Complementary Aspect-based Opinion Mining Across Asymmetric Collections

Yuan Zuo[1], Junjie Wu[2*], Hui Zhang[1], Deqing Wang[1], Hao Lin[2], Fei Wang[1], Ke Xu[1]

[1]School of Computer Science and Engineering, Beihang University, China
[2]School of Economics and Management, Beihang University, China
[*]*corresponding author: wujj@buaa.edu.cn*

***Abstract*—Aspect-based opinion mining is to find elaborate opinions towards an underlying theme, perspective or viewpoint as to a subject such as a product or an event. Nowadays, with rapid growing of opinionated text on the Web, mining aspect-level opinions has become a promising means for online public opinion analysis. In particular, the booming of various types of online media provide diverse yet complementary information, bringing unprecedented opportunities for public opinion analysis across different populations. Along this line, in this paper, we propose CAMEL, a novel topic model for complementary aspect-based opinion mining across asymmetric collections. CAMEL gains complementarity by modeling both common and specific aspects across different collections, and keeping all the corresponding opinions for contrastive study. To further boost CAMEL, we propose AME, an automatic labeling scheme for maximum entropy model, to help discriminate aspect and opinion words without heavy human labeling. Extensive experiments on synthetic multi-collection data sets demonstrate the superiority of CAMEL to baseline methods, in leveraging cross-collection complementarity to find higher-quality aspects and more coherent opinions as well as aspect-opinion relationships. This is particularly true when the collections get seriously imbalanced. Experimental results also show that the AME model indeed outperforms manual labeling in suggesting true opinion words. Finally, case study on two public events further demonstrates the practical value of CAMEL for real-world public opinion analysis.**

***Index Terms*—Aspect-based Opinion Mining; Topic Detection and Tracking; LDA Model; Maximum Entropy Model**

I. Introduction

With the dramatic growth of opinionated user generated content on the Web, to automatically understand, extract and summarize the public opinions expressed in different online media platforms has therefore become an important research topic and gained much attention in recent years [13], [24]. Aspect-based opinion mining, a technique proposed originally for finding elaborate opinions towards a perspective of a product [19], has become a promising means for mining aspect-level opinions for online public opinion analysis, where the concept of an aspect here has been extended to be an underlying theme, perspective or viewpoint as to a public event. For instance, for the annual key event Two Sessions (of the NPC and the CPPCC) 2015 in China, we would like to know the elaborate public opinions towards a plenty of relatively focused themes that have generated heated discussions, e.g., the downward pressure on GDP, the opportunities in Jing-Jin-Ji integration, the Hukou reform, anti-corruption, environment protection, etc. Aspect-based opinion mining technique becomes an intuitive candidate to fulfill this task.

Moreover, the diverse yet complementary information provided by rich online media of various types brings great opportunities for public opinion analysis across different collections. Indeed in the literature, there have been quite some excellent studies on cross-collection topic modeling [21], [1], [4], [5]. However, they either pay little attention to the complementarity of aspects across collections [4], or just focus on topics and aspects without considering the opinions [5]. Therefore, further study is still in great need for building cross-collection aspect-based opinion mining model, based on which diversity and complementarity in both aspects and opinions could be leveraged across collections containing substantially asymmetric information, e.g., the news collection with clear aspects versus the tweets collection with strong opinions.

To address the above challenge, in this paper, we propose CAMEL (Cross-collection Auto-labeled MaxEnt-LDA), a novel topic model for complementary aspect-based opinion mining across asymmetric collections. To our best knowledge, our work is among the earliest studies in this direction. CAMEL is essentially a type of cross-collection LDA model, which models aspect-level opinions and gains complementarity by modeling both common and specific aspects across different collections. By keeping all the corresponding opinions for both common and specific aspects, CAMEL is also capable of contrastive opinion analysis. Moreover, as a booster to CAMEL, we propose AME, an automatic labeling scheme for maximum entropy model. It helps discriminate aspect and opinion words without heavy human labeling.

We conducted extensive experiments on synthetic multi-collection data sets to evaluate the quality of aspects as well as opinions induced by CAMEL. Specifically, we design a sentence classification experiment to justify that CAMEL can find higher-quality aspects than baseline methods, and shows more robust performances especially with imbalanced collections in varying degrees. Besides, CAMEL exhibits obvious superiority in learning more coherent opinions and more relevant aspects and opinions in terms of the coherence measure. Also, the AME model for CAMEL indeed outperforms manual labeling in distinguishing aspect words from opinion ones. Finally, case study on two public events further demonstrates the practical value of CAMEL for real-world public opinion analysis.

II. Problem Definition

Our work in this paper focuses on public opinion mining across multiple media collections. Specifically, we aim to answer the following interesting questions: 1) What are the main concerns with a public event for users being active in different media platforms? Do these concerns share anything in common or just scatter in different media? 2) What are the public opinions to these concerns? Do the users in different media platforms have consistent or diversed opinions?

To answer these questions, we first need an aspect-based opinion mining model for capturing public concerns and opinions simultaneously from text collections. Here *aspect* means an underlying theme, perspective or viewpoint as to a subject like an event or a product, with the assumption that each sentence is generated by a single aspect. It has been reported that about 83% sentences in online reviews cover a single aspect [27], implying this assumption holds particularly for online social media such as Twitter or Chinese Weibo.

In addition, we need to build a cross-collection framework for the aspect-based opinion model so as to enable information integration from different collections. Actually a cross-collection model could benefit more. Suppose we want to mine opinions from both news and micro-blogs collections. Since micro-blogs are mostly generated by public users, referred as user generated content (UGC), we could expect sharper opinions from UGC but less clear concerns or aspects due to the more emotional and colloquial expression way. This is in sharp contrast to the news, where the concerns are usually very clear but the opinions are often monotonous and implicit. In other words, we have *asymmetric collections* or *complimentary collections* for public opinion analysis. Therefore, an intuitional way to display both sides' respective advantages is to use news to help tweets identify meaningful aspects and to use tweets to enrich news by diverse opinions. This is what we called *Complementary Aspect-based Opinion Mining* task defined as follows:

> *Given multiple text collections about a subject, design a cross-collection model that can leverage complementary information from different collections to form aspects-based opinions for comprehensive and contrastive public opinion analysis.*

Remark. There have been some studies on cross-collection topic modeling in the literature, and to our best knowledge the ones most related to our task include [4] and [5]. While [4] also studies contrastive opinion mining problem, it does not jointly model aspects and opinions and pays little attention to the complementarity of aspects across collections, which however is our main focus. The task of [5] is to summarize text across complementary collections, but it ignores public opinions totally, which is also the main theme of our study.

III. Model and Inference

In this section, we introduce the *Cross-collection Auto-labeled MaxEnt-Lda* model (CAMEL for short) for aspect-based opinion mining across complementary collections. Specifically, we first describe the key points of CAMEL as well as the generative process under CAMEL. We then introduce the Auto-labeled MaxEnt model (AME for short) that enables the subsequent joint modeling of aspects and opinions without expensive human labeling. We finally present the approximate posterior inference for CAMEL.

A. Model Description

As illustrated in Sect. II, our main task is to design a model that can leverage complementary information from diverse collections to jointly model aspects and opinions for public opinion analysis. Along this line, two key problems should be well addressed in the model. The first one is how to model aspects and opinions hidden in a collection in a simultaneous and automatical way, which will be discussed in detail in Sect. III-B. The second problem is how to capture the complementarity across multiple collections with possible severely asymmetric information, which is the main focus of this subsection.

Let us again consider the case of aspect-based opinion mining from two asymmetric collections, i.e., the news collection with clear aspects and the tweet collection with sharp opinions. In the *aspect* level, to help extract clear aspects from tweets, it is intuitive to share with the tweets side some similar aspects found from the news side. This can be done by mining aspects from the two collections separately, and then linking together similar aspects from different sides. However, it would be very difficult, if not impossible, to define a proper similarity measure and set a good threshold to it. Therefore, it would be better to design a cross-collection model that could directly mine some common aspects shared by different collections. In the *opinion* level, however, it would be more interesting to read public opinions from the tweet side, and compare them with the opinions from the news side, which are often regarded as the mainstream opinions from authoritative media. As a result, our cross-collection model should be able to mine the opinions separately from different sides for the purpose of comparison.

Based on the above reasoning, we now can describe our model CAMEL, and give the generative process under it. CAMEL is essentially a cross-collection LDA model with a maximum entropy model embedded to determine the priors for aspect and opinion words switching. CAMEL assumes that different collections not only share some common aspects but also have aspects of their own. Hereinafter, we call aspects shared across collections as *common aspects*, and call the aspects only contained in one collection as *specific aspects*. CAMEL also assumes that each specific aspect has a corresponding opinion, and each common aspect has multiple corresponding opinions, one for each collection. We now describe how to generate a document under CAMEL as follows.

Suppose there are several multinomial word distributions from a symmetric Dirichlet prior with parameter β, including: K^I common aspects $\{\phi_z^A\}_{z=1}^{K^I}$ shared by all collections, with C opinions $\{\phi_{z,c}^O\}_{c=1}^{C}$ for each ϕ_z^A , where C is the number of collections; K^S specific aspects $\{\psi_z^A\}_{z=1}^{K^S}$ and K^S cor-

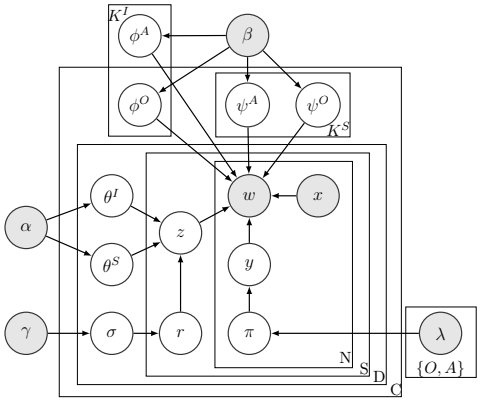


Fig. 1: Plate notation of CAMEL.

responding opinions $\{\psi_z^O\}_{z=1}^{K^S}$, one for each collection. All these are multinomial distributions over the vocabulary, which we assume has V words in total. Note that we here assume all collections have the same number of specific aspects for briefness, which could be relaxed to allow variant easily.

For sentence s in document d, we draw $r_{d,s}$ from a Bernoulli distribution over $\{0, 1\}$ parameterized by σ, which in turn is drawn from a symmetric $Beta(\gamma)$. $r_{d,s}$ is an indicator of whether sentence s is generated by common or specific aspects. Specifically, when $r_{d,s} = 0$, we assume a sentence is generated by a common aspect, otherwise by a collection-specific aspect. For word n in sentence s, we introduce an indicator variable $y_{d,s,n}$ for aspect and opinion switching, which is drawn from a Bernoulli distribution over $\{0, 1\}$ parameterized by π. Similar to $r_{d,s}$, we assume a word is generated by an aspect-word distribution when $y_{d,s,n} = 0$, otherwise by an opinion-word distribution. According to some previous studies [16], [14], topic models that set π with symmetric priors are unable to identify opinion words well. Therefore, to set π for word $w_{d,s,n}$, we utilize the weights learned by the maximum entropy component and the feature vector $\mathbf{x}_{d,s,n}$ of w, which give:

$$p(y_{d,s,n} = l|\mathbf{x}_{d,s,n}) = \pi_l^{d,s,n} = \frac{exp(\lambda_l \cdot \mathbf{x}_{d,s,n})}{\Sigma_{l'=0}^{1} exp(\lambda_{l'} \cdot \mathbf{x}_{d,s,n})},$$

where $\{\lambda_0, \lambda_1\}$ denote the weights learned by the maximum entropy model upon a set of training data, whose labels are obtained by an automatic procedure described in the next subsection.

The plate notation of CAMEL is shown in Fig. 1, with a summary of math notations provided in Table I. The generative process is thus described as follows:

1) For each common aspect z:
   a) Choose $\phi_z^A \sim Dir(\beta)$
2) For each collection c:
   a) Choose $\phi_{z,c}^O \sim Dir(\beta)$ for each common aspect z
   b) Choose $\psi_{z,c}^A, \psi_{z,c}^O \sim Dir(\beta)$ for each collection-specific aspect z
3) For each document d:
   a) Choose a collection indicator c
   b) Choose $\theta_d \sim Dir(\alpha)$
   c) Choose $\sigma_d \sim Beta(\gamma)$
   d) For each sentence s:
      i) Choose $r_{d,s} \sim Bern(\sigma_d)$
      ii) if $r_{d,s} = 0$ choose $z_{d,s} \sim Multi(\theta_d^I)$
          if $r_{d,s} = 1$ choose $z_{d,s} \sim Multi(\theta_d^S)$
      iii) For each word n:
          A) Choose $y_{d,s,n} \sim Bern(\pi_{d,s,n})$
          B) if $r_{d,s} = 0$ and $y_{d,s,n} = 0$ choose $w_{d,s,n} \sim Multi(\phi_{z_{d,s}}^A)$
             if $r_{d,s} = 0$ and $y_{d,s,n} = 1$ choose $w_{d,s,n} \sim Multi(\phi_{z_{d,s},c}^O)$
             if $r_{d,s} = 1$ and $y_{d,s,n} = 0$ choose $w_{d,s,n} \sim Multi(\psi_{z_{d,s},c}^A)$
             if $r_{d,s} = 1$ and $y_{d,s,n} = 1$ choose $w_{d,s,n} \sim Multi(\psi_{z_{d,s},c}^O)$

TABLE I: Math notations.

Notation	Description
K^I	the # of common aspects in total
K^S	the # of specific aspects for each collection
C	the # of collections
D	the # of documents in a collection
S	the # of sentences in a document
N	the # of words in a sentence
ϕ^A	common aspect-word distribution
ϕ^O	common opinion-word distribution
ψ^A	specific aspect-word distribution
ψ^O	specific opinion-word distribution
θ^I	common aspect mixture for a document
θ^S	collection specific aspect mixture for a document
σ	parameters for common and specific aspect switching for a sentence
π	parameters for aspect and opinion switching for a word
w	an observed word
z	aspect index for a sentence
x	feature vector for the maximum entropy (MaxEnt) model
λ	weights learned by the MaxEnt model
y	aspect and opinion switcher for a word
r	common and specific aspect switcher for a sentence
α	Dirichlet prior parameter for θ
β	Dirichlet prior parameter for all word distributions
γ	symmetric Beta prior parameters for σ
$Bern(\cdot)$	Bernoulli distribution with parameter$(\cdot)$
$Beta(\cdot)$	Beta distribution with parameter$(\cdot)$
$Multi(\cdot)$	Multinomial distribution with parameter$(\cdot)$
$Dir(\cdot)$	Dirichlet distribution with parameter$(\cdot)$

B. Auto-labeled MaxEnt Model

In order to obtain aspect-specific opinions, we adopt the MaxEnt-LDA model proposed by Zhao et al. in [27], where a maximum entropy (MaxEnt) model is trained with Part-Of-Speech (POS) tags of words serving as priors for aspect and opinion switching. This is motivated by the fact that aspect and opinion terms normally play different syntactic roles in a sentence, but it also suffers from the high cost of manual labeling of word tags.

To address this problem, we propose a procedure to label training data automatically, and thus form the so-called Auto-labeled MaxEnt model (AME). It is motivated by the observa-

tion that opinion words usually do not appear nearly to each other in a sentence. This in other words implies that a word nearby a known opinion word is likely to be a non-opinion word. The following gives the details of the procedure:

1) We first randomly select a set of M opinion words $\{V_m^{(O)}\}_{m=1}^M$ from a general opinion lexicon, which is usually publicly available for many languages. Those selected words are required to have relatively large document frequency in the target corpus.
2) We then randomly choose a set of sentences S such that each sentence in S contains at least one opinion word in $\{V_m^{(O)}\}_{m=1}^M$.
3) We label a word as an aspect word if it is not in $\{V_m^{(O)}\}_{m=1}^M$ and appears nearly to a known opinion word in a sentence contained by S. In this way, we finally obtain M opinion words $\{V_m^{(O)}\}_{m=1}^M$ and N aspect words $\{V_n^{(A)}\}_{n=1}^N$.

With the opinion and aspect words extracted via the above procedure, we can obtain the POS tag features from their context to train the MaxEnt model, of which we will not go into the details due to the page limit.

C. Approximate Posterior Inference

It is obvious that exact posterior inference is intractable in CAMEL, so we turn to a collapsed Gibbs sampling algorithm [6] for approximate posterior inference, which is simple to derive, comparable in speed to other estimators, and can approximate a global maximum. Due to the space limit, we leave out the derivation details and only present the sampling formulas. Note that the MaxEnt component is trained before we perform Gibbs sampling, which means $\{\lambda_0, \lambda_1\}$ are fixed during Gibbs sampling.

In CAMEL, we have three sets of latent variables: z, r and y. Given the assignments of all other hidden variables, we can jointly sample $(z_{d,s}, r_{d,s})$ as a block:

$$P(z_{d,s}=k, r_{d,s}=j|\mathbf{z}_{\neg(d,s)}, \mathbf{r}_{\neg(d,s)}, \mathbf{y}, \mathbf{w}, \mathbf{x})$$
$$\propto \frac{C^d_{(j)}+\gamma}{C^d_{(\cdot)}+2\gamma} \times \frac{C^{d,j}_{(k)}+\alpha}{C^d_{(j)}+K^j\alpha}$$
$$\times\left(\frac{\Gamma\left(C^{A,j,k}_{(\cdot)}+V\beta\right)}{\Gamma\left(C^{A,j,k}_{(\cdot)}+N^{A,j,k}_{(\cdot)}+V\beta\right)}\cdot\prod_{v=1}^{V}\frac{\Gamma\left(C^{A,j,k}_{(v)}+N^{A,j,k}_{(v)}+\beta\right)}{\Gamma\left(C^{A,j,k}_{(v)}+\beta\right)}\right)$$
$$\times\left(\frac{\Gamma\left(C^{O,j,k}_{(\cdot)}+V\beta\right)}{\Gamma\left(C^{O,j,k}_{(\cdot)}+N^{O,j,k}_{(\cdot)}+V\beta\right)}\cdot\prod_{v=1}^{V}\frac{\Gamma\left(C^{O,j,k}_{(v)}+N^{O,j,k}_{(v)}+\beta\right)}{\Gamma\left(C^{O,j,k}_{(v)}+\beta\right)}\right).$$

We first consider the case of $j=0$. With this condition, $C^d_{(j)}$ is the number of sentences assigned to common aspects in document d. $C^{d,j}_{(k)}$ is the number of sentences assigned to common aspect k in document d. K^j is the number of common aspects, i.e., K^I equivalently. $C^{A,j,k}_{(v)}$ is the number of times word v is assigned as an aspect word to common aspect k, and $C^{O,j,k}_{(v)}$ is the number of times word v is assigned as an opinion word to common aspect k. $C^{A,j,k}_{(\cdot)}$ is the total number of times any word is assigned as an aspect word to common aspect k, and $C^{O,j,k}_{(\cdot)}$ is the total number of times any word is assigned as an opinion word to common aspect k. $N^{A,j,k}_{(v)}$ is the number of times word v is assigned as an aspect word to aspect k in sentence s of document d, and similarly, $N^{O,j,k}_{(v)}$ is the number of times word v is assigned as an opinion word to aspect k in sentence s of document d. When $j=1$, all counts mentioned above refer to specific aspects. $C^d_{(\cdot)}$ is the number of sentences in document d. Note that all these counts represented by symbol C exclude sentence s of document d.

With assignments of $\mathbf{z}$ and $\mathbf{r}$, we can sample $y_{(d,s,n)}$ for $y_{(d,s,n)}=0$:

$$p(y_{(d,s,n)}=0|\mathbf{z}, \mathbf{r}, \mathbf{y}_{\neg(d,s,n)}, \mathbf{w}, \mathbf{x})$$
$$\propto \frac{exp(\lambda_0\cdot\mathbf{x}_{d,s,n})}{\Sigma^1_{l'=0}exp(\lambda_{l'}\cdot\mathbf{x}_{d,s,n})}\times\frac{C^{A,r_{d,s},z_{d,s}}_{(w_{d,s,n})}+\beta}{C^{A,r_{d,s},z_{d,s}}_{(\cdot)}+V\beta},$$

and for $y_{(d,s,n)}=1$:

$$p(y_{(d,s,n)}=1|\mathbf{z}, \mathbf{r}, \mathbf{y}_{\neg(d,s,n)}, \mathbf{w}, \mathbf{x})$$
$$\propto \frac{exp(\lambda_1\cdot\mathbf{x}_{d,s,n})}{\Sigma^1_{l'=0}exp(\lambda_{l'}\cdot\mathbf{x}_{d,s,n})}\times\frac{C^{O,r_{d,s},z_{d,s}}_{(w_{d,s,n})}+\beta}{C^{O,r_{d,s},z_{d,s}}_{(\cdot)}+V\beta}.$$

Here counts represented by symbol C indicate that word n in sentence s of document d has been excluded.

D. Discussions

We here briefly discuss the similarities and differences between CAMEL and two most related models, i.e., ccTAM [5] and CPT [4], from the perspective of model structure. ccTAM distinguishes common topics from specific ones as CAMEL did, but it does not separate opinions from topics. CPT explicitly separates opinions from topics. However, it needs strict rules to separate opinion and topic words, whereas we provide a more soft way for opinion-aspect switching over a word. Besides, CPT dose not distinguish common and specific topics, which limits its applicability on asymmetric collections.

IV. Experimental Results

In this section, we present extensive experimental results to evaluate CAMEL in both quantitative and qualitative ways. Hereinafter, we agree to use "CAMEL", "our method" and "ours" interchangeably in comparative studies.

A. Experimental Setup

1) Data Sets: Our method is tested on three real world data sets. One data set is a collection of electronic device reviews from Amazon[1], which is used to perform quantitative evaluations of our method. Two real world events data has been crawled and analysed for the qualitative analysis of our method.

The online reviews is collected by Jo et al. [12], which contains electronic device reviews with seven categories. To

[1]http://www.amazon.com

TABLE II: Statistical description of data sets.

Data	#Documents	#Sentences	#Words
Collection0	3,535	56,053	315,471
Collection1	3,659	60,935	339,192
Collection0&1	7,194	116,988	654,663
APEC News	9,662	251,023	3,561,412
APEC Tweets	82,366	144,149	1,136,422
Stampede News	1,015	42,651	262,613
Stampede Tweets	13,004	25,425	74,353

evaluate the quality of common and specific aspects as well as their opinions across collections, we create a new data set based on the reviews. We select reviews under three categories, namely coffee machine, canister vacuum and MP3 player, to build a data set with two collections. Intuitively, the reason for selecting these three categories is that the reviews under these categories have minimal overlap in contents, which can help to ensure aspects of common category can be well distinguished from those of specific category.

To build the new data set, we first place reviews labeled as canister vacuum into collection C_0 and reviews labeled as MP3 player into collection C_1, then randomly inject sentences of coffee machine reviews into reviews in C_0 or C_1. This way of making the new data set is according with characteristics of real world asymmetric collections, since common aspects are shared by collections and each collection has its own specific aspects. Here sentences from coffee machine is regarded as common part across collections, and canister vacuum or MP3 Player is regarded as specific part only occurs in C_0 or C_1 respectively. As to real word events, we crawled news and tweets related to event *2014 Shanghai stampede* and *2014 Beijing APEC*.

All data sets go through the same preprocessing process: first applying POS tagging and then get automatic labeled training data for Maximum Entropy(MaxEnt) model as illustrated in Sect.III-B. At last, we remove stop words and those with low document frequency. For tweets, we also remove URLs and Hashtags. We use Stanford POS Tagger[2] to tag English online reviews and LTP-Cloud[3] to tag Chinese news and tweets. The opinion lexicon used in auto-labeled MaxEnt model for English corpus is collected by Hu and Liu [9]. As to Chinese corpus we use an opinion lexicon merged from two widely used Chinese opinion lexicons. Details of preprocessed data sets are shown in Table II

2) Baseline Methods: In the quantitative experiments, we compare our method with three baseline methods. All baseline methods are MaxEnt-LDA [27] run over different collection configurations.

1) BL-0: Perform MaxEnt-LDA over collection C_0
2) BL-1: Perform MaxEnt-LDA over collection C_1
3) BL-2: Perform MaxEnt-LDA over collection C_0 and C_1.

Since we adopt the MaxEnt-LDA in our model to obtain aspect-specific opinions, comparing with these baselines can give us insights of whether aspects and opinions induced by our method can benefit from complementary aspect-based opinion mining by explicitly separating common and specific aspects.

3) Evaluation Measures: We briefly introduce two sets of measures used in experiments, one set is macro-averaged precision, recall and f-measure, the other set is opinion coherence and aspect-opinion coherence.

Since each online review has an category, we can leverage this supervised information to evaluate aspects learned by our method as well as baseline methods. As all methods assign one aspect to each sentence, therefore, we label a sentence with the category of the review it resides in. By manually mapping learned aspects to their corresponding category, we can take a review's category as "ground truth" to evaluate them by performing sentence classification.

To compare the quality of opinions, we choose an automatic measure called Topic Coherence [17], which has been widely used in evaluating topics and has been justified in according with human evaluations. We also slightly modify the coherence score to measure relevance of aspect and its opinion. Details please refer to Sect.IV-C1

B. Aspect Evaluation

We design a sentence classification experiment on reviews to evaluate the quality of aspects learned by our method. Note that all baseline methods as well as our method assigns one aspect to each sentence, and each of induced aspect often corresponds to one category, thus we can use sentence classification as the evaluation method for learned aspects. Specifically, better sentence classification results indicate better quality of aspects. In order to illustrate asymmetric aspect-based opinion mining, we created a dataset with common aspects across collections and specific ones only reside in each collection as described in Sect.IV-A1.

Each review in our data set has one category in the set $L = \{$coffee machine, canister vacuum, MP3 player$\}$. We use the category of the review to label its sentences. Sentences of coffee machine reviews are injected into reviews in collection C_0 or collection C_1 randomly. Therefore, coffee machine sentences exist across collections. Sentences of canister vacuum or MP3 player are only contained in C_0 or C_1. In other words, we expect aspects about coffee machine as common aspects, those related to remain two categories as specific aspects.

We run BL-0, BL-1, BL-2 and our method over the data to get inferred aspect set S^A and aspect assignment k of each sentence. In order to perform sentence classification evaluation, we have to manually map each aspect to one of the three categories, i.e., we created a mapping function $f(k) : S^A \rightarrow L$. Aspects can not be mapped to any category are labeled as *other*. Given the mapping, we can get predicted labels of sentences for each method. Then we evaluate all methods according to metrics, namely *precision*, *recall* and *f-measure*.

For all methods, we set $\alpha = 0.1$, $\beta = 0.01$. For our method, we set $\gamma = 0.1$. In order to keep all methods comparable, we have to set proper number of aspects for each method. Specifically, we set aspect number K_2 of BL-2 equals to the sum of K_0 and K_1, where K_0 is the aspect number for BL-0

[2]http://nlp.stanford.edu/software/tagger.shtml
[3]http://www.ltp-cloud.com/

TABLE III: Sentence classification results.

Method	$K=5$			$K=10$			$K=15$			$K=20$		
	precision	recall	f-measure	precision	recall	f-measure	precision	recall	f-measure	precision	recall	f-measure
BL-2*	**0.819**	0.783	0.799	0.745	**0.791**	0.765	0.795	0.777	0.785	**0.833**	0.747	0.787
Ours*	0.791	**0.833**	**0.820**	**0.861**	0.783	**0.820**	**0.829**	**0.812**	**0.820**	0.819	**0.820**	**0.818**
BL-0*	**0.866**	0.646	0.734	**0.856**	0.736	0.789	**0.864**	0.725	0.787	**0.846**	0.726	0.775
BL-2-C_0*	0.804	0.780	0.789	0.695	**0.794**	0.733	0.771	0.776	0.770	0.824	0.743	0.781
Ours-C_0*	0.791	**0.819**	**0.801**	**0.856**	0.768	**0.810**	0.820	**0.801**	**0.810**	0.816	**0.801**	**0.808**
BL-1*	**0.891**	0.727	0.800	0.858	0.780	0.815	**0.879**	0.780	0.826	**0.877**	0.767	0.818
BL-2-C_1*	0.836	0.787	0.809	0.826	0.787	0.806	0.827	0.779	0.802	0.843	0.751	0.793
Ours-C_1*	0.791	**0.847**	**0.813**	**0.865**	**0.798**	**0.829**	0.838	**0.822**	**0.830**	0.821	**0.833**	**0.827**
BL-2$\diamond$	0.825	0.836	0.827	0.797	0.761	0.767	0.810	0.814	0.809	0.810	0.852	0.829
Ours$\diamond$	**0.909**	**0.876**	**0.891**	**0.919**	**0.861**	**0.889**	**0.923**	**0.856**	**0.888**	**0.922**	**0.858**	**0.889**
BL-0$\diamond$	0.826	**0.943**	0.880	0.861	**0.927**	**0.892**	0.857	**0.934**	**0.893**	0.855	**0.913**	0.880
BL-2-$C_0\diamond$	0.889	0.841	0.862	0.882	0.689	0.753	0.886	0.798	0.836	0.892	0.846	0.868
Ours-$C_0\diamond$	**0.896**	0.871	**0.882**	**0.904**	0.865	0.884	**0.910**	0.852	0.880	**0.907**	0.860	**0.883**
BL-1$\diamond$	0.875	**0.954**	**0.912**	0.894	**0.933**	**0.913**	0.895	**0.946**	**0.920**	0.890	**0.945**	**0.917**
BL-2-$C_1\diamond$	**0.935**	0.832	0.880	**0.935**	0.833	0.880	**0.940**	0.830	0.880	0.930	0.857	0.891
Ours-$C_1\diamond$	0.921	0.881	0.899	0.934	0.857	0.894	0.934	0.859	0.895	**0.937**	0.856	0.894
BL-2	0.823	0.819	0.818	0.780	0.771	0.766	0.805	0.802	0.801	0.818	0.817	0.815
Ours	**0.869**	**0.862**	**0.863**	**0.869**	**0.860**	**0.863**	**0.868**	**0.859**	**0.862**	**0.869**	**0.860**	**0.863**
BL-0	0.846	0.794	0.807	0.859	**0.832**	0.841	0.860	**0.829**	0.840	0.851	0.819	0.828
BL-2-C_0	**0.847**	0.811	0.826	0.789	0.742	0.743	0.828	0.787	0.803	**0.858**	0.795	0.824
Ours-C_0	0.843	**0.845**	**0.842**	**0.880**	0.817	**0.847**	**0.865**	0.827	**0.845**	0.843	**0.831**	**0.846**
BL-1	0.883	0.841	**0.856**	0.876	**0.856**	**0.864**	**0.887**	**0.863**	**0.873**	0.884	**0.856**	**0.867**
BL-2-C_1	**0.886**	0.809	0.844	0.881	0.810	0.843	0.883	0.804	0.841	**0.886**	0.804	0.842
Ours-C_1	0.856	**0.864**	**0.856**	**0.899**	0.828	0.862	0.886	0.841	0.862	0.879	0.844	0.860

and K_1 is the aspect number for BL-1. Setting common aspect number K^I and specific aspect number K^S in our method needs to satisfy $K^I + CK^S = K_2$, where C is the number of collections and in this case $C = 2$. In this experiment, we keeps $K = K_0 = K_1$, and range K from 5 to 20, therefore K_2 ranges from 10 to 40. We set $K^I = 2, 4, 6, 8$ and $K^S = 4, 8, 12, 16$ for $K = 5, 10, 15, 20$ respectively. Here we set $K^S = 2K^I$, because we know the approximate proportion of common aspects in each collection. However, when dealing with data has no supervised information(or prior knowledge) as we did in our case study, one might has to try several configurations of K^I and K^S to find one appropriate setting, or resorts to nonparametric Bayesian inference. For all methods, we run 1000 iterations of Gibbs sampling, and each parameter configuration runs 10 samples.

The sentence classification results are reported in Table III. All results are the average of 10 samples. The marker $*$ indicates metrics are computed for the common category classification, and the marker $\diamond$ indicates metrics are computed for specific categories classification, and the results without marker are macro average of all categories, which reflect the overall performance of each method. The suffix C_0 or C_1 appends to BL-2 and our method indicates the measures are computed on the sentences in C_0 or C_1 respectively. With out suffix indicates the measures are computed on all sentences in C_0 and C_1.

From the comparison of BL-2 with our method on average classification results, as well as the common and specific categories classification results, we find our method consistently outperforms BL-2. Besides, our method performs quite stable with different K, however, the performance of BL-2 changes notably. Since BL-2 performs aspect-opinion mining direct on combined collections, it fails in leveraging the structure of common and specific aspects underneath the asymmetric collections that we focused in this paper. Our method explicitly separates aspects shared by collections and those specific to each collection, therefore, aspects in one collection serve as complementary information for aspects extraction in other collections in an mutual way.

The results of common category classifications confirm our above discussions. From the group of results marked with $*$, we can see our method still consistently outperforms all baseline methods on *recall* and *f-measure* in almost all cases. This promising result indicates that complementary aspect mining can great benefit the accuracy of learned common aspects. Although precision results of our method are better than most of those of BL-2, we also notice that they are slightly worse than those of BL-1 in most cases. The results of specific category classifications, i.e., results marked as $\diamond$, are in contrary to those in common category classifications, as our method gains relative high precision but low recall as compared to BL-0 or BL-1.

In order to illustrate this phenomenon, we check the mapping of aspect to category for our method, and find there is always an common aspect be mapped to category *other*, when $K > 5$. Then we look into those aspects and find they are all related to *after-sales service*. We give one sample of after-sales service aspect-opinions in TableIV, where opinion0(or opinion1) means opinions induced from C_0(or C_1). Note that after-sale service is independent to review categories and obviously an common aspect across review collections. There is no label named after-sale service, therefore, sentences related to after-sale service but originally labeled as other categories are misclassified by our method. Since BL-0 and BL-1 mixed after-sales service aspect with other aspects, therefore, no aspect is mapped to *other*. We guess this might be the reason why the precisions of our model in common category classifications as well as the recalls in specific category classifications are

TABLE IV: After-sales service aspect and opinions.

aspect	service call customer back return product problem amazon buy warranty
opinion0	send free good great ship local wrong long happy fast
opinion1	send work great free good local defective creative original easy

less comparative with BL-0's or BL-1's. However, except for the common aspect we called after-sale service, other aspects can be clearly mapped to an category, unless it is not ratable. Therefore, we can conclude from the classification results that our method can learn more accuracy aspects by performing complementary aspect-based opinion mining.

C. Opinion Evaluation

In this subsection, we give evaluations of opinions learned by our method. In most previous works [3], [27], the evaluation of opinions is highly relied on the judgment of humans, which is costly and some times bias from the data. Therefore, we try to use some automatic measures to evaluate opinions. Recently, an automatic measure named Topic Coherence [17] has been proposed to evaluate the coherence of distributions learned by topic models, which has been justified in according with human evaluations. Therefore we use coherence score to evaluate opinions, and further propose a new measure based on it to evaluate the coherence(or relevance) between aspect and its corresponding opinion.

1) Opinion Coherence: Coherence score measures a single word distribution by computing the semantic similarity degree between high probability words in it. Higher score often indicates better quality. Given T high probability words of an opinion, the coherence score for the opinion is defined as

$$C_O(k; V^{(k)}) = \sum_{t=2}^{T} \sum_{l=1}^{t-1} \log \frac{D(v_t^{(k)}, v_l^{(k)}) + \epsilon}{D(v_l^{(k)})},$$

where $V^{(k)} = (v_1^{(k)}, \ldots, v_T^{(k)})$ is a list of T most probable words in opinion k. $D(v)$ counts the number of documents containing the word v, and $D(v, v')$ counts the number of documents containing both v and v'. ϵ is a smoothing variable used to avoid taking the log of zero for words that never co-occur.

2) Aspect-Opinion Coherence: Coherence score introduced above can only evaluate the quality of one opinion. It can not evaluate the coherence between aspect and its opinion. Since our method learns pairs of aspect and opinion, we propose a new measure for the evaluation of aspect opinion pairs. Given T high probability words of an aspect and its opinion respectively, the coherence score for one pair of aspect and opinion is defined as

$$C_{A,O}(k; V^{A,(k)}, V^{O,(k)}) = \sum_{t=1}^{T} \sum_{l=1}^{T} \log \frac{D(v_t^{A,(k)}, v_l^{O,(k)}) + \epsilon}{D(v_t^{A,(k)})},$$

where $V^{A,(k)}$ is a list of T most probable words in aspect k, and $V^{O,(k)}$ is a list of T most probable words in opinion k. Loosely speaking, the value of $\frac{D(v^{A,(k)}, v^{O,(k)})}{D(v^{A,(k)})}$ estimates the probability one could observe the opinion word $v^{O,(t)}$ if he or she already have observed the aspect word $v^{A,(t)}$ in an document. For both coherence scores, we set $\epsilon = 1^{-12}$ to reduce the score for completely unrelated words as suggested in [22].

TABLE V: Results of opinion and aspect-opinion coherence.

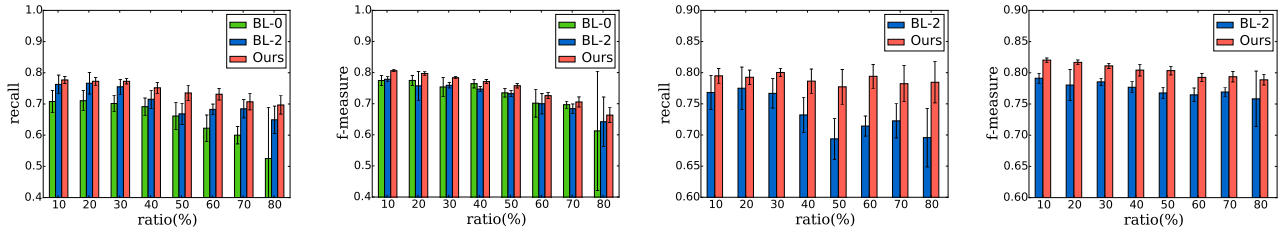
Method	opinion coherence		aspect-opinion coherence	
	$T = 10$	$T = 15$	$T = 10$	$T = 15$
BL-0	-121.1±3.6	-307.6±7.2	-229.3±5.3	-572.1±7.1
Ours-C_0	**-119.4±4.3**	-307.7±9.0	**-226.2±6.2**	-571.3±15.7
BL-1	-129.8±3.7	-334.9±11.9	-246.7±5.5	-621.1±11.6
Ours-C_1	**-127.7±1.7**	**-326.7±5.4**	-245.1±3.2	**-613.0±4.8**
BL-2	-134.8±2.7	-345.6±7.9	-255.2±3.5	-640.7±12.1
Ours	**-129.8±2.0**	**-329.7±4.1**	**-240.6±3.4**	**-604.0±6.4**
Ours$*$	**-103.7±2.6**	**-260.9±3.9**	**-192.3±3.0**	**-483.5±4.8**
Ours-C_0*	-127.5±4.6	-309.6±6.5	-246.6±6.5	-598.8±20.1
Ours-C_1*	-126.0±4.5	-320.2±9.6	-245.9±6.0	-612.1±10.2

We compare our method with baseline methods based on average opinion coherence score and average aspect-opinion coherence score of all induced aspect-based opinions. For all methods, we set $\alpha = 0.1$, $\beta = 0.01$. For our method, we set $\gamma = 0.1$. To make all methods comparable, we have to set proper number of opinions for each method. Note that for baseline methods, setting number of opinions is the same as setting number of aspects, details in Sect.IV-B. As our method learns C opinions for each common aspect, where C is the number of collections, We have to set $C(K^I + K^S)$ equals to K_2. In this experiment, we set $K = K_0 = K_1 = 15$, $K_2 = 30$, and $K^I = 5, K^S = 10$.

The average opinion coherence and aspect-opinion coherence results with $T = 10$ and $T = 15$ are illustrated in Table V. The suffix C_0 or C_1 appends to our method indicates we evaluate coherence of common aspect-opinions and specific ones on C_0 or C_1. The marker $*$ indicates the average coherence score for common aspect-opinions only. From the results, we can see our method outperforms BL-2 on both opinion coherence and aspect-opinion coherence with different T. This result indicates taking complementary aspect-opinions shared across collections into account can not only help to find more accuracy aspects but also improve the opinion coherence and enhance the relevance of aspect and opinion. We can also find the overall performance of our method surpass BL-0 and BL-1. Since BL-0 and BL-1 run over single collection, which limits its ability in separating common aspect-opinions and specific ones, and therefore hurts its opinion coherence and aspect-opinion relevance. We also compare the common aspect-opinions of our method on collection C_0, C_1 and both. The result is promising, as it illustrates those common aspect-opinions are more coherence on combined collections than on single one, which further demonstrates the complementary aspect-based opinion mining is necessary for asymmetric collections.

D. Complementary Aspect Mining Evaluation

We compare performance of baseline methods and ours, given situations that common aspects across collections are highly imbalanced. This experiment can give us insights about how collections lack of common aspects can benefit from complementary informations in other collections. In order to



(a) C_0: recall. (b) C_0: f-measure. (c) C_0 and C_1: recall. (d) C_0 and C_1: f-measure.

Fig. 2: Common category classification results with different imbalance ratios on C_0 or C_0 combined with C_1.

simulate the imbalance circumstances, we reform the online reviews data used in Sect.IV-B by removing sentences labeled as coffee machine(the common category) from collection C_0, while collection C_1 remains the same.

To creat different levels of imbalance, we remove coffee machine sentences according to ratios from 10% to 80%, which is the proportion of coffee machine sentences we remove from collection C_0. With reformed data set, we can perform coffee machine sentences classification on C_0 and C_0 combined with C_1 to discuss the complementary aspect mining. We set $\alpha = 0.1$, $\beta = 0.01$ for all methods, $\gamma = 0.1$ for our method. $K = K_0 = K_1 = 15$, $K_2 = 30$, $K^I = 6$ and $K^S = 12$. The resulting *recall* and *f-measure* are illustrated in Fig. 2.

In Fig.2a and Fig. 2b we can see without complementary information, the recall and f-measure of BL-0 both drop rapidly and become extremely unstable when ratio reaches 80%. Since BL-2 and our method both utilize complementary common aspects in C_1, therefore is relative stable than BL-0. As we could see in Fig 2, the performance of BL-2 on C_0 as well as on C_0 and C_1, also become rather poor and unstable as ratio reaches 80%. This promising result strongly suggested the explicit separation of common aspects and specific aspects is necessary for complementary aspect mining.

E. Validity of Auto-labeled MaxEnt

To justify the validity of our auto-labeled MaxEnt component, we give comparisons of manual labeled MaxEnt(MME for short) with auto-labeled MaxEnt(AME for short). We compare the $Precision@n$($P@n$ for short) of MME with AME on reviews data with varying number of training sentences S. Here $P@n$ means how many words are precisely opinion words other than aspect words given the top n probability words of an opinion, which is judged by human. For MME, we select 50 sentences with opinions words and manually label them. As to AME, we use our automatic procedure to acquire the same number of training sentences. We increase the size of training data inputted into the MaxEnt, and compare the $P@5$, $P@10$ and $P@20$ of MaxEnt-LDA [27] with different sources of training data. The results are reported in Table VI.

From the results, we can find that AME is inaccuracy when training size is small. However, when training size is larger than 30, AME is even more accuracy than MME. Note that the training data of AME is obtained fully automatically, therefore, we can set a larger training size without the concern of costly in time and other resources. Thus, we conclude our AME is an effective method to avoid manually label training data for the MaxEnt component in our model.

TABLE VI: Average $P@n$ of aspect-specific opinion words with different sizes of training data from MME and AME.

Size	$P@5$		$P@10$		$P@20$	
	MME	AME	MME	AME	MME	AME
$S = 10$	**0.90**	0.64	**0.82**	0.58	**0.71**	0.51
$S = 20$	**0.80**	0.64	**0.74**	0.56	**0.67**	0.50
$S = 30$	**0.86**	0.84	**0.82**	0.81	0.70	**0.76**
$S = 40$	0.80	**0.84**	0.75	**0.85**	0.71	**0.78**
$S = 50$	0.82	**0.90**	0.81	**0.83**	0.71	**0.76**

TABLE VII: Complementarity between news & tweets.

Tweets:	谢谢你还有和你一起喊往后退的那些好人! Thanks to you and those who also shouted fallback, you are good man! 冷静机智的后退哥，太棒了! How calm and wise the fallback shouters are, great!
News:	事发时现场齐声喊出“后退！”“后退！”声音者大约有上百人。 When the event took place, there are hundreds of people shouting fallback! fallback! 事后，网友给这个群体取名为“后退哥”。 After the event, this group of people are called “fallback shouters” by Internet users.

F. Case Study

In this subsection, we apply our method to news and tweets from two real world events, namely *2014 Shanghai Stampede* and *2014 Beijing APEC*, to illustrate its advantages in the complementary aspect-opinion mining. Hereinafter, we refer to *2014 Shanghai Stampede* as Stampede and refer to *2014 Beijing APEC* as APEC. We show some discovered common aspect-opinions to demonstrate how concrete aspects in news can help tweets to find more clear aspects and how rich opinions resides in tweets supplement to aspects of news. At last, we give brief discussion about what aspects are those concerned only in news or tweets.

To illustrate how concrete news aspects supplement to unclear tweets aspects, as well as how tweets enrich opinions of news aspects, we list several tweets and sentences in news, as shown in Table VII.

As we can see, the aspect of above tweets is less clear, since one may not know who yelled fallback, when and why, however opinions are very intense. The sentences of news are just the opposite. Obviously, those tweets and sentences of news share the same aspect, and their are assigned to an

TABLE VIII: Sample specific aspect and opinion from tweets.

Aspect	素质(competence) 提高(improvement) 国民(citizen) 有待(needs) 国人(countrymen) 安全(safety) 素养(attainment) 民众(the_public) 秩序(order) 国家(nation) 意识(awareness)
Opinion	提高(improve) 调查(investigate) 客观(objective) 需要(need) 文明(civilized) 宽慰(comfort) 安全(safe) 教育(educate) 偶然(accidental) 无序(unordered)

TABLE IX: Sample specific aspect and opinion from news.

Aspect	事件(event) 教训(lesson) 上海(Shanghai) 跨年(new_year) 事故(accident) 原因(reason) 政府(government) 报告(report) 生者(survivor) 台阶(stairs)
Opinion	调查(investigate) 客观(objective) 处理(dispose) 公布(announce) 严查(strictly_investigate) 尊重(respect) 真实(true) 宽慰(comfort) 关注(care) 认真(serious)

common aspect by our method. We show this common aspect and its opinions as well as several other samples induced by our method from Stampede and APEC in Table.X.

The common aspect 0 and 1 are extracted from Stampede, the rest two are extracted from APEC, and opinion 0 means the opinion words are extracted from tweets, otherwise from news. Common aspect 1 is about the *fallback shouters* that refers to a bunch of people who yelled fallback to alert the crowd the break out of stampede, which saved lots of lives. From the opinion words in tweets, we can find the public is mainly touched by fallback shouters and appreciates their behavior. While there is no obvious opinions from news. Another case of Stampede is common aspect 2, which is about the announcement of namelist of victims in the event. Most people feels sympathize with those who dead young, while some people criticizes those young people died because of their ignorance. Common aspect 3 is about event APEC, and it is about the reform of China lead by Chairman Xi. Opinions from tweets reflect people's confidence and expectation on this government. Common aspect 4 is about the free-trade agreement in APEC, which also raises the concern of the public, since they can benefit from this agreement.

From all those common aspect-opinions, we can see the opinions extracted from tweets are obviously more emotional, while those extracted from news is monotonous. The common aspects can be easily recognized, which is less possible if we directly extract them from tweets. Besides those showed in this paper, there are also some other common aspects, such as *air pollution*, *visa-free* and *APEC holiday* etc. in APEC, and *rescue*, *penalty* etc. in Stampede. All of them are concerned by public and reported frequently by news media. Due to the limit of pages, we are not going to list them all.

With the induced specific aspect-based opinions from news and tweets, we can find aspects concerned by the public while less reported by news, as well as those widely covered by news while less focused by the public. We give two examples, one for tweets and one for news. In the Table.VIII, we can see the public talks a lot about the populace's cultivation, and its opinion words show the quality of the nation needs to be improved and education should be enhanced. While in Table.IX, we can find news reports about the government will draw lessons from the event, however, in tweets people only cares about the punishment of person in charge.

V. Related Work

A. Aspect-based Opinion Mining

Two subtasks are usually involved in this problem, namely, *aspect or feature identification* and *opinion extraction*. Most of the early works on aspect identification are feature-based approaches [9], [18], e.g., applied frequent itemset mining to identify product aspects [15], which normally exert some constraints on high-frequency noun phrases to find aspects. As a result, they usually subject to the risk of producing too many non-aspects and missing low-frequency aspects [7]. Several early works have applied supervised learning to identify both aspects and opinions [10], [11], [25], which, however, needs hand-labeled training sentences and thus is very costly.

In recent years, with the popularity of topic models, more unsupervised methods are proposed for aspect-based opinion mining. For instance, Titov and McDonald propose a multi-grain topic model to learn both global topics and local topics, in which local topics correspond to rateable aspects [23]. Another approach to discover aspects is to fit a topic model to sentences instead of documents. For instance, Brody and Elhadad run the latent dirichlet allocation(LDA) [2] model over sentences instead of documents to extract aspects [3]. Zhao et al. [27] and Jo et al. [12] assume that all words in a single sentence are generated from one topic.

Some researchers take approaches that model topic and sentiment in a unified way. For instance, Lin and He propose a joint topic-sentiment(JTM) model to detects sentiment and topic simultaneously from text [14]. The aspect and sentiment unification model(ASUM) proposed by Jo et al. [12] is similar to JST, the major difference lays in that ASUM assumes each single sentence only covers one topic. The above two models do not explicitly separate topic words and sentiment words. Mei el al. [16] propose a topic sentiment mixture model, which represents positive and negative sentiments as language models separate from topics, but both sentiment language models capture general opinion words only. Brody et al. [3] take a two-step approach by first detecting aspects and then identifying aspect-specific opinion words. Zhao et al. [27] propose a topic model integrating with a maximum entropy model(MaxEnt-LDA) to jointly capture both aspects and aspect-specific opinion words within a topic model. [20]gives detailed discussions about aspect-specific opinion models based on LDA.

B. Cross Collection Text Mining

Zhai et al. [26] introduce a task called "comparative text mining" and proposed a cross-collection mixture(ccMix) model based on probabilistic latent semantic index(pLSA) [8]. The goal of the task is to discover the common themes across all collections and the ones unique to each collection. Paul et al. [21] extend ccMix to the ccLDA model based on LDA for cross-culture topic analysis. Gao et al. [5] propose a cross collection topic aspect model(ccTAM) to perform event summarization across news and social media steams. They assume aspects contained only in tweets can server as a supplementation to those in news. Fang et al. [4] propose

TABLE X: A sample of common aspect-opinions induced from APEC and Stampede data.

Common Aspect 1		**Common Aspect 2**		**Common Aspect 3**		**Common Aspect 4**	
人群 现场 外滩 警察 后退哥 新年 事件 游客 秩序 声音		遇难者 名单 事件 外滩 上海 身份 踩踏 广场 上午 年龄		中国 社会 外交 改革 世界 经济 推进 国际 习近平 法制		中国 关税 APEC 产品 协定 自贸 价格 商品 企业 澳大利亚	
crowd scene bund policeman fallback_shouters new_year event tourist order voice		victims namelist event bund Shanghai identity stampede square forenoon age		China society diplomacy reformation world economy advance international Jinping_Xi legislation		China tariff APEC products agreement free-trade price commodity enterprise Australia	
opinion 0	**opinion 1**	**opinion 0**	**opinion 1**	**opinion 0**	**opinion 1**	**opinion 0**	**opinion 1**
后退(recede)	上面(above)	默哀(grieve)	公布(announce)	加油(cheer)	改革(reform)	进口(import)	投资(invest)
好样的(great)	摔倒(tumble)	同情(sympathize)	核实(verify)	发展(progress)	发展(develop)	便宜(cheap)	降低(reduce)
感动(touched)	疑似(suspected)	祈福(blessing)	受伤(injure)	深入(thorough)	全面(overall)	期待(expect)	取消(cancel)
冷静(calm)	根本(at all)	痛惜(regret)	确认(check)	伟大(great)	深化(deepen)	带来(bring)	开放(open)
希望(hope)	下来(down)	敷衍(perfunctory)	拥挤(crowed)	全面(overall)	创新(innovate)	发展(develop)	重要(significant)
脆弱(tender)	周围(around)	无知(ignorant)	发生(happen)	幸福(happy)	和平(peaceful)	便利(convenient)	审查(investigate)
减少(reduce)	拥挤(crowded)	心痛(distressed)	慰问(sympathise)	文明(civil)	重要(significant)	恢复(recover)	服务(serve)
关键(important)	稳定(stable)	安抚(appease)	哀悼(grieve)	美好(braw)	伟大(great)	分享(share)	出口(export)
危险(dangerous)	吵杂(noisy)	简单(simple)	深切(heartfelt)	强大(strong)	复兴(revival)	自由(free)	促进(promote)
感谢(thank)	附近(nearby)	庆幸(fortunately)	悼念(mourn)	公平(fair)	公平(fair)	影响(influence)	增加(improve)

a cross-perspective topic model(CPT) to perform contrastive opinion modeling. They view opinions with the same topic yet from different news sources as different perspectives, and learns topics(not *aspects*) across collections by perform LDA over aggregated collections, which has no guarantee of finding shared topics, especially when collections is less comparative as news versus tweets in our case.

VI. Conclusions

In this paper, we propose CAMEL, a novel topic model for complementary aspect-based opinion mining across asymmetric collections. By modeling both common and specific aspects and keeping contrastive opinions, CAMEL is capable of integrating complementary information from different collections in both aspect and opinion levels. An automatically labeling scheme is also introduced to further boost the applicability of CAMEL. Extensive experiments and real-world case study on public events demonstrate the effectiveness of CAMEL in leveraging complementarity for high-quality aspect and opinion mining.

Acknowledgment

Deqing Wang was supported by National Natural Science Foundation of China (NSFC) (71501003) and China Postdoctoral Science Foundation funded project (2014M550591). Junjie Wu was supported by National High Technology Research and Development Program of China (SS2014AA012303), NSFC (71322104,71171007,71531001,71471009), National Center for International Joint Research on E-Business Information Processing (2013B01035), Foundation for the Author of National Excellent Doctoral Dissertation of China (201189), and Fundamental Research Funds for Central Universities.

References

[1] Yang Bao, Nigel Collier, and Anindya Datta. A partially supervised cross-collection topic model for cross-domain text classification. In *CIKM '13*, pages 239–248. ACM, 2013.

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.

[3] Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In *HLT '10*, pages 804–812, 2010.

[4] Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. Mining contrastive opinions on political texts using cross-perspective topic model. In *WSDM '12*, pages 63–72, 2012.

[5] Wei Gao, Peng Li, and Kareem Darwish. Joint topic modeling for event summarization across news and social media streams. In *CIKM '12*.

[6] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS'04*.

[7] Honglei Guo, Huijia Zhu, Zhili Guo, XiaoXun Zhang, and Zhong Su. Product feature categorization with multilevel latent semantic association. In *CIKM '09*, pages 1087–1096, 2009.

[8] Thomas Hofmann. Probabilistic latent semantic indexing. In *SIGIR '99*.

[9] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *KDD '04*, pages 168–177, 2004.

[10] Wei Jin and Hung Hay Ho. A novel lexicalized hmm-based learning framework for web opinion mining. In *ICML '09*, pages 465–472, 2009.

[11] Wei Jin, Hung Hay Ho, and Rohini K. Srihari. Opinionminer: A novel machine learning system for web opinion mining and extraction. In *KDD '09*, pages 1195–1204, 2009.

[12] Yohan Jo and Alice H. Oh. Aspect and sentiment unification model for online review analysis. In *WSDM '11*, pages 815–824, 2011.

[13] Kar Wai Lim and Wray Buntine. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *CIKM '14*, pages 1319–1328. ACM, 2014.

[14] Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *CIKM '09*, pages 375–384, 2009.

[15] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *WWW '05*.

[16] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *WWW '07*, pages 171–180, 2007.

[17] David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *EMNLP '11*, pages 262–272, 2011.

[18] Samaneh Moghaddam and Martin Ester. Opinion digger: An unsupervised opinion miner from unstructured product reviews. In *CIKM '10*.

[19] Samaneh Moghaddam and Martin Ester. Aspect-based opinion mining from product reviews. In *SIGIR '12*, pages 1184–1184, 2012.

[20] Samaneh Moghaddam and Martin Ester. On the design of lda models for aspect-based opinion mining. In *CIKM '12*, pages 803–812, 2012.

[21] Michael Paul and Roxana Girju. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *EMNLP '09*.

[22] Keith Stevens, Philip Kegelmeyer, David Andrzejewski, and David Buttler. Exploring topic coherence over many models and many topics. In *EMNLP-CoNLL '12*, pages 952–961, 2012.

[23] Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models. In *WWW '08*, pages 111–120, 2008.

[24] Yao Wu and Martin Ester. Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *WSDM '15*.

[25] Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. Phrase dependency parsing for opinion mining. In *EMNLP '09*.

[26] ChengXiang Zhai, Atulya Velivelli, and Bei Yu. A cross-collection mixture model for comparative text mining. In *KDD '04*.

[27] Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *EMNLP'10*.