Leveraging Only the Category Name for Aspect Detection through Prompt-based Constrained Clustering

Yazheng Li[1], **Pengyun Wang**[3*], **Yasheng Wang**[3], **Yong Dai**[1], **Yadao Wang**[3], **Lujia Pan**[3], **Zenglin Xu**[2,4*]
[1]University of Electronic Science and Technology of China, Chengdu, China
[2] Harbin Institute of Technology, Shenzhen, China
[3]Huawei Noah's Ark Lab [4]Peng Cheng Lab, Shenzhen, China
[1]uestc_liyazheng@163.com [2]zenglin@gmail.com
[3]{wangpengyun, wangyasheng, wangyadao, panlujia}@huawei.com

Abstract

Aspect category detection (ACD) aims to automatically identify user-concerned aspects from online reviews, which is of great value for evaluating the fine-grained performance of a product. The most recent solutions tackle this problem via weakly supervised methods, achieving remarkable improvement over unsupervised methods. However, a closer look at these methods reveals that the required human efforts are nontrivial and can sometimes be hard to obtain. In this study, we explore the possibility of minimizing human guidance while improving detection performance, with a deep clustering method that relies merely on the category name of each aspect and a pretrained language model (LM). The LM, combined with prompt techniques, is employed as a knowledge base to automatically generate constraints for clustering, as well as to provide a representation space to perform the clustering. Our method (1) extracts extensive keywords to expand our understanding of each aspect, (2) automatically generates instance-level and concept-level constraints for clustering, and (3) trains the clustering model with the above constraints. We demonstrate the capability of the proposed framework through extensive experiments on four benchmark datasets across nine domains. Our model not only performs noticeably better than existing unsupervised approaches but also considerably surpasses weakly supervised methods that require more human efforts. Our code is available at: https://github.com/liyazheng/PCCT.

1 Introduction

With the rapid development of online shopping platforms, the number of product reviews increases exponentially. Aspect-based sentiment analysis (ABSA) can analyze the attitudes and preferences of customers towards the detailed aspects of goods,

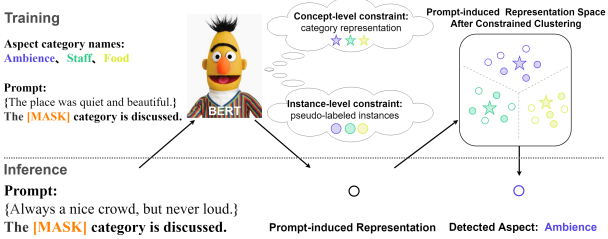


Figure 1: An overview of our Prompt-based Constrained Clustering method PCCT. Merely aspect names and a language model are required to category review sentences into predefined aspects.

which is key to promoting the sale for dealers. Aspect category detection (ACD), which categories review segments (i.e., a part of the reviews) into the corresponding predefined aspects (e.g., Ambience, Staff, and Food), is an essential step for ABSA.

Since labeled data is labor-intensive and time-consuming, unsupervised and weak supervised ACD have become the focus of the research community (García-Pablos et al., 2018; Karamanolakis et al., 2019; Huang et al., 2020; Shi et al., 2021). Latent Dirichlet Allocation (LDA)-based topic models (Brody and Elhadad, 2010; Chen et al., 2014; Özyurt and Akcayol, 2021) exploit the co-occurrence statistic of words to generate aspect keywords. Clustering-based methods (Chen et al., 2016; Xiong and Ji, 2016; Zhao et al., 2014) employ traditional clustering algorithms (e.g., k-means) to group words in review segments and design task-specific constraints for better performance. These unsupervised methods typically suffer from poor aspect detection performance, propelling researchers to explore weakly supervised methods. JASen (Huang et al., 2020) uses expert-designed keywords to learn latent representations for the user's concerned aspects and then uses neural models to distill the word-level discriminative knowledge. SSCL (Shi et al., 2021) adopts a neural network to extract high-quality aspect keywords, which are categorized into their target aspects by

* Co-corresponding authors.

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 1352 - 1364

mapping rules designed by domain experts.

To reduce the reliance on expert knowledge while improving the detection performance, we propose the Prompt-based Constrained ClusTering method (aliased as PCCT) for aspect category detection, merely based on the category name of each aspect (i.e., aspect names), as Figure 1 illustrates. The constraints in constrained clustering (Basu et al., 2008) refer to clustering with prior information, such as pairwise constraints, cluster size constraints, instance difficulty constraints (Zhang et al., 2020), which can be exploited to benefit the clustering process. Inspired by Meng et al. (2020); García-Pablos et al. (2018),we adopt aspect names as prior information to generate informative keywords for each category, which are converted into category-specific constraints to guide clustering.

More precisely, we elaborate on the following designings as shown in Figure 2. (1) We extend the single aspect name into a set of keywords by exploiting the linguistic and world knowledge contained in the pre-trained model. The aspect keywords act as category-centered semantics to guide the clustering process. (2) We explore a novel way to embed category-centered semantics into instance- and concept-level constraints. The instance-level constraint refers to a small set of pseudo-labeled instances that can influence the clustering process with discriminative category information. The concept-level constraint is built by generating cluster centroids corresponding to predefined aspect categories. (3) We train the constrained clustering model and perform predictions by computing the similarity between query instances and cluster centroids. Importantly, we propose to perform the knowledge extraction and clustering in a prompt-induced space, which is verified effective for narrowing the gap between LM's task-agnostic pre-training and task-specific fine-tuning (Petroni et al., 2019; Schick and Schütze, 2021; Liu et al., 2021a). We demonstrate the effectiveness of PCCT by performing experiments on nine benchmark datasets and show that our model can achieve significantly better aspect detection performance than both unsupervised and weakly supervised state-of-the-art methods. Our contributions can be summarized as follows:

- We propose a deep-constrained clustering method, PCCT, for aspect category detection. PCCT does not need any human effort but only the category name of each aspect.
- We explore a novel way to encode prior information contained in aspect name into constraints for clustering. To achieve the maximum capacity of LMs for the ACD task, we propose to conduct the generation of constraints and clustering over a prompt-induced space.
- Extensive experiments are carried out on four benchmark datasets and show that PCCT surpasses both unsupervised and weakly supervised state-of-the-art methods by a large margin.

2 Related Works

2.1 Aspect Category Detection

Unsupervised aspect category detection methods are based primarily on Latent Dirichlet Allocation (LDA) topic models (Brody and Elhadad, 2010; Chen et al., 2014; Pablos et al., 2018; Zheng et al., 2020; Özyurt and Akcayol, 2021) or traditional clustering algorithms (Chen et al., 2016; Xiong and Ji, 2016; Zhao et al., 2014). They typically suffer from poor aspect detection performance.

Several weakly-supervised methods have been proposed recently. Typical forms of weak supervision include hand-crafted mapping rules (He et al., 2017; Luo et al., 2019; Shi et al., 2021; Chebolu et al., 2022), and a few seed words per class (Angelidis and Lapata, 2018a; Huang et al., 2020; Karamanolakis et al., 2019). The neural topic model aspect-based autoencoder (ABAE) (He et al., 2017) and its variants (Luo et al., 2019; Shi et al., 2021) learns substantially high quality aspect-related words by capturing word cooccurrence patterns. However, manual mapping is needed when categorizing model-inferred words into aspects. To avoid manual mapping, MATE (Angelidis and Lapata, 2018a) takes a small set of seed words to initialize the autoencoder. JASen (Huang et al., 2020) utilizes seed words to generate aspect representations and enables a CNN model to align reviews to aspects. Although much weaker than a fully annotated corpus, the above forms of supervision signal still require non-trivial, corpus-specific knowledge from experts.

Our method, in comparison, relies just on the label name of each aspect category and has very little supervision.

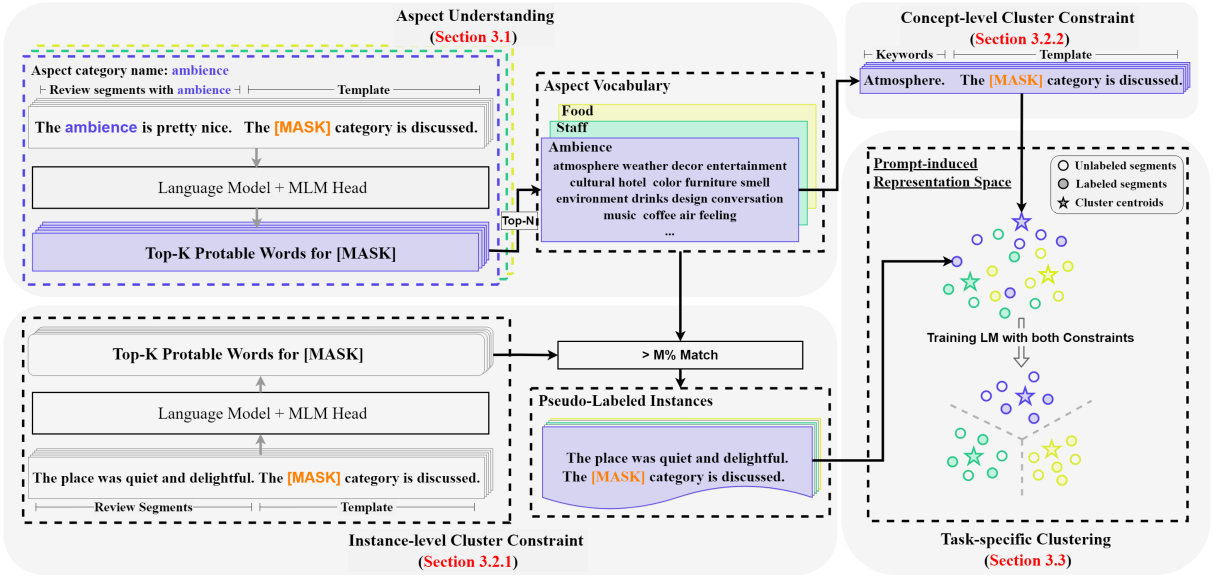


Figure 2: An illustration of PCCT (better viewed in color). We first utilize the review segments with an explicit category name to acquire the aspect vocabularies. Then, on one hand we match each aspect vocabulary with top-k protable words at the [**MASK**] position of each review segment to build the instance-level constraint. On the another hand, we construct clustering centroids based on these vocabularies to build the concept-level constraint. At last, constrained clustering is implemented to obtain the final segment representation in the prompt-induced space.

2.2 Deep Constrained Clustering

Constrained Clustering (Basu et al., 2008) refers to clustering with external or side information from other sources. Early works construct standard must-link/cannot-link constraints for traditional algorithms such as k-means. Recently, works that combine clustering with deep representation have drawn much attention (Xie et al., 2016; Jiang et al., 2017; Zhang et al., 2017). In particular, the deep embedded clustering model (DEC) (Xie et al., 2016) iteratively minimizes the within-cluster KL-divergence over the representation space of an autoencoder. SSCL (Zhang et al., 2021) leverages a pretrained language model as the backbone and achieves promising improvements in short text clustering.

A new deep constrained clustering framework is outlined in Zhang et al. (2020), which shows that deep clustering can handle not only basic pairwise constraints, but more complex constraints generated from new types of prior information, such as high-level domain knowledge. Following this line, our proposed method explores a general method to encode aspect name information into clustering constraints.

2.3 Prompt-based Learning

A prompt is usually a task description sentence with an unfilled token and has shown huge potential to mine knowledge from PLMs with few or even zero labeled data (Petroni et al., 2019; Brown et al., 2020; Liu et al., 2021b). Prompt-based learning has been applied for many natural language tasks, such as question answering (Lewis et al., 2019), common sense reasoning (Yang et al., 2020), text classification (Seoh et al., 2021; Schick and Schütze, 2021), and so on. A recent work (Liu et al., 2021a) casts the ACD task into a generation task, using natural language prompts to query the BART (Lewis et al., 2020) to generate task-specific tokens corresponding the output.

In contrast to previous works, we not only adopt prompts to induce knowledge from the LM, but also use prompts to map review segments into a task-specific representation space for clustering. To our knowledge, we are the first to perform clustering in a prompt-induced embedding space.

3 Methodology

Figure 2 shows our clustering framework for aspect detection, which utilizes only the label names of aspect categories and a pre-trained LM $\mathcal{H}$. The key to high performance is to effectively induce aspect-detection-specific knowledge from $\mathcal{H}$. Our framework contains three steps: (1) an expansive understanding of the aspect categories; (2) the generation of two task-specific cluster constraints based on aspect understanding; (3) a constrained clustering

training method over the prompt-induced representation space. We use BERT for LM in this study, but our framework can be compatible with more powerful pre-trained LMs.

3.1 Aspect Understanding

Given the label name of each aspect, our aim is to expand the understanding of the aspects in the form of a set of keywords (that is, aspect vocabulary) that can comprehensively represent their semantics. A prompt-based method is proposed for the purpose.

Specificially, after locating a label name in the corpus, we concatenate its context with a template $\mathcal{T}$, which describes our task. For example, assuming we have a review segment $\mathcal{X}$ = "The ambience is pretty nice" containing the exact label name "ambience", we wrap it into

$$\mathcal{X}_{\text{p}} = \text{The ambience is pretty nice. } \mathcal{T}.$$

We choose

$$\mathcal{T} = \textit{The } [\textbf{MASK}] \textit{ category is discussed}$$

which has shown to be the best performance template in (Liu et al., 2021a). The LM $\mathcal{H}$ with a mask language model head (**H-MLM**) gives the probability that each word v in the BERT vocabulary is filled in the token [**MASK**]. Aspect vocabulary V_c for category c is generated as following:

$$W_{ic} = \textit{TOP-K}(\textbf{H-MLM}(\mathcal{X}_{\text{p}})); \quad (1)$$

$$W_c = \bigcup_{i=1}^{|\mathcal{D}_c|} W_{ic}; \quad (2)$$

$$V_c = \textit{TOP-N}(\textit{SORT}(W_c)). \quad (3)$$

$\mathcal{D}_c$ is a set of review segments containing the exact label name of category c. We sort words in W_c by the number of times they occurred in every W_{ic}. The top-N words in sorted W_c are chosen for V_c.

3.2 Automatic Constraints Generation

It is difficult for unsupervised clustering to group review segments into predefined aspects directly (Xiong and Ji, 2016). Here, we encode the aspect keywords into instance- and concept-level constraint to guide clustering for the ACD task.

3.2.1 Instance-level Clustering Constraint

Based on keywords in aspect vocabularies, we search for reliable instances for each aspect in the training corpus. The generated pseudo-labels indicate to the model the desired cluster assignment for the reliable instances.

We propose to examine each review segment by its predictions on the [**MASK**] token in the template and attempt to match them with each of the aspect vocabularies. A satisfactory match indicates a reliable training example, with its label assigned to the matched category. Specifically, after reforming the review segment $\mathcal{X}$ into $\mathcal{X}_{\text{p}}$, we obtain its replacements W from Eq.(1). We assign $\mathcal{X}$ to aspect c if W covers more than $M\%$ of V_c. By examining every example in the unlabeled trainset as above, we will obtain a set of pseudo-labeled data S. Above generation process is also shown in Algorithm 2 (Appendix B).

3.2.2 Concept-level Clustering Constraint

We build a general representation for each aspect to function as cluster centroids. Once centroids are identified, clustering can be restricted to group review segments into user-interested aspects.

Aspect keywords from V_c can be naturally used to build the cluster centroid $\boldsymbol{\mu}_c$. In order to embed keywords into a prompt-induced representation space, we wrap the keywords into a template $\mathcal{T}$ and feed it into a LM as before. Assuming that we have a keyword **w=** "atmosphere", the sentence sent into LM should be

$$\mathcal{W}_{\text{p}} = \textbf{atmosphere. } \mathcal{T}.$$

We define **w**'s representation $\boldsymbol{k}_i$ as the hidden output vector at [**MASK**] position. In particular, the representation of the exact label name, that is, **w=** "ambience", is denoted as $\boldsymbol{k}_l$. We adopt an attention mechanism to aggregate keywords' representations into $\boldsymbol{\mu}_c$:

$$\begin{aligned} \boldsymbol{\mu}_c &= \sum_{i=1}^{|V_c|} a_i \mathbf{k}_i, \\ a_i &= \frac{\exp\left(\mathbf{k}_i \cdot \mathbf{k}_l\right)}{\sum_{j=1}^{N} \exp\left(\mathbf{k}_j \cdot \mathbf{k}_l\right)}, \end{aligned} \quad (4)$$

where a_i measures the similarity between each keyword and its corresponding label name.

3.3 Constrained Clustering

For constrained clustering, we first take the LM to transform review segments into a prompt-induced representation space, and then train the LM for better clustering results with the generated centroids $\{\boldsymbol{\mu}_c\}_{c=1}^{C}$ and pseudo-labeled instances $\mathcal{S}$.

In the first step, we wrap review segment x into $\mathcal{X}_{\text{p}}$ with the same template $\mathcal{T}$ used for centroids and take the LM output embedding at the position of [**MASK**] as x's representation $\boldsymbol{h}$. We find that it is empirically important to use exactly the same $\mathcal{T}$ for keywords and review segments. Also noted that we further pre-train the origin LM on the abundant unlabeled trainset when obtaining $\boldsymbol{h}$ and $\boldsymbol{\mu}$ before clustering, due to the origin LM suffers from domain-awareness challenge – no review data is pre-trained on it (Xu et al., 2019).

Concept Constrained Clustering Once $\boldsymbol{\mu}$ and $\boldsymbol{h}$ are identified, we use p_{ic} to denote the probability of assigning the review segment x_i to the c-th cluster, which is calculated as

$$p_{ic} = P(y_c|\boldsymbol{h}_i) = \frac{\exp\left(\boldsymbol{\mu}_c^T \boldsymbol{h}_i\right)}{\sum_{c'=1}^{C} \exp\left(\boldsymbol{\mu}_{c'}^T \boldsymbol{h}_i\right)}. \quad (5)$$

For unsupervised clustering, we iteratively refine the LM by a target cluster assignment probability proposed by (Xie et al., 2016). Specifically, the target assignment q_{ic} is defined as:

$$q_{ic} = \frac{p_{ic}^2 / f_c}{\sum_{c'=1}^{C} \left(p_{ic'}^2 / f_{c'}\right)}; f_c = \sum_i p_{ic}, \quad (6)$$

where q_{ic} first sharpens p_{ic} and then normalizes it by frequency. By doing so, the LM is able to learn from current high-confidence assignments while demoting low-confidence ones. We push the current assignment probability towards the target probability with a KL divergence loss between them:

$$\mathcal{L}_{Cluster} = \sum_{i=1}^{D} \sum_{c=1}^{C} q_{ic} \log \frac{q_{ic}}{p_{ic}}, \quad (7)$$

where D is the size of the unlabeled corpus and C is the number of categories. q_{ic} is updated by Eq.(6) every 50 batches and is terminated when the hard label selected from q_{ic} remain unchanged after update.

Instance Constrained Clustering The above clustering objective focuses on high-level semantic concepts of each aspect. We further adopt $\mathcal{S}$ to aid clustering performance. For each $(x, c) \in \mathcal{S}$, we compute the cross-entropy loss over LM's current assignment $P\left(y_c|\boldsymbol{h}_x\right)$ from Eq.(5):

$$\mathcal{L}_{CE} = -\sum_{(x,c)\in\mathcal{S}} \log P\left(y_c|\boldsymbol{h}_x\right). \quad (8)$$

Overall objective Our final training objective is

$$L = (1 - \alpha_t) \cdot L_{Cluster} + \alpha_t \cdot L_{\text{CE}}. \quad (9)$$

α_t balances between the clustering loss and the cross-entropy loss. In preliminary experiments, we found a schedule to increase α_t linearly from 0 to 1 during training consistently gives good results, so we use it in all our experiments.

Inference The resulting LM model $\mathcal{H}$ and cluster centroids $\{\boldsymbol{\mu}_c\}_{c=1}^{C}$ can be used to classify any unseen reviews. Given a review segment x, we obtain its representation $\boldsymbol{h}$ with $\mathcal{H}$, and predict the output category $\hat{y}$ for x:

$$\hat{y} = \arg\max_c \left(\boldsymbol{\mu}_c^T \boldsymbol{h}\right). \quad (10)$$

Algorithm 1 summarizes the proposed framework.

Algorithm 1: PCCT Framework

Input: Unlabeled review corpus $\mathcal{D}$; aspect label names $\mathcal{L}$; a pre-trained LM $\mathcal{H}$; number of aspect categories C.
Output: Cluster centroids $\{\boldsymbol{\mu}_c\}_{c=1}^{C}$; cluster assignments $\{\hat{y}_i\}_{i=1}^{n}$;
// Build constraints:
Aspect keywords $\{V_c\}_{c=1}^{C} \leftarrow$ Section 3.1;
Pseudo-labeled Instances $\mathcal{S} \leftarrow$ Section 3.2;
Cluster centroids $\{\boldsymbol{\mu}_c\}_{c=1}^{C} \leftarrow$ Section 3.2;
// Cluster with constraints:
$B \leftarrow$ Total number of batches ;
repeat
  for $i = 0$ *to* $B - 1$ **do**
    $P \leftarrow Eq.(5)$;
    if i *mod* $50 = 0$ **then**
      $Q \leftarrow$ Eq.(6);
    end
    Update $\mathcal{H}$ with Eq.(9);
  end
until *Hard labels from* Q *unchanged*;
// Inference:
for $x_i \in \mathcal{D}$ **do**
  $\boldsymbol{h} \leftarrow$ Task-specific embedding of x_i;
  $\hat{y}_i \leftarrow \arg\max_c(\boldsymbol{\mu}_c^T \boldsymbol{h})$;
end

4 Experiments

4.1 Datasets

We train and evaluate our methods on nine datasets: Semeval Reviews in two different domains (**Restaurant** and **Laptop**), **CitySearch**

Aspect	Prompt-based keywords	Non-prompt keywords
staff	job relevant women work **individual** sports senior security customer vip management staff technical **professional** class client education **appropriate** language personality current services	staff crew team personnel employees floor head management line inside workers set brand site department guard guests club ward men unit canteen ladies faculty
ambience	**cultural** hotel opposite color furniture culture shopping dance architecture design **visual lifestyle** previous house original	music atmosphere weather decor entertainment coffee air space work view art money reception furniture style lighting fun party beer smell environment drinks design conversation feeling
food	food fish salad cheese foods **nutritional** **edible** meal chicken eat flavor dishes fruit lunch vegetable snack beverage soup consumption **asian** eats flavour	food eat fare foods meat chow diet bread land eating seafood dish course breakfast product salad chicken burger pasta total meals store fruit back

Table 1: The generated aspect keywords of *CitySearch* dataset with prompt-based and non-prompt schemes.

restaurant reviews and OPOSum product reviews in six different domains (**Bags**, **Bluetooth**, **Boots**, **Keyboards**, **TV** and **Vacuums**). Reviews are segmented into elementary discourse units (EDUs) for training and evaluating following previous work (Angelidis and Lapata, 2018a; Shi et al., 2021). For OPOSum datasets, we only consider the typical aspects since the category name "general" cannot provide useful prior information. The detailed information on datasets and experiment settings are shown in Appendix A and D.

4.2 Competetive Methods

We compare our method with both unsupervised and weakly-supervised aspect detection methods.

Unsupervised Methods: (1) **SERBM** (Wang et al., 2015) extracts review aspects in an unsupervised manner based on Boltzmann machines. (2) **CAt** (Tulkens and van Cranenburgh, 2020) adopts a contrastive attention mechanism based on Radial Basis Function kernels for aspect assignment.

Weakly Supervised Methods with Hand-crafted Aspect Mapping: (3) **ABAE** (He et al., 2017) uses the auto-encoder framework to extract aspect keywords from review segments. These keywords are then mapped into predefined aspect categories by experts. (4) **SSCL** (Shi et al., 2021) improves the quality of ABAE-extracted aspect keywords with contrastive learning and takes BERT as a student model in knowledge distillation to boost performance.

Weakly Supervised Methods with Aspect Seed Words: (5) **W2VLDA** (García-Pablos et al., 2018) is a topic model-based approach, which automatically pairs discovered topics with predefined aspect seed words. (6) **TS** (Karamanolakis et al., 2019) is a student-teacher co-training framework, where the teacher is a bag-of-words classifier based on seed words and the student is a BERT classifier. (7) **UCE** (Nguyen et al., 2021) refines an encoder with uncertainty-aware loss, which approximates the similarity between segments and aspects with the ground truth similarity generated from seed words. (8) **JASen** (Huang et al., 2020) learns aspect embeddings using seed words and then trains a CNN as a classifier to learn discriminative information at the word level.

4.3 Aspect Keywords Visualization

Table 1 shows the learned aspect keywords of *CitySearch* Dataset from Section 3.1. These keywords expand the semantic meaning of the aspect name. Take the "food" aspect as an example, "fish, salad, fruit..." cover a variety of different foods, and "nutritional, edible, Asian..." describes the properties of the food. To further illustrate the advantages of our prompt-based aspect understanding, we compare our strategy with an alternative solution without prompts from (Meng et al., 2020). Instead of taking the probable words of the [**MASK**] token in prompt, this alternative way directly takes the probable words at each appearance of the aspect name. We can observe from Table 1 that the prompt-based scheme generates more meaningful adjectives such as "professional" and "cultural" than the non-prompt scheme. These adjectives are helpful for PCCT to identify implicit aspect expressions such as "always a nice crowd, but never loud." where no explicit expression is related to its target aspect "ambience". We can also observe that PCCT has the ability of word sense disambiguation. Given the category name, we can obtain a set of

Method	Restaurant		Laptop		CitySearch	
	Accuracy	macro-F1	Accuracy	macro-F1	Accuracy	macro-F1
Unsupervised Methods						
SERBM (2015)	-	-	-	-	83.8	74.5
CAt (2020)	66.3	46.2	58.0	58.6	83.6	82.5
Weak-supervised Methods with Aspect Mapping						
ABAE (2017)	67.3	45.3	59.8	56.2	85.7	77.5
SSCL (2021)	-	-	-	-	89.7	87.0
Weak-supervised Methods with Aspect Seed Words						
W2VLDA (2018)	70.8	51.4	64.9	63.4	70.7	72.0
JASen (2020)	83.8	66.3	71.0	69.7	87.3	86.2
UCE (2021)	77.5	58.8	71.3	71.3	-	-
PCCT	**85.3**	**79.2**	**74.3**	**73.4**	**90.6**	**89.8**

Table 2: Performance Comparison on *Semeval* and *CitySearch* datasets. We choose to report macro-F1 due to the imbalance issue (Appendix C). The original papers report the weighted macro averages as well as the F1 scores for each category. We averaged these F1 scores for each category to obtain the macro F1 score presented in our paper.

Model	Bags	B/T	Boots	Kbs	TVs	Vcs	AVGS
	Weak-supervised Methods						
TS (2019)	41.2	42.0	31.6	26.9	40.4	40.5	37.1
SSCL (2021)†	50.6	52.7	44.0	38.7	48.7	37.2	45.3
UCE (2021)	49.4	48.4	45.7	48.0	47.6	41.2	46.7
PCCT	**69.8**	**66.8**	**59.6**	**55.4**	**72.8**	**48.9**	**60.2**

Table 3: Performance comparision (Weighted-F1 score) on typical aspects of *OPOSum* Dataset. Baseline results without the mark † are taken from (Nguyen et al., 2021). The results marked by † are reproduced by running their code, since these experiments are not conducted in the original articles (details are shown in Appendix D).

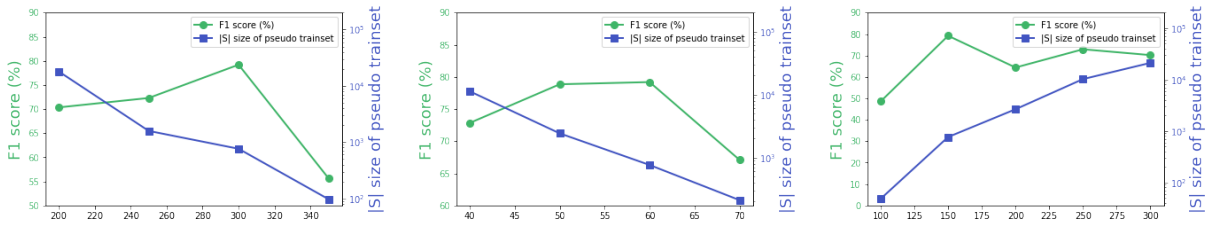
Domain	Keywords
BOOTS	comfort touch pain stress balance bedroom attachment
KEYBOARD	comfort curve wrists minutes vertical split wrist feel pure softer action zoom hardcore

Table 4: Different aspect keywords for category "Comfort".

keywords for each domain according to the procedures described in Section 3.1. As shown in table 4, given the category name "Comfort", we obtain a list of key words for the domains of BOOTS and KEYBOARD, respectively. We can see that "comfort" has different meanings for these two domains of BOOTS and KEYBOARD.

4.4 Performance Comparison

The comparison results are shown in Table 2 and Table 3 respectively. Note that we report Accuracy and Macro-F1 for *Restaurant*, *Laptop* and *CitySearch* datasets, while Weighted-F1 is reported for the other datasets. This is due to the issue of consistency with the existing published results. In general, weakly supervised methods obtain much better performance than unsupervised ones. And among all unsupervised and weakly-supervised methods, PCCT achieves the best performance across all nine datasets, despite of being an almost unsupervised method. For the first three datasets, it outperforms SSCL, JASen, and UCE by **0.9%**, **2.7%** and **5.4%** in the absolute value of precision, on average and respectively. If we examine the Macro-F1 score, the overall improvement in performance is about the same size as above, except for the Restaurant dataset. PCCT increases the Macro-F1 score from the second best of **66.3%** to **79.2%** for the *Restaurant* dataset. Such an improvement is potentially due to the effectiveness of PCCT at addressing the issue of aspect imbalance in the data, as shown in a more detailed analysis in Appendix C. For the other six datasets, PCCT demonstrates an even more remarkable improvement. On average, PCCT



(a) Aspect Vocab Size **N**. (b) Match Threshold **M** (%). (c) Number of Protable Words **K**.

Figure 3: (On *SemEval Restaurant* Dataset.) A trade-off between the quality and quantity of generated pseudo-labeled instances. Both detection performance (green) and the total number of pseudo-labeled instances (blue) are reported. The x-axises indicates three parameters have impact on the pseudo labeled instances, respectively.

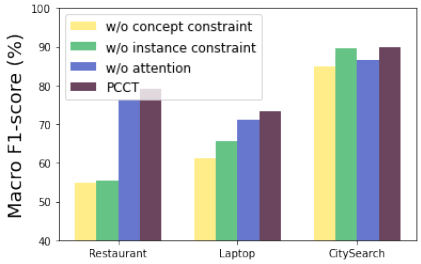


Figure 4: (On *SemEval* and *CitySearch* datasets). Ablation study of our proposed PCCT model.

increases the Weighted-F1 score by **22.9%**, **14.9%** and **13.5%**, over TS, SSCL and UCE, respectively. A closer look at the datasets reveals that, the reviews (e.g., *Despite the slender profile, it stays put on my desk, no slipping or sliding or rocking.*) in *OPOSum* are more detailed in describing how users use the product (e.g., *Keyboard*) and how they feel about it, and it is difficult to determine the aspect category (e.g., *Feel Comfortable*) by a specific word. Unlike baseline methods, PCCT effectively extracts the semantic information at the sentence level using prompts and clusters prompt-induced sentence embeddings instead of word embeddings in the representation space, thus allowing better handling of the implicit aspect representations described above.

4.5 Ablation Study

To validate the effectiveness of our constrained clustering, we train our model (1) without the instance-level constraint. (2) without concept-level constraint. (3) without attention. For the first variant, only the clustering loss is involved during training. In the second variant, we design a variant of our model that adopts a linear clustering head of size 768 × C with C indicating the number of clusters following the setting of (Zhang et al., 2021). This method only restricts the number of clusters, and the cluster centroids are randomly initialized by the cluster head without any prior information. Lastly, for the third variant, a simple method of mean pooling is used to aggregate aspect keywords into clustering centroids.

As shown in Figure 4, PCCT performs better than all alternative variants. Although both concept- and instance-level constraints are important, the former consistently exhibits a more significant impact on model performance than the latter. On average, the variant without the former decreases the performance by **13.83%**, while the variant without the latter diminish the performance by **10.63%**. This observation might be indicative of the importance of the choice of centroids in deep clustering (Zhang et al., 2020). Furthermore, the effect of the constraints varies from dataset to dataset. For simple datasets such as *CitySearch*, the improvement is relatively limited (∼**2%**). For harder datasets, such as *Restaurant* and *Laptop*, the increase in performance can be quite remarkable (on average ∼**15%**). Specifically, the *CitySearch* dataset involves only 3 aspect categories, *Restaurant* involves 5, and *Laptop* involves 8. Meanwhile, the *Restaurant* dataset suffers from a severe category imbalance (Appendix C). This illustrates that the constraints we designed can better guide the clustering process when the task is more difficult. Lastly, the adopted attention mechanism for constraint generation is effective as well, as it consistently performs better than the method of mean-pooling aggregation.

4.6 Parameter Analysis

We analyze three parameters: aspect vocab size N, matching threshold M, and number of protable words K when generating pseudo-labeled instances. They all have a direct influence on the quality and quantity of our pseudo labels (Section 3.2.1). It should be expected that, given fixed K and N (or

M) (defined in Section 3.1), increasing M (or N) improves the credibility, but decreases the number of pseudo-labeled instances. On the contrary, given a fixed M and N, increasing K will improve the quality, but decrease the number of pseudolabeled instances. Experimental results in Figure 3 confirm this trade-off and show that the performance of our model increases with the level of credibility before dropping due to the issue of sparse pseudo trainset.

5 Conclusion

In this paper, we propose a prompt-based constrained clustering method for aspect category detection. First, we automatically extract a comprehensive set of keywords for each aspect. Based on the keywords, we construct instance- and concept-level clustering constraints. Finally, a joint training objective is proposed for optimal aspect detection performance. Experiments show that our method not only learns high-quality aspect keywords, but also significantly outperforms both unsupervised and weakly supervised methods. In the future, we plan to extend our framework to jointly perform aspect and sentiment detection for ABSA.

6 Limitations

Here, we discuss the limitations of PCCT. First, only the category name is insufficient for accurate categorisation in certain complex situations. For example, some review texts describe the "Ambience" of a restaurant by describing the specific scene or people inside the restaurant : “Aside from the bearded, courdoroy blazer professor type with the nyu student he’s sleeping with that week, you will also see a strange mix of hipsters, frat boys and Will Smith types in this restaurant.” Incorporating domain knowledge from external knowledge bases into PTM is a possible direction to explore. Second, our method is based on a simple template and we directly use the optimal template validated in (Liu et al., 2021a). However, our preliminary experiments show that different templates with the same semantic might lead to poorer performance and that the best templates for certain review domain (e.g., Laptop or Restaurant) vary. We also note that, we filtered out the "general" aspect for the OPOSum dataset. This is because PCCT extends the names of aspects to obtain category-centered semantics, and the word "general" does not represent its heterogeneous content well. In terms of training efficiency, PCCT still has to update all parameters of the LM and requires at least two 1080ti GPUs per training session, which lasts about 60 minutes. Finally, in the experimental setup, we verified the effectiveness of PCCT only for the aspect category detection task, and, in fact, PCCT can be extended to aspect sentiment analysis or to more general text classification tasks.

Acknowledgments

Zenglin Xu and Yazheng Li was partially supported by National Key Research and Development Program of China (No. 2018AAA0100204), a key program of fundamental research from Shenzhen Science and Technology Innovation Commission (No. JCYJ20200109113403826), and the Major Key Project of PCL (No. PCL2021A06). We appreciate the helpful and informative feedback provided by the anonymous reviewers.

References

Stefanos Angelidis and Mirella Lapata. 2018a. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium. Association for Computational Linguistics.

Stefanos Angelidis and Mirella Lapata. 2018b. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium. Association for Computational Linguistics.

Sugato Basu, Ian Davidson, and Kiri Wagstaff. 2008. *Constrained clustering: Advances in algorithms, theory, and applications*. CRC Press.

Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 804–812. The Association for Computational Linguistics.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei.

2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Siva Uday Sampreeth Chebolu, Paolo Rosso, Sudipta Kar, and Thamar Solorio. 2022. Survey on aspect category detection. *ACM Comput. Surv.* Just Accepted.

Lu Chen, Justin Martineau, Doreen Cheng, and Amit P. Sheth. 2016. Clustering for simultaneous extraction of aspects and features from reviews. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 789–799. The Association for Computational Linguistics.

Zhiyuan Chen, Arjun Mukherjee, and Bing Liu. 2014. Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 1: Long Papers*, pages 347–358. The Association for Computer Linguistics.

Gayatree Ganu, Noemie Elhadad, and Amélie Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In *12th International Workshop on the Web and Databases, WebDB 2009, Providence, Rhode Island, USA, June 28, 2009*.

Aitor García-Pablos, Montse Cuadros, and German Rigau. 2018. W2vlda: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91:127–137.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 388–397, Vancouver, Canada. Association for Computational Linguistics.

Jiaxin Huang, Yu Meng, Fang Guo, Heng Ji, and Jiawei Han. 2020. Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6989–6999, Online. Association for Computational Linguistics.

Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou. 2017. Variational deep embedding: An unsupervised and generative approach to clustering. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1965–1972. ijcai.org.

Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. 2019. Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4610–4620. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.

Patrick Lewis, Ludovic Denoyer, and Sebastian Riedel. 2019. Unsupervised question answering by cloze translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4896–4910, Florence, Italy. Association for Computational Linguistics.

Jian Liu, Zhiyang Teng, Leyang Cui, Hanmeng Liu, and Yue Zhang. 2021a. Solving aspect category sentiment analysis as a text generation task. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 4406–4416. Association for Computational Linguistics.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021b. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *CoRR*, abs/2107.13586.

Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019. Unsupervised neural aspect extraction with sememes. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5123–5129. ijcai.org.

Yu Meng, Yunyi Zhang, Jiaxin Huang, Chenyan Xiong, Heng Ji, Chao Zhang, and Jiawei Han. 2020. Text classification using label names only: A language model self-training approach. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9006–9017, Online. Association for Computational Linguistics.

Thi-Nhung Nguyen, Kiem-Hieu Nguyen, Young-In Song, and Tuan-Dung Cao. 2021. An uncertainty-aware encoder for aspect detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*, pages 797–806. Association for Computational Linguistics.

Baris Özyurt and Muhammet Ali Akcayol. 2021. A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Syst. Appl.*, 168:114231.

Aitor García Pablos, Montse Cuadros, and German Rigau. 2018. W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Syst. Appl.*, 91:127–137.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 486–495. The Association for Computer Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 255–269. Association for Computational Linguistics.

Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. Open aspect target sentiment classification with natural language prompts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 6311–6322. Association for Computational Linguistics.

Tian Shi, Liuqing Li, Ping Wang, and Chandan K. Reddy. 2021. A simple and effective self-supervised contrastive learning framework for aspect detection. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 13815–13824. AAAI Press.

Stéphan Tulkens and Andreas van Cranenburgh. 2020. Embarrassingly simple unsupervised aspect extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3182–3187, Online. Association for Computational Linguistics.

Linlin Wang, Kang Liu, Zhu Cao, Jun Zhao, and Gerard de Melo. 2015. Sentiment-aspect extraction based on restricted boltzmann machines. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 616–625, Beijing, China. Association for Computational Linguistics.

Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 478–487. JMLR.org.

Shufeng Xiong and Donghong Ji. 2016. Exploiting flexible-constrained k-means clustering with word embedding for aspect-phrase grouping. *Inf. Sci.*, 367-368:689–699.

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 2324–2335. Association for Computational Linguistics.

Jheng-Hong Yang, Sheng-Chieh Lin, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. 2020. Designing templates for eliciting commonsense knowledge from pretrained sequence-to-sequence models. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3449–3453, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Dejiao Zhang, Feng Nan, Xiaokai Wei, Shang-Wen Li, Henghui Zhu, Kathleen R. McKeown, Ramesh Nallapati, Andrew O. Arnold, and Bing Xiang. 2021. Supporting clustering with contrastive learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 5419–5430. Association for Computational Linguistics.

Dejiao Zhang, Yifan Sun, Brian Eriksson, and Laura Balzano. 2017. Deep unsupervised clustering using mixture of autoencoders. *CoRR*, abs/1712.07788.

Hongjing Zhang, Sugato Basu, and Ian Davidson. 2020. A framework for deep constrained clustering - algorithms and advances. In *Machine Learning and Knowledge Discovery in Databases*, pages 57–72, Cham. Springer International Publishing.

Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng, and Xiaoyan Zhu. 2014. Clustering aspect-related phrases by leveraging sentiment distribution consistency. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1614–1623. ACL.

Lu Zheng, Zhen He, and Shuguang He. 2020. A novel probabilistic graphic model to detect product defects from social media data. *Decis. Support Syst.*, 137:113369.

A Dataset Information

- **CitySearch Restaurant Reviews**. This data has been collected by (Ganu et al., 2009). Reviews are processed into 279,862 segments for training and 1490 for test. The *Citysearch* dataset has only train and test sets. Following (Shi et al., 2021), we use restaurant subsets of SemEval 2014 (Pontiki et al., 2014) and SemEval 2015 (Pontiki et al., 2015) as a development set that contains 2686 segments in total.

- **SemEval**. Following (Huang et al., 2020), we have 17,027 unlabeled reviews from Yelp Dataset Challenge and 14,683 unlabeled Amazon reviews in the laptop category as training corpus for Restaurant and Laptop, respectively. We take SemEval-2016 (Pontiki et al., 2016) and SemEval-2015 (Pontiki et al., 2015) as development sets and test sets.

- **OPOSUM**. This is a new dataset introduced in (Angelidis and Lapata, 2018b) contains Amazon reviews from six product domains: For evaluation, it randomly samples 600 reviews to be used for development (300) and testing (300) each domain. The "General" aspect is not considered in our settings.

The annotated aspects for all datasets are shown in Table 5.

Dataset	Aspects
Citysearch	Food, Ambience, Staff
SE-Rest	location, drinks, food, ambience, service
SE-Laptop	support, os, display, battery, company, mouse, software,keyboard
Bags	Compartments, Customer Service, Handles, Looks, Price, Quality, Protection, Size/Fit.
Bluetooth	Battery, Comfort, Connectivity, Durability, Ease of Use, Look, Price, Sound
Boots	Color, Comfort, Durability, Look, Materials, Price, Size, Weather, Resistance
Keyboards	Build Quality, Extra Function, Feel Comfort, Layout, Looks, Connectivity, Noise, Price
TVs	Apps/Interface, Connectivity, Image, Ease of Use, Size/Look, Sound, Price, Customer Service,
Vacuums	Accessories, Build Quality, Customer Service, Noise, Price, Ease of Use, Suction Power, Weight

Table 5: The annotated aspects for different datasets.

B Pseudo-labels generation.

Algorithm 2 shows the generation process of pseudo-labeled instances.

C Data statistics for SemEval datasets

Figure 5 shows the distribution of categories on the trainset of *SemEval-Restaurant* and the F1 values of our PCCT method and the JASen baseline method on each category. Since the trainset is unlabeled, we cannot directly perform statistics on its category distribution. Therefore, we use some predefined keywords to retrieve the training set and simulate the category distribution of the training set by the number of instances retrieved. We can observe that the *SemEval-Restaurant* dataset suffers from a significant category imbalance problem, which is not mitigated by the JASen model. JASen performs extremely poorly on the category with few examples, hence the low macro-F1 score. The PCCT model, on the other hand, has a more even performance across categories, with macro-F1 getting a significant boost. We speculate that this is due to the fact that our scheme does not directly rely on an unbalanced training set for training, but rather generates the same number of category keywords for each category to aid in training.

Algorithm 2: Generate hard-selective pseudo labels

Input: Unlabeled Review corpus $\mathcal{D}$;aspect vocabularies ; a pre-trained LM $\mathcal{H}$; match threshold $M\%$.
Output: Partial pseudo-labeled instances $\mathcal{S}$.

```
S ← {};
for x_i ∈ D do
    w_i ← Eq.(1)
    for j ∈ 1, ..., C do
        U_ij ← w_i ∩ V_j
        if |U_ij| / |V_j| > M% then
            c_i ← j;
            S ← S ∪ {(x_i, c_i)};
        end
    end
end
```

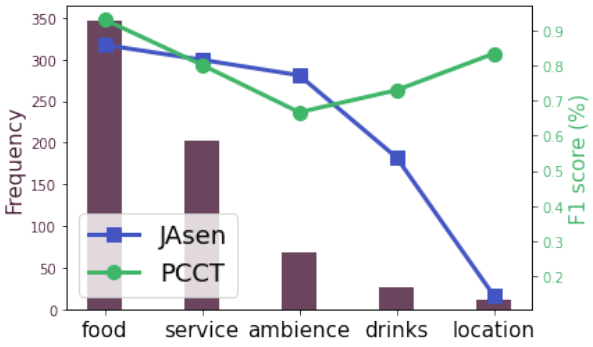


Figure 5: (On *SemEval Restaurant* dataset). Both detection performance of JASen and PCCT and the class distribution are reported. The x-axis indicates each aspect category.

D Experimental settings.

D.1 PCCT

We use Adam optimizer with a batch size of 64. We use the BERT-base-uncased as our backbone model and set the maximum input length to 512. When generating aspect keywords, we involve aspect vocabulary size N, the matching threshold M, and the number of predicted words for each [**MASK**] token K. We set $N = 100, M = 60, K = 50$ for the *CitySearch* dataset, $N = 300, M = 60, K = 150$ for *SemEval* datasets and $N = 150, M = 60, K = 100$ for the *OPOSUM* dataset. The model runs on 4 NVIDIA GeForce GTX 1080 Ti GPUs. All reported results are averaged on five runs with different seeds.

D.2 SSCL

The performance of SSCL is reported by running their code available at https://github.com/tshi04/AspDecSSCL. We use their default parameter setting on all datasets. We should mention that the manual mapping step has a great influence (around 20% on the F1 score) on its final performance. We report the best F1 score among results with 20 different mappings for each dataset.