Procesamiento del Lenguaje Natural, Revista nº 69, septiembre de 2022, pp. 117-127 recibido 31-03-2022 revisado 21-05-2022 aceptado 23-05-2022

A Semantic-Proximity Term-Weighting Scheme for Aspect Category Detection

Ponderación de Términos basada en Proximidad Semántica para la Detección de Categorías de Aspecto

Monserrat Vázquez-Hernández, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, México
{mvazquez, villasen, mmontesg}@inaoep.mx

Abstract: Aspect category detection is a subtask of aspect-level sentiment analysis, which aims at identifying the aspect categories present in an opinion. It is a difficult task because the category must be inferred from the terms of the opinion, and also because each opinion may include judgments for more than one aspect category. In recent years, the use of attention mechanisms has improved performance in different tasks, allowing the identification and prioritization of terms that mostly contribute to the classification. However, in multi-label problems, such as aspect category detection, different terms must be selected based on each category, which is a drawback for these models. Motivated by the same idea of identifying and highlighting the importance of terms, this paper proposes a weighing scheme that emphasizes terms in an opinion based on their *semantic proximity* to each aspect category. The proposed scheme has been evaluated on different SemEval datasets, demonstrating its effectiveness in this multi-label scenario. Moreover, it can be applied in scenarios with limited training data and can be combined with different classification models, including deep neural networks.
Keywords: semantic proximity, weighting of terms, aspect category detection.

Resumen: La detección de categorías de aspecto es una subtarea dentro del análisis de sentimientos a nivel de aspecto. Esta subtarea aborda la identificación de aquellas categorías de aspecto presentes en una opinión. Se trata de una tarea desafiante pues la categoría debe inferirse de los términos de la opinión, aunado a esto, una opinión puede incluir evaluaciones de más de una categoría de aspecto. En los últimos años, el uso de mecanismos de atención ha permitido mejorar los resultados en distintas tareas, éstos permiten identificar y priorizar los términos clave que contribuyen a la clasificación. Sin embargo, en problemas multi-etiqueta, como la detección de categorías de aspecto, se deben seleccionar diferentes términos dependiendo de cada categoría lo cual es un inconveniente para estos modelos. Motivados por esta misma idea de identificar y destacar la importancia de términos clave, en este trabajo se propone un esquema que permite enfatizar los términos de una opinión en función de su *proximidad semántica* a cada categoría de aspecto. El esquema propuesto se evaluó en distintos conjuntos de datos de SemEval demostrando su efectividad en este escenario multi-etiqueta. Además, es posible aplicarlo a pesar de contar con pocos datos de entrenamiento, y puede combinarse con distintos modelos de clasificación, incluyendo redes neuronales profundas.
Palabras clave: proximidad semántica, ponderación de términos, detección de categorías de aspecto.

ISSN 1135-5948. DOI 10.26342/2022-69-10

1 Introduction

Sentiment Analysis aims to identify emotions, attitudes or opinions in a subjective text about a product, service or topic of interest (Liu and Zhang, 2012). Different consumers may have different opinions about the same product or service. Through their opinions, consumers express their approval or rejection on particular aspects that they wish to highlight, which poses the challenge of grouping the opinions into different predefined aspect categories in order to identify relevant groups (López Ramos and Arco García, 2019). This challenge is tackled by the sub-task of Aspect Category Detection.

Aspect Category Detection (ACD) attempts to identify the general concepts to which each of the different aspects named in an opinion belong (Pontiki et al., 2016). For example, given the opinion: "the spaghetti was tasteless but the staff was nice", the aspects named are "spaghetti" and "staff", and the corresponding categories are "food" and "service". ACD is a multi-label problem since more than one aspect can be evaluated in an opinion and each one corresponds to a specific category.

Identifying the terms associated with the different categories is a difficult task because the category must be inferred from the context. One possibility for this is to observe different modifiers in an opinion. Through the nature of each category, it is possible to infer the associated terms. For example, in a restaurant, the term "tasteless" is used to describe the "food" but not the "staff". In this context, a common approach used to address this subtask is the use of lexicons (Mowlaei, Saniee Abadeh, and Keshavarz, 2020). Methods based on this approach perform category detection using sets of words to identify the corresponding categories. However, the construction of these lexicons is difficult, expensive, domain and language dependent.

Recently, deep learning approaches using attention mechanisms have been applied to address this task. These mechanisms examine the context of a sentence and identify and prioritize the most relevant terms for its classification. In same way, they are like lexicon-based approaches in that they emphasize the most relevant terms associated with a category, except that these approaches do so automatically without relying on external resources (Chaudhari et al., 2021). Unfortunately, these approaches have some drawbacks for this application. On the one hand, this is a multi-label problem, so the terms related to all different aspect categories mentioned in a single text must be jointly identified (Movahedi et al., 2019). On the other hand, like any deep learning approach, they require large training sets to achieve good results (Chaudhari et al., 2021), and for this subtask datasets are usually limited and also highly imbalanced.

Similar to previous works, in this paper we propose a new term weighting scheme whose aim is also to identify and prioritize terms associated with each aspect category. Based on a given set of category-oriented lexicons, which may have been manually or automatically defined, the proposed scheme weights each term in an opinion according to its semantic proximity to the different aspect categories. To do this, it considers pre-trained word embeddings; in a first step it computes a representative vector for each category, and then, in a second step, it measures the similarity of each term vector with respect to each category vector. Accordingly, the terms that contribute the most to identify each category are highlighted before feeding the classification algorithm, acting as a kind of non-supervised pre-attention mechanism. In this manner, the solution proposed can deal with multi-label problems, is less sensitive to data scarcity and distribution, and can be combined with different classification models, including neural networks.

The evaluation of the proposed approach was carried out on datasets from SemEval (Pontiki et al., 2016), considering English and Spanish languages, two different application domains, as well as several works for comparison purposes.

Summarizing, the two main contributions of this paper are:

- A new term weighting scheme specially suited to the aspect category detection task, which acts as a kind of non-supervised attention mechanism.
- A detailed study on the effectiveness and adaptability of the proposed weighting scheme, considering different languages, domains and classification models.

The remainder of the paper is organized as follows. Section 2 presents a brief overview of previous work on aspect category detection. Section 3 describes the proposed weighting scheme. Sections 4 reports the experiments, results, and their analysis. Finally, Section 5 points out our conclusions and future work.

2 Related work

Aspect category detection is a subtask of aspect-based sentiment analysis, which attempts to assign a subset of categories from a set of predefined aspect categories to a given opinion (López Ramos and Arco García, 2019). This subtask was introduced and defined at the SemEval workshop: "An aspect category expresses, in a general way, the characteristics evaluated of an entity. Aspect categories are usually not defined by terms present in opinions instead they are inferred through terms used to evaluate different aspects. Category detection is a challenging problem due to the existence of overlapping categories" (Pontiki et al., 2016).

Previous research works in this topic can be organized according to their classification strategy in: lexicon-based, unsupervised, supervised and hybrid (Liu and Zhang, 2012). Despite the existence of successful lexicon-based and unsupervised methods (Ghadery et al., 2018), the majority of the proposed approaches follow a supervised learning approach, considering hand-crafted representations and using classification algorithms such as SVM, K-NN, Logistic Regression or ensembles of them (Xenos et al., 2016), (Hercig et al., 2016), (Hetal and others, 2021).

Over the last few years, deep learning models have brought significant advances to the aspect category detection task. For example, in (Toh and Su, 2016) it is described the construction of a set of binary classifiers, one for each category, considering a variety of lexical and syntactic features, along with extra features learned from a Convolutional Neural Network (CNN). This approach was the best performer at SemEval 2016, and based on the analysis of its results, their authors concluded that the CNN output probabilities were the most relevant features, and that the combination of two different machine learning methods is a feasible approach for the task. In (Xue et al., 2017) the multi-label aspect classification was also handled by multiple one-vs-all binary classifiers, implemented through a neural network with BiLSTM and CNN layers. In addition, this work models the task as a multi-task learning problem, jointly solving the detection of aspect categories and the extraction of aspect terms. Its results showed an important performance improvement, confirming the synergy between both tasks. In (He et al., 2017) it is presented a deep neural network approach based on an attention mechanism, which was later modified and improved in (Movahedi et al., 2019). In this last work, instead of training several one-vs-all models, the authors proposed a single model, namely Topic-Attention Network, which detects aspect categories of a given review sentence by attending to different parts of the sentence based on different topics. Their results confirmed that a single attention may not be able to provide a good representation for reviews containing multiple aspects, and, therefore, pointed out the relevance of learning to weigh the terms based on the different categories. More recently, and in this direction, in (Zhang et al., 2021) it is presented a multilayer self-attention model to deal with aspect category detection. Particularly, it is a BERT-based multi-self-attention model, which uses multiple attentions to obtain relevant information of the multiple target categories. Despite obtaining competitive results, its authors pointed out the difficulties of the attention model to correctly handle short texts, since they provide very limited contextual information.

Following the previous ideas, our approach seeks to emphasize the opinion terms by weighting them in accordance to their semantic proximity to each of the aspect categories, thus generating as many weights for each term as the number of categories. In this way, the terms that could contribute the most to identifying any given category are highlighted prior to feeding the corresponding binary classifier, acting as a kind of non-supervised attention mechanism.

3 Proposed method

Figure 1 shows the general diagram of the proposed method. It follows a *one-vs-all* approach, which means that it uses as many classifiers as aspect categories in the training set. The main components of our method are: *i*) the weighting of terms, *ii*) the construction of the opinions' representations, and *iii*) the classification process. The following subsections detail the first two components since they represent the core contribution of our work. For the classification process we consider traditional as well as deep leaning models, which are described in Section 4.

3.1 New term weighting scheme

The purpose of the proposed term weighting scheme (named as SP for *semantic proximity*) is to emphasize the contribution of each opinion term for the detection of each aspect category. Accordingly, we calculate as many weights for each term as the number of aspect categories.

Given a set of aspect categories, $\mathbf{C} = \{C_1, C_2, ..., C_n\}$, where each category is represent by a pre-defined lexicon or set of terms, $C_i = \{t_{i1}, ..., t_{in}\}$, and using pre-trained word embeddings from GloVe (Pennington, Socher, and Manning, 2014)[1] to represent each term, with $emb(t)$ indicating the embedding vector of term t, we compute the semantic proximity weight of term t_i for category C_j as follows:

1. Define the representative vector of the category C_i, referred as $emb(C_i)$. This vector is computed as the average of the term vectors of the category lexicon:

$$emb(C_i) = \frac{1}{|C_i|} \sum_{\forall t \in C_i} emb(t) \quad (1)$$

2. Measure the semantic proximity of term t_i with respect to category C_j. This proximity is computed by the cosine similarity between the term and category embeddings:

$$SP_{C_j}(t_i) = \cos(emb(t_i), emb(C_j)) \quad (2)$$

According to this new term weighting scheme, the opinion's words that are strongly related to the lexicon of the aspect category that is under analysis will have a greater weight than those from less related words. Figure 2 shows two opinions along with the semantic proximity weights of their words relative to the ambience-general and food-quality categories, respectively.

[1]The experiments were carried out using pre-trained embeddings on Wikipedia 300d and Twitter 200d. Twitter GloVe embeddings were chosen due to their orientation towards language and text size in social networks, Wikipedia GloVe were used to extend the coverage of the vocabulary. Nonetheless, alternative options could be considered.

3.2 Opinion representations

As we previously mentioned, the proposed weighting scheme, which acts as a kind of unsupervised pre-attention mechanism, can be used in combination with different classification models, including traditional classifiers such as the SVM, as well as deep neural networks like a CNN. In the first case, the SP weights are integrated under the Bag of Words representation, while in the second case these weights are used to alter the embeddings that feed the networks. Both cases of opinion representation are described below.

Representation for a traditional classifier. In this case, opinions are represented using a Bag of Words model. Accordingly, each opinion or document is represented by a vector $d = \langle w_1, w_2, ..., w_m \rangle$, where m is the size of the training vocabulary and w_i indicates the weight of term t_i in the opinion. We propose to define these weights as a combination of the term frequency and the term semantic proximity to the aspect category of interest[2] C_j as follows:

$$w_i = tf(t_i) \times SP_{C_j}(t_i) \quad (3)$$

Representation for a deep neural network. In this case, opinions are represented by the array of their word embeddings. Thus, an opinion or document having k terms will be represented by an array of the form $d = [emb(t_1), emb(t_2), ..., emb(t_k)]$. We propose to alter each of these embeddings by multiplying them by a scalar that indicates the relevance of each term, that is, by the semantic proximity SP of each term to the category of interest C_j. Based on this, the new embedding of a term t_i, denoted as $emb'(t_i)$, is computed as indicated in Formula 4, and the new representation of the opinion d is as indicated in Formula 5.

$$emb'(t_i) = SP_{C_j}(t_i) \times emb(t_i) \quad (4)$$

$$d = [emb'(t_1), emb'(t_2), ..., emb'(t_k)] \quad (5)$$

[2]Please note that we follow a one-vs-all classification approach, and, thus, we have a different classifier for each aspect category.

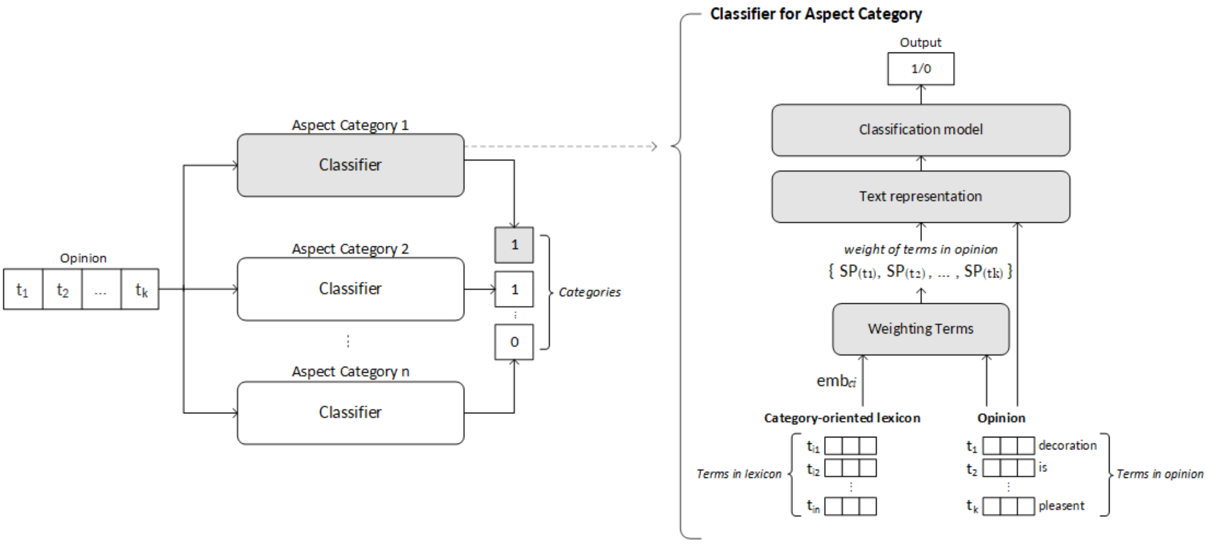

Figure 1: The proposed method. The left part depicts its general view, which is based on a one-vs-all classification approach; the right part details the main components of each aspect category classifier.

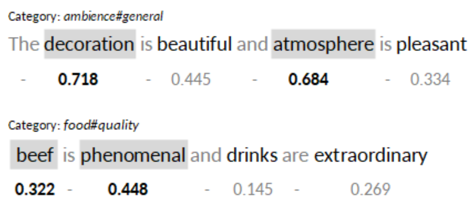

Figure 2: Two opinions with their respective SP weights. Words such as "decoration" and "atmosphere" are highly related to the ambience-general category, whereas "beef" and "phenomenal" to the food-quality category.

4 *Experiments*

4.1 Datasets

For the experiments we considered two datasets from SemEval 2016 (Pontiki et al., 2016). Particularly, we used the collections for the restaurant domain in both English and Spanish. Table 1 describes the distribution of these collections. As pre-processing operations, we eliminated punctuation marks and special characters; when using the SVM classifier we also removed stopwords.

4.2 Category-Oriented Lexicons

The category-oriented lexicons that our method takes as input can be manually or automatically defined. For the experiments reported, we extracted these lexicons from the training sets using the SS3 method recently proposed in (Burdisso, Errecalde, and Montes-y Gómez, 2019). This method associates each vocabulary term with a confidence value for each of the categories. This value is a number in the interval [0,1] and represents the degree of confidence with which a term is believed to exclusively belong to given category. For example, the term "tequila" will have a confidence value close to 1 for the "drinks" category, whereas the term "of" will have a value equal or close to 0 because it is similarly distributed across all categories.

	instances			
	spanish		**english**	
Categories	*traing*	*test*	*traing*	*test*
ambience#general	293	126	255	66
drinks#quality	31	10	47	22
drinks#style_options	29	11	32	12
drinks#prices	14	10	20	4
food#quality	845	291	849	313
food#style_options	192	69	137	55
food#prices	127	41	90	23
restaurant#general	540	222	422	142
restaurant#miscellaneous	14	13	98	33
restaurant#prices	115	39	80	21
service#general	504	222	449	155
location#general	15	18	28	13

Table 1: Aspect categories in restaurant datasets.

As defined in the SemEval workshop, aspect categories are formed by an entity and attribute pair (e.g., for the category food#quality, the entity is "food" and the attribute is "quality"). In particular, for the restaurant data set there are 12 predefined aspect categories which are listed in Table

1. From these categories, 6 different entities E = {ambience, drinks, food, restaurant, service, location} and 5 different attributes A = {general, miscellaneous, prices, quality, style_options} are derived. We extracted lexicons for each one of the entities and for each one of the attributes, and then we made the corresponding unions to define the lexicons for the 12 aspect categories.

As stated above, the SS3 method (Burdisso, Errecalde, and Montes-y Gómez, 2019) determines a confidence value for all terms with respect to all categories. In order to only include in the lexicons the terms most strongly associated with each category, we propose to filter them using the following criteria: we consider confidences are normally distributed, and thus we keep the terms whose confidence values are equal or above β standard deviations from the mean, considering $\beta = \{1, 2, 3\}$. Table 2 shows five terms for four different lexicons, two in Spanish and two in English.

ambience#general		**food#quality**	
term	*value*	*term*	*value*
ambiente	1.000	calidad	1.000
mesas	1.000	caros	1.000
tranquilo	0.976	platos	0.977
terraza	0.921	excesivos	0.809
bonito	0.846	proporcion	0.749

a) Category-oriented lexicons in Spanish

location#general		**drinks#style_options**	
term	*value*	*term*	*value*
located	1.000	champagne	1.000
chart	0.870	martinis	1.000
block	0.870	well	0.922
sidewalk	0.870	generously	0.922
conveniently	0.870	guaranteed	0.802

b) Category-oriented lexicons in English

Table 2: Example of category-oriented lexicons. For each term its confidence value according to SS3 is indicated.

4.3 Experimental settings

For the experiments, we employed the method proposed in combination with two classification algorithms, a SVM and a CNN. For each classification model, we trained 12 binary classifiers, one for each aspect category. For the sake of simplicity, for all classifiers we used the same settings; the used hyperparameter values are as follows:

- **SVM**: C = 2.5, kernel = linear, and Degree = 2.5. For this particular case, the terms considered for the BOW representation were those present in training set opinions from the category under analysis, without considering empty words[3].
- **CNN**: We used a combination of kernels of sizes 1,2,3 in the convolutional layer to create different feature maps (Gehrmann et al., 2018). The rest of settings for them are: activation=relu, pool-size = max_length - kernel-size + 1, strides=1. The general settings of CNN architecture are: epochs = 9, filters = 256, dropout = 0.5, activation function = sigmoide, loss function = binary crossentropy, and optimizer = adam.

Due to the fact that the datasets show a high level of imbalance, particularly because the task was approached under the one-vs-all approach, we carried out additional experiments applying oversampling over the minority categories, which indeed correspond to the aspect category under analysis for each binary classifier. In particular, we applied an oversampling technique that consists in randomly replicate instances of the minority class until reaching the size of the majority class.

4.4 Baseline results

As baseline results we used those obtained by the same two classifiers but using the traditional representations. That is, for the SVM we used a BOW with tf-idf weigths, without including our SP weights. In the case of the CNN, we fed it with the pretrained Glove embeddings without having altered them with our SP weights. We also consider the baseline results reported for each SemEval 2016 task (Pontiki et al., 2016).

Additionally, we compared our results against those from state-of-the-art constrained[4] methods. In particular, we considered the top 3 results for the Spanish and the English datasets. The works considered are:

- *GTI* (Alvarez-López et al., 2016). It uses a support vector machine, but also

[3]Empty words defined in the Python NLTK library are considered.

[4]Works that use external resources such as datasets or dictionaries are classified as *U: Unconstrained* and those that do not use any type of additional resource are classified as *C: Constrained.*

a manually debugged list of words obtained from the training set to remove inter-category noise.

- *TGB* (Çetin et al., 2016). It uses a two-layer approach. In first layer considers a one-vs-all classification approach, where probabilities for entities and attributes are computed. Then, in a second layer, these probabilities are combined to understand which is the best combination of entity and attribute in order to determine the target aspect categories.

- *UWB* (Hercig et al., 2016). it implements a classifier per category using maximum entropy approach. A large number of features are considered for the construction of classifiers.

- *BUTkn* (Macháček, 2016). It uses a set of word n-grams manually compiled for each aspect category and then classifiers the opinions by looking for the occurrence of these n-grams.

- *XRCE* (Brun, Perez, and Roux, 2016). It adapts a component that extracts semantic information about entities and attributes. A dependency graph is created in which the relationships of a term with respect to categories are represented. The classification is performed by looking for word matches in the opinions and their relationships with categories.

4.5 Experimental Results

Table 3 presents a summary of the best results obtained with our method as well as with the baseline configurations. The second and third columns refer to the configuration of the classifier (kind and whether or not oversampling was used), while the fourth and third columns indicate the configuration used to calculate the SP term weights. When comparing the results obtained with and without using the SP weights, the usefulness of the proposed method is clearly appreciated. For the Spanish collection, our best result is achieved with the SVM classifier using oversampling, and with $\beta = 2$ and using the Twitter Glove embeddings of 200 dimensions for the computation of the SP weights. For English, the best results were achieved with the CNN architecture with oversampling, and with $\beta = 2$ and using Wikipedia Glove embeddings of 300 dimensions for the computation of the SP weights.

Regarding the comparison against state-of-the-art methods, Table 4 shows our method's results as well as the best constrained results from SemEval 2016 in both Spanish and English. These comparisons evidence the relevance of the proposed method, since, in addition to its simplicity and generality, it shows competitive results in both scenarios; in particular it outperforms the best results previously reported in the Spanish dataset.

4.5.1 Discussion

In contrast to most previous works, the proposed method does not necessarily need to use external resources for its implementation. For example, the lexicons used by the proposed weighting scheme can be automatically extracted from the training set, as we did in the experiments. For those categories with a considerably reduced number of instances, the number of terms extracted for their lexicons is also reduced. However, our experiments showed that the number of terms per lexicon is not directly related to the results obtained per category. Tables 5 and 6 contrast the number of instances, number of terms in lexicons, and results achieved for each of the aspect categories of the two datasets.

Despite the small number of terms in the lexicons, the proposed weighing scheme demonstrates its usefulness by paying different levels of attention to those terms strongly related to the categories, which helps to determine whether or not a category is present in an opinion. Analyzing the terms that constitute all category-oriented lexicons (refer to Figure 3), we observe that they are clearly representative of the different categories, despite they were automatically extracted. In consequence, an advantage of this approach is that it can be easily adapted to other domains or languages, as was observed in the experiments.

To demonstrate the generality of our method, we performed another experiment using the English laptop dataset. The achieved results are shown in Table 7. In this case, only the best constrained work is taken as a reference for comparison. Our best result was achieved using the SVM classifer, applying oversampling for class balancing, and considering $\beta = 2$ with GloVe Wikipedia em-

Spanish	Classifier	Oversampling	Embeddings	Threshold	micro-F1
Our method	**SVM**	**Yes**	**Twitter 200d**	$\boldsymbol{\beta = 2}$	**71.11**
Baseline	SVM	Yes	-	-	64.30
Our method	CNN	Yes	Twitter 200d	$\beta = 1$	69.99
Baseline	CNN	Yes	Wikipedia 300d	-	60.50
Baseline SemEval	-	-	-	-	54.68

English	Classifier	Oversampling	Embeddings	Threshold	micro-F1
Our method	SVM	Yes	Twitter 200d	$\beta = 2$	66.50
Baseline	SVM	No	-	-	64.30
Our method	**CNN**	**Yes**	**Wikipedia 300d**	$\boldsymbol{\beta = 2}$	**68.50**
Baseline	CNN	Yes	Wikipedia 300d	-	66.80
Baseline SemEval	-	-	-	-	58.92

Table 3: Micro-F1 results of the aspect category detection task, for the restaurant domain in Spanish and English. For each method the result of its best configuration is included; the best overall result is highlighted in bold.

	micro-f1
SP+SVM	**71.111**
GTI	70.027
TGB	63.551
UWB	61.968

a) Spanish

	micro-f1
BUTKn.	71.494
XRCE	68.701
SP+CNN	**68.500**
UWB	67.817

b) English

Table 4: Comparison against SOTA results by constrained methods in the aspect category detection tasks from SemEval 2016. The results of our method correspond to the SP+SVM and SP+CNN configurations for Spanish and English, respectively.

Category	Training Intances	#Terms	Test Instances	%Errors
ambience#general	293	40	126	30.73
drinks#quality	31	17	10	90.00
drinks#style_options	29	9	11	45.45
drinks#prices	14	8	10	55.55
food#quality	845	17	291	20.66
food#style_options	192	62	69	58.06
food#prices	127	8	41	20.00
restaurant#general	540	23	222	29.22
restaurant#miscellaneous	14	28	13	100.00
restaurant#prices	115	23	39	52.63
service#general	504	31	222	16.55
location#general	15	47	18	83.33

Table 5: Number of opinions to classify and lexicon terms for each aspect category, as well as the percentage of incorrectly classified instances by our best result in the Spanish dataset.

Category	Training Instances	#Terms	Test Instances	%Errors
ambience#general	255	17	66	78.79
drinks#quality	47	10	22	45.45
drinks#style_options	32	9	12	58.33
drinks#prices	20	7	4	75.00
food#quality	849	6	313	19.17
food#style_options	137	6	55	50.91
food#prices	90	6	23	47.29
restaurant#general	422	4	142	42.14
restaurant#miscellaneous	98	4	33	78.79
restaurant#prices	80	10	21	47.62
service#general	449	4	155	21.29
location#general	28	1	13	53.85

Table 6: Number of opinions to classify and lexicon terms for each aspect category, as well as the percentage of incorrectly classified instances by our best result in the English dataset.

beddings of 300 dimensions for the computation of the SP weights. As can be noticed, our result is significantly higher than the reported SemEval's baseline result, and it also outperforms the best reported constrained result. It is important to highlight that, in spite of the large number of categories in this dataset, 67 aspect categories derived from 22 entities and 9 attributes, for the experiments we did not carried out any additional hyperparameter adjustment.

	micro-f1
SP+SVM	**62.10**
UWB	60.45
Baseline SemEval	52.68

Table 7: Micro F1 results in the aspect category detection using the English laptop dataset.

4.5.2 Error analysis

Doing a detailed analysis of the errors (refer to Tables 5 and 6), it can be noticed that

Figure 3: Terms of three category-oriented lexicons of the restaurant domain.

there is no a clear relationship between the characteristics of the datasets and error rates. For the Spanish dataset, the correlation between the number of training instances and the percentage of errors is $r = -0.690$, while for the English dataset $r = -0.664$. Surprisingly, these values indicate that there was a slight tendency to misclassify instances from categories with many examples. This is partly because these more frequent categories are more diverse and also because they usually appeared together with others.

On the other hand, when correlating the size of the categories' lexicons with the classification errors, for Spanish we obtained a $r = 0.214$ and for English $r = 0.355$, suggesting little influence of this variable on the results. However, analyzing these lexicons in greater depth, we found that although they seemed related and relevant to the different categories, they showed a high overlap. In the Spanish dataset, for the smallest categories more than 80% of their terms are also included in others. One interesting example is the entity category "drinks", for which any term was unique; even more, 40% of its terms are also in the lexicons of four or more categories. For English, a similar behavior was observed, but, in addition, we noticed that for the category "location#general", with a single term but exclusive to this category, we obtained a better result than for other categories with larger lexicons. This suggests us the influence of the category-oriented lexicons in whole process, as well as the need to define them more precisely.

5 Conclusions

The main contribution of this work is the proposal of a new term weighting scheme specially suited to the aspect category detection task. It is based on the evaluation of the semantic proximity of each term in an opinion with respect to the categories' description, acting as a kind of non-supervised attention mechanism.

The proposed weighting scheme relies on the availability of a set of category-oriented lexicons, nonetheless, they can be automatically extracted from the training dataset. This latter characteristic makes the method easily adaptable to different domains and languages.

From the results obtained, it is possible to conclude that the proposed term weighting scheme has a positive impact on the identification of the categories of aspects expressed in an opinion. Moreover, it has the advantage of being able to be combined with different classification models, including traditional machine learning classifiers as well as deep neural networks.

On the other hand, although the method is less sensitive to small and imbalanced datasets than other supervised approaches, it is affected by these conditions. As it was observed, the method achieved better results in Spanish than in English, being the latter the collection with less instances and high imbalance rates.

As working directions, we plan to evaluate the method using different kinds of category lexicons, both manually and automatically generated. Besides that, we seek to evaluate our method in collections having different volumes of information, as well as using different contextual word embeddings such as those from BERT. Furthermore, due to the generality of the proposed method, we plan to apply it in other text classification tasks such as polarity classification, author profiling and fraud detection.

Acknowledgments

The present work was supported by CONACyT/México (scholarship 756974 and grant CB-2015-01-257383). In addition, the authors thank CONACYT for the computational resources provided by the Deep Learning Platform for Language Technologies.

References

Alvarez-López, T., J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castano. 2016. Gti at semeval-2016 task 5: Svm and crf for aspect detection and unsupervised aspect-based sentiment analysis.

In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 306–311.

Brun, C., J. Perez, and C. Roux. 2016. XRCE at SemEval-2016 task 5: Feedbacked ensemble modeling on syntactico-semantic knowledge for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 277–281, San Diego, California, June. Association for Computational Linguistics.

Burdisso, S. G., M. Errecalde, and M. Montes-y Gómez. 2019. A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182–197.

Çetin, F. S., E. Yıldırım, C. Özbey, and G. Eryiğit. 2016. Tgb at semeval-2016 task 5: multi-lingual constraint system for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 337–341.

Chaudhari, S., V. Mithal, G. Polatkan, and R. Ramanath. 2021. An attentive survey of attention models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(5):1–32.

Gehrmann, S., F. Dernoncourt, Y. Li, E. T. Carlson, J. T. Wu, J. Welt, J. Foote Jr, E. T. Moseley, D. W. Grant, P. D. Tyler, et al. 2018. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PloS one*, 13(2):e0192360.

Ghadery, E., S. Movahedi, H. Faili, and A. Shakery. 2018. An unsupervised approach for aspect category detection using soft cosine similarity measure. *arXiv preprint arXiv:1812.03361*.

He, R., W. S. Lee, H. T. Ng, and D. 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, July. Association for Computational Linguistics.

Hercig, T., T. Brychcín, L. Svoboda, and M. Konkol. 2016. Uwb at semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 342–349.

Hetal, V. et al. 2021. Ensemble models for aspect category related absa subtasks. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(13):2348–2364.

Liu, B. and L. Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*. Springer, pages 415–463.

López Ramos, D. and L. Arco García. 2019. Aprendizaje profundo para la extracción de aspectos en opiniones textuales. *Revista Cubana de Ciencias Informáticas*, 13(2):105–145.

Macháček, J. 2016. BUTknot at SemEval-2016 task 5: Supervised machine learning with term substitution approach in aspect category detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 301–305, San Diego, California, June. Association for Computational Linguistics.

Movahedi, S., E. Ghadery, H. Faili, and A. Shakery. 2019. Aspect category detection via topic-attention network. *CoRR*, abs/1901.01183.

Mowlaei, M. E., M. Saniee Abadeh, and H. Keshavarz. 2020. Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, 148:113234.

Pennington, J., R. Socher, and C. D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Pontiki, M., D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *10th International Workshop on Semantic Evaluation (SemEval 2016)*.

Toh, Z. and J. Su. 2016. NLANGP at SemEval-2016 task 5: Improving aspect based sentiment analysis using neural network features. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 282–288, San Diego, California, June. Association for Computational Linguistics.

Xenos, D., P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis, and I. Androutsopoulos. 2016. Aueb-absa at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In **SEMEVAL*.

Xue, W., W. Zhou, T. Li, and Q. Wang. 2017. Mtna: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 151–156.

Zhang, X., X. Song, A. Feng, and Z. Gao. 2021. Multi-self-attention for aspect category detection and biomedical multilabel text classification with bert. *Mathematical Problems in Engineering*, 2021.