |
|---|---|---|
| Chinese | 不错 (not bad), 方便 (convenient), 干净 (clean), 热情 (friendly), 好 (good), 满意 (satisfied), 高兴 (happy), 漂亮 (beautiful), 喜欢 (like), 赞 (great) | 不好 (not good), 惨 (miserable), 差 (bad), 陈旧 (old), 丑 (ugly), 恶劣 (wicked), 慢 (slow), 失望 (disappointed), 脏 (dirty), 糟(terrible) |
| French | bien (good), propreté (clean), calme (quiet), bon (good), propre (clean), aimable (friendly), confort (comfort), bonne (good), agréable (pleasant), excellent (excellent) | cher (expensive), bruyant (nosiy), bruyante (noisy), panne (break-down), mal (evil), mauvaise (bad), sale (dirty), manque (lack), difficile (difficult), désagréable (unpleasant) |
| German | gut (good), gute (good), ruhig (quite), freundlich (friendly), freundliche (friendly), sauber (clean), hilfsbereit (helpful), schön (beautiful), guter (good), saubere (clean) | teuer (expensive), lärm (noise), schlecht (bad), schmutzig (dirty), enttäuscht (disappointed), hässlich (ugly), schlechte (poor), schmutzige (dirty), unzufrieden (unhappy), schlechtes (bad) |
| Spanish | buena (good), limpieza (clean), limpia (clean), buen (good), amable (friendly), excelente (excellent), limpio (clean), bueno (good), perfecto (perfect), contentos (happy) | caro (expensive), pobre (poor), malo (bad), peor (worse), sucio (dirty), sucios (dirty), costoso (expensive), ruidoso (noisy), decepcionado (disappointed), grosero (rude) |
| Italian | ottima (good), pulizia (clean), buona (good), gentile (kind), comodo (comfortable), ottimo (excellent), bene (good), bella (nice), pulito (clean), cortesia (courtesy) | rumore (noise), scarsa (poor), sporco (dirty), male (bad), scarso (poor), cattivo (bad), rumoroso (noisy), scortese (rude), deluso (disappointed), sgradevole (unpleasant) |
| Dutch | goed (good), vriendelijk (friendly), prima (fine), goede (well), schoon (clean), mooi (beautiful), uitstekend (excellent), behulpzaam (helpful), netjes (neatly), lekker (nice), | lawaai (noise), jammer (pity), slecht (bad), gehorig (noisy), duur (expensive), vies (dirty), pover (poor), slechter (worse), teleurstellend (disappointing), lelijke (ugly) |

for training and 1000 labeled reviews for testing, among which the percentages of positive and negative reviews are half and half. In the case of unsupervised sentiment classification, each dataset contains 1000 labeled reviews for testing and the rest unlabeled data is used for topic and sentiment modelling. All of the datasets are preprocessed with sentence segmentation. For Chinese, we further use ICTCLAS[6] to conduct word segmentation over sentences.

As aforesaid, in the proposed CLJST and CLASUM models, a bilingual dictionary is used to bridge two languages and an opinion word lexicon is used to identify which words are sentiment words. In a bilingual dictionary, one source word usually corresponds to multiple target words (i.e., translation words in the target language). But, only one target word is adopted each time. The candidate target word and the topic are selected simultaneously in Gibbs sampling, which is a random process. In order to overcome the interference of irrelevant target words, we employ the following simple selection strategy. Given a source word, we first compute the Term Frequency (TF) for each corresponding target word in the target corpora; Next, if the number of target words exceeds a threshold $K$, we choose top $K$ terms with the highest TF values as candidate target words, otherwise all target words are taken into account; Finally, a translation probability is assigned to each candidate target word $w_i$ according to $TF(w_i)/\sum_{i=1}^{K} TF(w_i)$. In this study, the translation probabilities are incorporated as prior knowledge, which makes the target word with a high translation probability has more chances to be selected; but other target words may also be selected. Generally, given a source word, the number of target words that can be found in the target corpus is less than that in a bilingual dictionary, since some candidate translation words are infrequently used. According to our statistics, there are about 87% and 95% source words in hotel reviews and product reviews, respectively, which have the number of corresponding target words less than or equals to 5. Therefore, we set the threshold $K$ for candidate target words as 5. And, for each language pair, a bilingual dictionary is used. Specifically, for English-Chinese, an existing dictionary with 41,814 entries is employed. Other bilingual dictionaries with English as the source language and other five additional languages as the target language are acquired by *Google Translate* from the English vocabulary.

In the experiments, the opinion word lexica used for the source language, English, are obtained from the widely used knowledge base, HowNet[7]. For each target language, we use seed words as prior knowledge. The seed words are obtained by term frequency. Specifically, we select 10 positive and 10 negative sentiment words with the highest term frequency as seed words. Table V presents the full list of seed words of all target languages, where for each seed word a English translation word is provided in subsequent brackets, which is obtained by *Google Translate*.

*2) Parameters Settings:* We set $\alpha_k = 50/T$ for each aspect $k$ as in [?]. We incorporate the knowledge of seed sentiment words by setting hyper parameter $\beta$, i.e., priors for word distributions of sentiment-topic pairs. This means that we initially predict that no negative seed word appears in a positive sentiment-topic pair, and vice versa. Specifically, for each positive seed word $w$ with language label $x$, $\beta_w^{x,ng}$ is set to 0; for each negative seed word $w$ with language label $x$, $\beta_w^{x,ps}$ is set to 0; for each non-seed word with language label $x$ and sentiment label $l$, the default value of $\beta_w^{x,l}$ is set to 0.1. Besides, at the initialization step, we assign the seed words with their seed sentiments, whilst the sentiments of the non-seed words are initialized randomly. The value for $\mu_l$ is set to 0.01 for each sentiment label $l$. For the target language, the values of $\gamma_{sc}^{tg}$ and $\gamma_{tg}^{tg}$ are asymmetrical. In the experiments, we fix $\gamma_{tg}^{tg}$ but vary $\gamma_{sc}^{tg}$ to adjust the amount of knowledge learned from the source language to improve topic extraction or sentiment classification

[6]http://ictclas.cn/index.html

[7]http://www.keenage.com/html/e_index.html

TABLE VI
THE EXAMPLE ASPECTS DISCOVERED BY THE CLLDA MODEL

| | | |
|---|---|---|
| Food | topic | breakfast, coffee, room, food, restaurant, buffet, fruit, eggs, service, morning, pastries, wine, cheese, dinner, bar |
| | Pos. | good, great, nice, excellent, fresh, lovely, best, wonderful, delicious |
| | Neg. | bad, poor, cold, hot, hard, terrible |
| 食物(food) | topic | 品种 (variety), 服务员 (waiter), 送(deliver), 份 (dish), 菜 (vegetable), 鸡蛋 (egg), 早餐 (breakfast), 自助 (buffet), 味道 (taste), 每天 (every), 粥 (porridge), 东西 (stuff), 早饭 (breakfast), 西式 (western), 人气 (popularity), 咸菜 (pickles) |
| | Pos. | 不错 (good), 丰盛 (rich), 饱 (full), 丰富 (abundant), 很好 (nice), 美味 (yummy), 好吃 (yummy), 新鲜 (fresh) |
| | Neg. | 难吃 (horrible), 一般 (just so-so), 没什么 (nothing),差 (bad), 单调 (tedious), 失望 (disappointing) |
| Service | topic | staff, location, English, service, desk, breakfast, room, reception, stay, spoke, restaurant, restaurants, concierge, people, help |
| | Pos. | helpful, friendly, great, good, clean, excellent, nice, pleasant, wonderful, polite, courteous, comfortable |
| | Neg. | trouble, rude, unfriendly |
| 服务(service) | topic | 服务 (service), 态度 (attitude), 前台 (receptionist), 酒店 (hotel), 服务员 (waiter), 候 (wait), 客人 (customer), 大堂 (hall manager), 感觉 (feeling), 员工 (staff), 服务生 (waiter), 小姐 (waitress), 微笑 (smile), 素质 (quality), 行李 (luggage), 打招呼 (greet), 效率(efficiency) |
| | Pos. | 不错 (good), 很好 (nice), 热情 (friendly), 高 (high), 礼貌 (polite),到位 (proper), 赞 (excellent), 笑容 (smile), 结账 (check-out) |
| | Neg. | 差 (bad), 慢 (slow), 恶劣 (terribile), 久(long) |

in the target language. For the source language, the values of both $\gamma_{sc}^{sc}$ and $\gamma_{tg}^{sc}$ are fixed to 0.01.

*3) Baselines:* For the purpose of performance comparison, we adopt four representative baselines. We first choose USL (universal sentiment lexicon) and SVM, because USL is a fundamental approach to unsupervised sentiment classification and SVM is a typical model for supervised sentiment classification. Furthermore, we choose JST and ASUM as baselines, because they are two state-of-the-art models jointly considering aspects and sentiments. But, they both are monolingual models. CLJST is an extended model for JST, whilst CLASUM is an extended model for ASUM. Thus, we compare CLJST with JST and CLASUM with ASUM, respectively. Through comparing the cross-lingual model with the corresponding monolingual model, we can validate the effectiveness of the proposed cross-lingual mechanism.

## B. Experimental Results

*1) Aspect Discovery:* This experiment is to investigate the performance of the developed models on discovering aspects on given datasets. For the purpose of illustration, Table VI presents two major aspects discovered by CLLDA on English and Chinese hotel reviews, where, for each aspect, top 30 words are listed, which are further divided into topic words, positive opinion words, and negative opinion words. Note that for the sake of space limitation, the aspects discovered on datasets in other languages are not presented. It can be seen in Table VI that the proposed model can effectively detect major aspects from the reviews in both source and target languages. The extracted words under each aspect are quite coherent and meaningful. Moreover, the proposed model can obtain bilingual aligned aspects due to the mechanism of sharing the same topic distribution. This mechanism enables us to improve the aspect modeling on the data in the target language by leveraging the rich resources in the source language. Besides, Table VI clearly shows that our model can effectively extract both aspects and aspect-dependent sentiment knowledge. The discovered opinion words (either positive or negative) are quite specific and informative with respect to the aspects.

*2) Cross-Lingual Mechanism:* This experiment aims to evaluate the proposed cross-lingual mechanism. For this purpose, we calculated the perplexity of the test dataset under different $\gamma_{sc}^{tg}$. In information theory, perplexity is a measurement of how well a probability distribution or probability model predicts a sample. In [1], it is employed to examine the performance of topic detection. Mathematically, for a dataset with $M$ documents, its perplexity is defined as

$$perplexity(D) = \exp\left\{-\frac{\sum_{d=1}^{M}\log p(\mathbf{w_d})}{\sum_{d=1}^{M} N_d}\right\}, \tag{22}$$

where $N_d$ indicates the number of words in document $d$, and $p(\mathbf{w_d})$ indicates the probability of document $d$ [1],

$$p(\mathbf{w_d}) = \sum_{z} p(z)\prod_{n=1}^{N_d} p(w_n|z). \tag{23}$$

In this context, the lower the perplexity, the better the generalization performance of the topic model. For more details, readers are refereed to [1].

We first evaluate the proposed cross-lingual framework on different domains. Table VII presents the perplexity values of the test datasets in Chinese under different domains. As seen in Table VII, with the increase of $\gamma_{sc}^{tg}$, the perplexity under different domains basically decreases in a monotonic manner. Therefore, we can say that the performance of CLLDA was improved by introducing the cross-lingual mechanism. Actually, the improvement mainly benefits from the knowledge transferred from the data in the source language. More specifically, within a certain range, the larger the $\gamma_{sc}^{tg}$, the more improvement that can be achieved.

We further evaluate the proposed cross-lingual framework on different languages. Taking English as the source language,

TABLE VII
THE PERPLEXITY OF CHINESE TEST DATASETS IN DIFFERENT DOMAINS WITH EXPONENTIALLY INCREASING $\gamma_{sc}^{tg}$

| $\gamma_{sc}^{tg} \rightarrow$ | 0.001 | 0.01 | 0.1 | 1 |
|---|---|---|---|---|
| Electronics | 891.577 | 886.291 | 875.624 | 829.403 |
| Kitchen | 806.246 | 804.469 | 788.822 | 751.710 |
| Network | 585.608 | 582.498 | 577.552 | 562.346 |
| Health | 696.527 | 690.537 | 680.129 | 654.075 |
| Hotel | 133.922 | 129.507 | 130.738 | 125.999 |
| Average | 534.835 | 529.835 | 524.656 | 508.012 |

TABLE VIII
THE PERPLEXITY OF TEST DATASETS IN DIFFERENT LANGUAGES IN THE HOTEL DOMAIN WITH EXPONENTIALLY INCREASING $\gamma_{sc}^{tg}$

| $\gamma_{sc}^{tg} \rightarrow$ | 0.001 | 0.01 | 0.1 | 1 |
|---|---|---|---|---|
| French | 664.795 | 661.59 | 659.498 | 658.047 |
| German | 1090.84 | 1067.38 | 1062.16 | 1044.94 |
| Spanish | 657.153 | 651.546 | 649.757 | 629.508 |
| Dutch | 868.732 | 866.921 | 855.115 | 831.888 |
| Italian | 1038.45 | 1031.68 | 1035.33 | 1024.94 |
| Average | 863.994 | 855.823 | 852.372 | 837.865 |

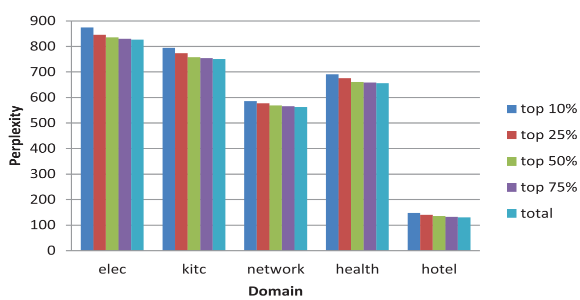



Fig. 4. The perplexity of Chinese test sets in different domains with varying-sized bilingual dictionaries.

Table VIII presents the perplexity values of test datasets in different languages under the hotel domain. It can be observed from Table VIII that, with the increase of $\gamma_{sc}^{tg}$, the perplexity monotonically decreases under different languages. Therefore, we can claim that the proposed cross-lingual mechanism is effective for different domains and for different languages.

As aforesaid, in CLLDA a bilingual dictionary is requested to bridge the source and target languages. In order to investigate the dependency of the performance of the CLLDA model on the quality of the bilingual dictionary, we compared the perplexity of the same test datasets of different domains in Chinese with a group of bilingual dictionaries with different sizes and thus different quality. For this purpose, we first ranked the original bilingual dictionary containing 41,814 entries according to the frequency of source words, and then selected top $N$ (i.e., $N =$ 10%, 25%, 50%, 75%, 100%) source words as new dictionaries, respectively. In this way, the size of the new dictionaries can well indicate their quality. The experimental results are presented in Fig. 4. From this figure, we can observe that with the increase of the size of the dictionary, the performance of the perplexity is decreased gradually, indicating that the performance of the CLLDA model is improved. When the number of source words contained in the new dictionary exceeds 50%, the performance of the CLLDA model is only very slightly affected by the bilingual dictionary.

*3) Sentiment Classification:* In order to evaluate the performance of the CLJST and CLASUM models, we apply the sentiment-aspect pairs extracted by these two models in real sentiment classification tasks. We quantitatively study the quality of the sentiment-aspect discovery according to how much they can improve the performance of sentiment classification, as compared to the four baselines presented in Section IV-A3. The sentiment of a document $d$ is determined according to $P(l|d)$, i.e., the probability of sentiment label $l$ given document $d$, which is approximated using Equation (11)/(19) in CLJST/CLASUM. Specifically, a document $d$ is classified as a positive one, if its probability of positive sentiment label $l_{ps}$ (i.e., $P(l_{ps}|d)$) is greater than its probability of negative sentiment label $l_{ng}$ (i.e., $P(l_{ng}|d)$), or vice versa. Following the experimental setup of JST [2] and ASUM [3], we employ accuracy as the criterion, which is the ratio of the number of correctly classified reviews to the total number of reviews. Without loss of generality, we conducted 2-fold cross-validation tests in this paper.

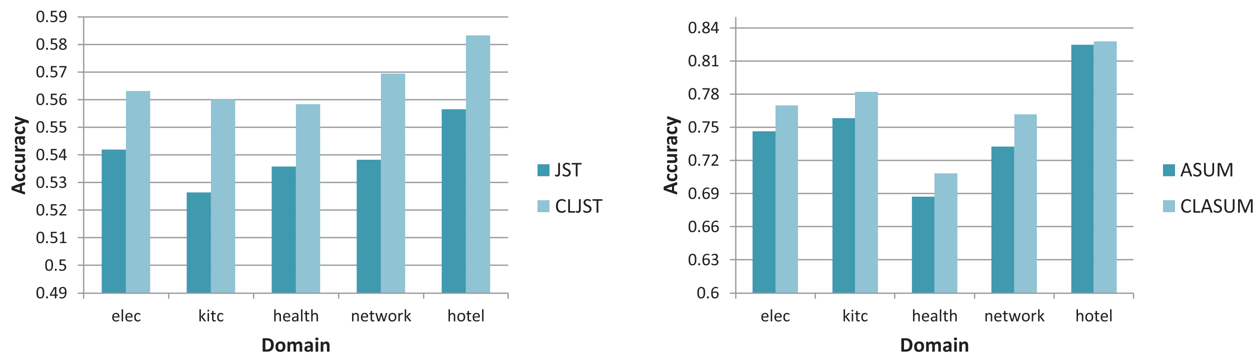
We first compare CLJST and CLASUM with the unsupervised USL and supervised SVM models. All the models are evaluated on the same testing datasets. In USL, we adopt a universal sentiment lexicon collected from HowNet to determine the sentiment polarity of a review. In SVM, we employ LibSVM[8] to train a sentiment classifier, then predict the sentiment polarities on test data. Besides, we tried different kernel functions to refine the performance of SVM, and finally adopt a linear one as the kernel function, with which LibSVM achieves the best performance. Table IX presents the accuracy of the CLJST and CLASUM models and the two baselines, USL and SVM, in sentiment classification of different domains. In this table, for each model and each domain, the accuracy of sentiment classification on both positive and negative reviews as well as their combination are listed. The bottom row presents the accuracy of sentiment classification averaged on different domains. From Table IX, we can observe: First, CLASUM has comparable performance to the supervised method, SVM, which well demonstrates the value of CLASUM, as an unsupervised model; Second, CLASUM performs better than USL, indicating that unsupervised sentiment classification can perform well without a universal sentiment lexicon; Third, although CLJST and CLASUM are both topic based models, CLASUM performs much better than CLJST. This is mainly because of the constraint in CLASUM, where all words in a sentence are generated from the same topic; However, in CLJST different words in a sentence may be generated from different topics. As compared to CLJST, the constraint in CLASUM introduces a fault-tolerant mechanism for sentiment-aspect discovery. This may be not appropriate for all circumstances, but it really holds up well for sentiment classification.

Since the purpose of this study is to leverage the resources in the source language to improve sentiment classification in the target language, in the next experiment we compare

[8]http://www.csie.ntu.edu.tw/cjlin/libsvm/

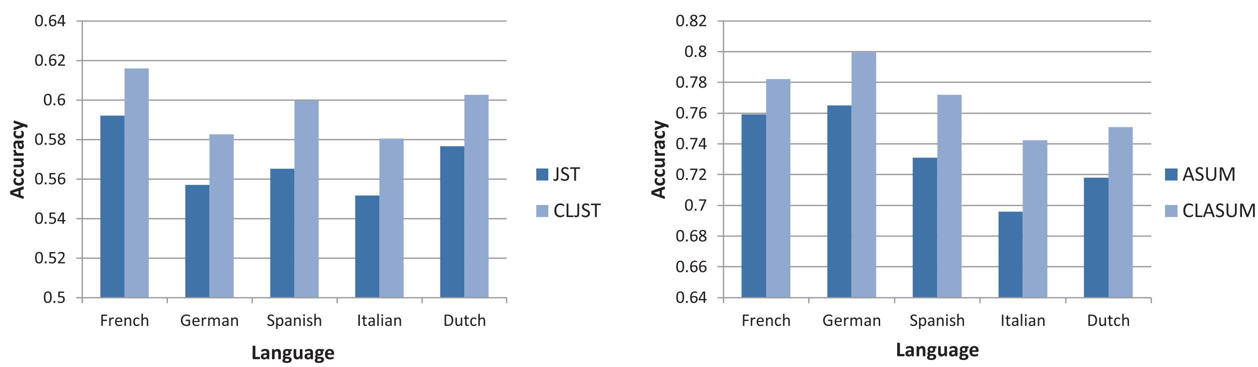
TABLE IX
THE ACCURACY OF SENTIMENT CLASSIFICATION WITH DIFFERENT MODELS

| | USL | | | SVM | | | CLJST | | | CLASUM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ps | ng | all | ps | ng | all | ps | ng | all | ps | ng | all |
| Electronics | 0.7182 | 0.7255 | 0.7208 | 0.7920 | 0.7916 | 0.7918 | 0.5671 | 0.5812 | 0.5752 | 0.7889 | 0.7565 | 0.7710 |
| Kitchen | 0.7636 | 0.6937 | 0.7255 | 0.7791 | 0.7851 | 0.7820 | 0.5685 | 0.5891 | 0.5726 | 0.7981 | 0.7692 | 0.7853 |
| Network | 0.7157 | 0.6976 | 0.7065 | 0.7595 | 0.7883 | 0.7738 | 0.5787 | 0.5693 | 0.5731 | 0.7826 | 0.7569 | 0.7647 |
| Health | 0.6793 | 0.6824 | 0.6808 | 0.7068 | 0.6939 | 0.7003 | 0.5675 | 0.5650 | 0.5662 | 0.7333 | 0.7085 | 0.7104 |
| Hotel | 0.7176 | 0.7785 | 0.7433 | 0.8327 | 0.8308 | 0.8315 | 0.5936 | 0.5768 | 0.5864 | 0.8177 | 0.8305 | 0.8286 |
| Average | 0.7189 | 0.7155 | 0.7154 | 0.7740 | 0.7779 | 0.7759 | 0.5751 | 0.5763 | 0.5747 | 0.7841 | 0.7643 | 0.7720 |




(a) (b)

Fig. 5. The accuracy of sentiment classification of the models on datasets of different domains. (a) JST vs. CLJST (b) ASUM vs. CLASUM.




(a) (b)

Fig. 6. The accuracy of sentiment classification of the models on datasets in different languages. (a) JST vs. CLJST (b) ASUM vs. CLASUM.

CLJST and CLASUM with their monolingual versions, JST and ASUM, which are directly applied to test datasets of different domains in Chinese for sentiment classification. Fig. 5 presents the experimental results of the four models on data in Chinese, but of different domains. As seen in Fig. 5, in all of the five domains, CLJST outperforms JST and CLASUM outperforms ASUM, which convincingly demonstrates the validity of the proposed cross-lingual mechanism in CLJST and CLASUM and thus highlight the value of these models in practical applications. Note that the improvement made by CLJST and CLASUM mainly lies in the joint modelling of two datasets in different languages. In this way, discovering aspects and sentiments on the target language can be refined via transfer learning from auxiliary datasets in the source language. More specifically, the larger the datasets, the more word co-occurrence information that can be used. Therefore, modelling on two datasets can achieve better results than modelling on one of them.

We also compare CLJST and CLASUM with JST and ASUM on hotel review datasets in different languages. Fig. 6 presents the experimental results, i.e., the accuracy of sentiment classification of the four models. It can be seen in Fig. 6 that for all datasets in different languages, CLJST and CLASUM perform better than JST and ASUM, respectively, which is consistent with the above results on datasets of different domains. Besides, in these two experiments, it can be observed that ASUM performs better than JST, and CLASUM performs better than CLJST. The reason lies in that in JST/CLJST, each word is regarded as a separate unit, whilst in ASUM/CLASUM, each sentence is regarded as a separate unit. In reviews, each sentence tends to represent one aspect and one sentiment. Therefore, viewing each sentence as a separate unit is more reasonable and helpful for sentiment classification than viewing each word as a separate unit.

Based on these two experiments, it can be concluded that no matter in what domains and in what languages, the proposed

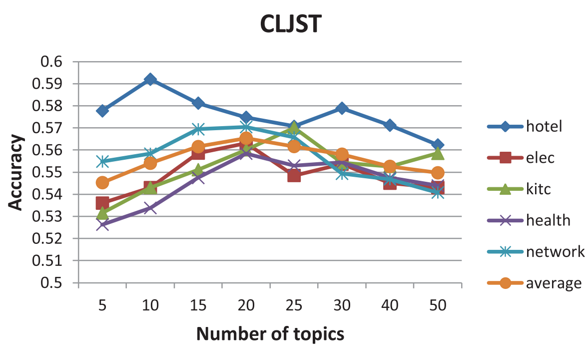
TABLE X
THE ACCURACY OF SENTIMENT CLASSIFICATION BY CLJST AND CLASUM WITH EXPONENTIALLY INCREASING $\gamma_{sc}^{tg}$.
(A) CLJST (B) CLASUM

| $\gamma_{sc}^{tg} \rightarrow$ | 0.01 | 0.1 | 1 | 10 |
|---|---|---|---|---|
| Electronics | 0.5360 | 0.5474 | 0.5584 | 0.5511 |
| Kitchen | 0.5430 | 0.5529 | 0.5661 | 0.5586 |
| Network | 0.5315 | 0.5593 | 0.5694 | 0.5439 |
| Health | 0.5408 | 0.5474 | 0.5518 | 0.5405 |
| Hotel | 0.5715 | 0.5733 | 0.5755 | 0.5584 |
| Average | 0.5446 | 0.5561 | 0.5636 | 0.5505 |

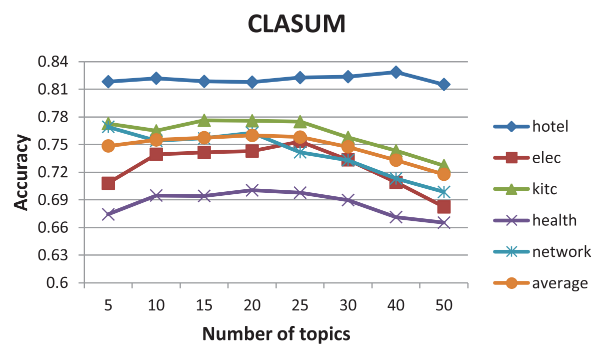
(a)

| $\gamma_{sc}^{tg} \rightarrow$ | 0.01 | 0.1 | 1 | 10 |
|---|---|---|---|---|
| Electronics | 0.7243 | 0.7547 | 0.7546 | 0.7264 |
| Kitchen | 0.7626 | 0.7713 | 0.7838 | 0.7713 |
| Network | 0.7443 | 0.7575 | 0.7552 | 0.7225 |
| Health | 0.6892 | 0.6978 | 0.7100 | 0.7007 |
| Hotel | 0.8105 | 0.8265 | 0.8175 | 0.8056 |
| Average | 0.7462 | 0.7616 | 0.7642 | 0.7453 |

(b)




(a)




(b)

Fig. 7. The accuracy of sentiment classification by CLJST and CLASUM with different numbers of topics. (a) CLJST (b) CLASUM.

cross-lingual models, CLJST and CLASUM, are effective and robust.

*4) Impact of Model Parameters:* As two important parameters of the proposed cross-lingual CLJST and CLASUM models, in what follows we investigate the impact of the hyperparameters $\gamma_{sc/tg}^{tg}$ and the number of topics on their performance.

Table X presents the results of sentiment classification of CLJST and CLASUM with $\gamma_{sc}^{tg}$ exponentially increasing from 0.01 to 10. In this experiment, $\gamma_{tg}^{tg}$ is fixed to be 0.01 such that $\gamma_{sc}^{tg}$ is greater than or equals to $\gamma_{tg}^{tg}$, which guarantees that more knowledge is learned from the source language than from the target one. From Table X, we can note that for both CLJST and CLASUM, the accuracy of sentiment classification increases as $\gamma_{sc}^{tg}$ increases from 0.01 to 0.1 and 1. However, the accuracy decreases, as $\gamma_{sc}^{tg}$ increases to a quite large value, 10. This phenomenon suggests that with a relatively large $\gamma_{sc}^{tg}$, the model can learn more knowledge from the source language to help sentiment classification for the target language. However, a too large $\gamma_{sc}^{tg}$ may lead the model to excessively depending on the source language and overlooking the features of the target language. As a consequently, the performance of the models decreases.

Fig. 7 presents the results of sentiment classification of both CLJST and CLASUM in different domains with different numbers of topics. The results averaged on different domains are also presented in Fig. 7. From this figure, we can see that, by and large, the accuracy of sentiment classification first increases, as the number of topics increases. When the number of topics falls into the region (10,30), the CLJST and CLASUM models achieve the best performance. Beyond this region, the performance of the models decreases. For this reason, in all of the experiments presented in the above subsections the number of topics is set to 20. Besides, the main reason why the model on hotel domain performs better than on the other domains lies in that the hotel domain contains the largest number of reviews compared with the rest.

## V. CONCLUSIONS

Cross-lingual sentiment classification, aiming to leverage the information/knowledge learned from a source language to benefit sentiment classification in a target language, is an important research issue with many practical applications. In this paper, we investigate for the first time an unsupervised cross-lingual topic model framework for sentiment classification at aspect level. Specifically, we propose a cross-lingual LDA-based topic model framework (CLLDA). By embedding the state-of-the-art models JST and ASUM into CLLDA, we further present two unsupervised cross-lingual joint aspect/sentiment models, namely, CLJST and CLASUM, for sentiment classification at aspect level, which make use of the information/knowledge obtained from a source language to help sentiment classification in a target language. Through experiments on realistic review datasets in different domains and different languages, we examine the performance of the CLLDA framework on discovering aspects and validate the effectiveness of its cross-lingual mechanism. We further study the performance and merits of the CLJST and CLASUM models by applying them to sentiment classification tasks and comparing to typical baselines, which make them useful and practical tools for carrying out cross-lingual sentiment classification. Finally, we investigate the impact of the hyperparameters of the CLJST and CLASUM models and the number of topics on their performance.

## REFERENCES


[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.

[2] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 375–384.

[3] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," *Proc. WSDM '11*, 2011, pp. 815–824.

[4] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual subjectivity analysis using machine translation," *Proc. Conf. Empir. Meth. Nat. Lang. Process.*, 2008, pp. 127–135.

[5] L. Shi, R. Mihalcea, and M. Tian, "Cross language text classification by model translation and semi-supervised learning," *Proc. Conf. Empir. Meth. Nat. Lang. Process.*, 2010, pp. 1057–1067.

[6] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Knowledge-based approaches to concept-level sentiment analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 12–14, Mar.-Apr. 2013.

[7] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay, "Enhanced senticnet with affective labels for concept-based opinion mining," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 31–38, Mar.-Apr. 2013.

[8] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," *Proc. 52nd Annu. Meeting Assoc. Comput. Linguist.*, 2014, vol. 1, pp. 1555–1565.

[9] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of twitter," *Proc. Inf. Process. Manage.*, 2015.

[10] X. Wan, "Co-training for cross-lingual sentiment classification," *Proc. Joint Conf. 47th Annu. Meeting ACL and 4th Int. Joint Conf. Nat. Lang. Process. AFNLP: 1-Vol. 1*, 2009, pp. 235–243.

[11] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 45–63, 2014.

[12] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, pp. 723–762, 2014.

[13] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2015, pp. 959–962, ACM.

[14] S. Li, R. Wang, H. Liu, and C.-R. Huang, "Active learning for cross-lingual sentiment classification," *Proc. Nat. Lang. Process. and Chinese Comput.*, 2013, pp. 236–246, Springer.

[15] Q. Chen, W. Li, Y. Lei, X. Liu, and Y. He, "Learning to adapt credible knowledge in cross-lingual sentiment analysis," *Proc. 53rd Annu. Meeting Assoc. Comput. Linguist.*, 2015.

[16] S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," *Proc. Annu. Conf. North Amer. Chap. Assoc. Comput. Linguist.*, 2010, pp. 804–812.

[17] Y. Lu, M. Castellanos, and U. Dayal, *et al.* "Automatic construction of a context-aware sentiment lexicon: An optimization approach," *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 347–356.

[18] S. Kim, J. Zhang, Z. Chen, A. H. Oh, and S. Liu, "A hierarchical aspect-sentiment model for online reviews," *Proc. Assoc. Adv. Artif. Intell.*, 2013.

[19] S. Mukherjee, G. Basu, and S. Joshi, "Joint author sentiment topic model," *Proc. 14th Int. Conf. Data Mining (SDM)*, 2014, pp. 370–378.

[20] S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. Bu, C. Chen, and X. He, "Interpreting the public sentiment variations on twitter," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1158–1170, 2014.

[21] E. Erosheva, S. Fienberg, and J. Lafferty, "Mixed-membership models of scientific publications," *Proc. Nat. Acad. Sci. USA*, 2004.

[22] D. Mimno, H. M. Wallach, J. Naradowsky, D. A. Smith, and A. McCallum, "Polylingual topic models," *Proc. Conf. Empir. Meth. Nat. Lang. Process.: 2-Vol. 2*, 2009, pp. 880–889, Association for Computational Linguistics.

[23] J. Boyd-Graber and D. M. Blei, "Multilingual topic models for unaligned text," *Proc. 25th Conf. Uncertainty Artif. Intell.*, 2009, pp. 75–82.

[24] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, "Opinionit: A text mining system for cross-lingual opinion analysis," *Proc. 19th ACM Int. Conf. Inf. knowl. Manage.*, 2010, pp. 1199–1208, ACM.

[25] J. Boyd-Graber and P. Resnik, "Holistic sentiment analysis across languages: Multilingual supervised latent Dirichlet allocation," *Proc. Conf. Empir. Meth. Nat. Lang. Process.*, 2010, pp. 45–55.

[26] Z. Lin, X. Jin, X. Xu, W. Wang, X. Cheng, and Y. Wang, "A cross-lingual joint aspect/sentiment model for sentiment analysis," *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 1089–1098, ACM.


**Zheng Lin** obtained the Ph.D. degree in the Institute of Computing Technology, the Chinese Academy of Sciences in 2014. She is an Assistant Professor in the Institute of Information Engineering, the Chinese Academy of Sciences. Her current research interests include Sentiment Analysis and Natural Language Processing.

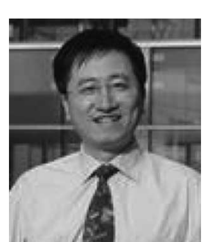
**Xiaolong Jin** obtained the Ph.D. degree in computer science from Hong Kong Baptist University in 2005. He is an Associate Professor in the Institute of Computing Technology, the Chinese Academy of Sciences. His research interests include social computing, social network, performance modelling and evaluation, multi-agent systems. Dr. Jin has published two monographs and more than 120 scientific papers.

**Xueke Xu** is an Assistant Professor in the Institute of Computing Technology, the Chinese Academy of Sciences. His research interests include web opinion retrieval and mining.

**Yuanzhuo Wang** is an Associate Professor in the Institute of Computing Technology, the Chinese Academy of Sciences (ICT-CAS). His major research interests include Network Security and Social Network.

**Xueqi Cheng** is a Professor in the Institute of Computing Technology, the Chinese Academy of Sciences, and the director of the Key Lab of Network Data Science and Technology. His main research interests include network science, web search and data mining, big data processing and distributed computing architecture, etc.

**Weiping Wang** is a Professor in the Institute of Information Engineering, the Chinese Academy of Sciences. His main research interests include big data storage & processing, distributed computing architecture and data mining, etc.



**Dan Meng** is a Professor and the director of the Institute of Information Engineering, the Chinese Academy of Sciences. His main research interests include computer architecture, information security and big data storage & processing, etc.