Clustering Multilingual Aspect Phrases for Sentiment Analysis

Lucas Rafael Costella Pessutto
Institute of Informatics - UFRGS
Porto Alegre, RS – Brazil
lrcpessutto@inf.ufrgs.br

Danny Suarez Vargas
Institute of Informatics - UFRGS
Porto Alegre, RS – Brazil
dsvargas@inf.ufrgs.br

Viviane P. Moreira
Institute of Informatics - UFRGS
Porto Alegre, RS – Brazil
viviane@inf.ufrgs.br

***Abstract*—The area of sentiment analysis has experienced significant developments in the last few years. More specifically, there has been growing interest in aspect-based sentiment analysis in which the goal is to extract, group, and rate the overall opinion about the features of the entity being evaluated. Techniques for aspect extraction can produce an undesirably large number of aspects – with many of those relating to the same product feature. This problem is aggravated when the reviews are written in many languages. In this paper, we address the novel task of multilingual aspect clustering which aims at grouping together the aspects extracted from reviews written in several languages. We contribute with a proposal of techniques to tackle this problem and test them on reviews written in five languages. Our experiments show that our unsupervised clustering technique achieves results that outperform a semi-supervised baseline in many cases.**

***Index Terms*—Aspect-Based Sentiment Analysis, Multilingual Aspect Clustering, Unsupervised Learning, Word Embeddings**

I. Introduction

The dawn of the Web 2.0 changed the way users interact on the Internet, enabling more content production as people express their opinions over many subjects on multiple platforms. E-commerce systems allow users to give opinions about the products that are being sold. This information, in turn, becomes useful to other users as they can rely on previous shopping experiences from other people as a basis for their own purchases. Also, companies can take advantage of the opinions to measure the acceptance of a product and improve it according to the users' taste. While useful and valuable, reviews are difficult to process because they are often represented as large amounts of unstructured text. Moreover, in systems accessed on a global scale, opinions can be found in different languages, posing further difficulties to automatic processing.

Sentiment Analysis is the field of study which aims to process the information conveyed by unstructured texts, providing structured information that facilitates the understanding of the opinions, attitudes, or emotions towards a particular entity [1]. The main tasks in sentiment analysis include polarity attribution, aspect extraction, and opinion summarization. Polarity attribution consists in determining if the opinion expressed in a review is positive, negative, or neutral. Aspect Extraction is a more fine-grained task as its goal is to extract the features of the entity being reviewed. Opinion summarization aims to build a concise text that synthesizes the opinions about an entity from a large set of review texts [2].

In spite of the good results achieved by modern aspect extraction techniques, they can produce an undesirably large number of aspects. This happens because people use different words to express the same aspect of an entity [1]. For example, the words *screen*, *display*, and *touchscreen* refer to the same feature in the smartphone domain. In order to group together the terms that refer to the same feature, *aspect clustering* is employed. This is a fundamental step to allow the construction of summaries containing a small list of representative aspects that convey the users' overall opinion.

Reviews in multiple languages are abundant in a number of important sources such as TripAdvisor, AirBnB, Amazon *etc.* As a result, dealing with multilingual data becomes necessary. In this scenario, one language will be less represented than others lacking the required amount of data to allow for the algorithms to yield good results. In such cases, it is useful to rely on languages with a greater density of reviews. The combined use of multiple languages for sentiment analysis has proven useful and enabled reaching results that are significantly better than when a single language is considered [3].

The focus of this article is on *multilingual aspect clustering*, which can be defined as the task of grouping together equivalent aspects across multiple languages. To the best of our knowledge, this is the first work to address this problem. Our solution combines unsupervised clustering, syntactic term similarity, and word embeddings. We carried out experiments on restaurant reviews written in English, Spanish, Russian, Dutch, and Turkish and compared our performance against an established baseline. The results show that our unsupervised clustering technique achieves results that are better than the results of the semi-supervised baseline in many cases.

II. Problem Description and Solution Overview

In the context of multilingual aspect clustering, the set of input reviews can be defined as $R = \{R_1, R_2, ..., R_l\}$, where R_l corresponds to the subset of reviews in language l. All reviews belong to the same domain, for example, if R_l contains opinions about smartphones, all other subsets will also have reviews on smartphones. Each subset is composed of review sentences $R_l = \{s_{l_1}, s_{l_2}, ..., s_{l_m}\}$, where s_{l_m} denotes the m^{th} sentence in language l.

978-1-5386-7325-6/18/$31.00 ©2018 IEEE
DOI 10.1109/WI.2018.00-91

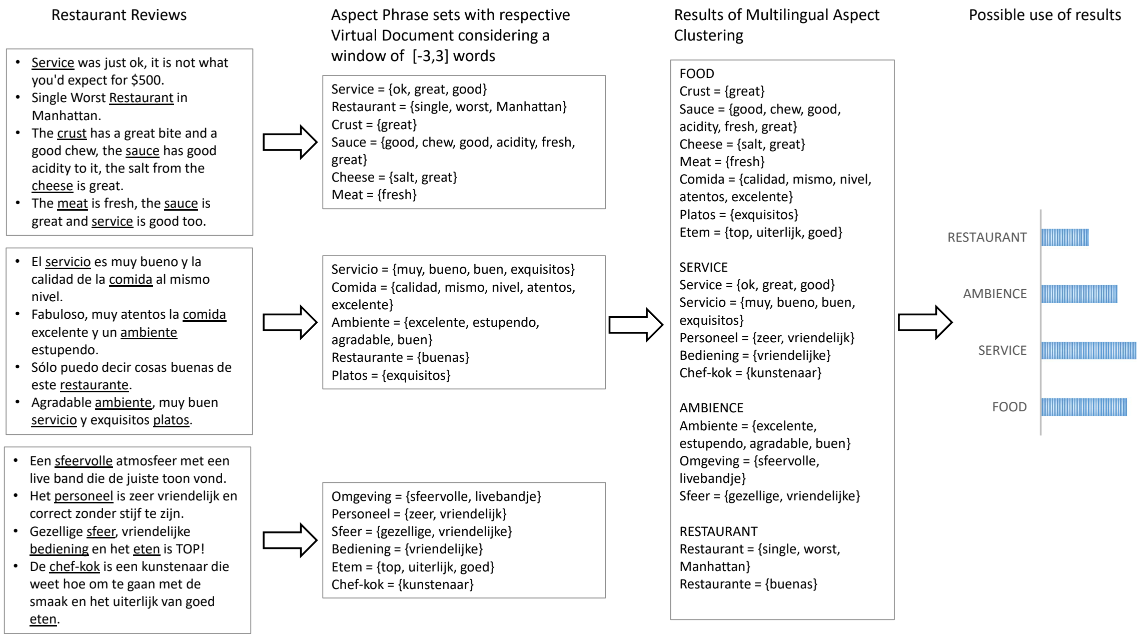


Figure 1: An example of Multilingual Aspect Clustering

Aspect extraction techniques may be employed on reviews to extract the properties of the target entity (please refer to Section III-A for more details on aspect extraction). The explicit properties of a target that occur in a set of reviews are referred to as *aspect phrase* (also called *product feature* or *surface form* in the literature). An aspect phrase is composed of one or more terms (*e.g., battery*, *battery life*).

Users can express the same product features using different words or phrases. Thus, a clustering step is necessary to group the aspect phrases that belong to the same category. Each of these groups will be called *Aspect Group* and will consist of a set of aspect phrases in multiple languages.

An example of Multilingual Aspect Clustering can be seen in Figure 1. Initially, the set of reviews (R) is formed by the union of the subsets of reviews in three languages (English, Spanish, and Dutch) with the aspect phrases already extracted. The next step is to build the virtual documents for each aspect phrase. Then, we can apply a clustering algorithm on the aspects and on the virtual documents in order to group together semantically similar aspect phrases. In our example, four clusters are formed representing the aspects Food, Service, Ambience, and Restaurant. Once the data is clustered, we can use it to build summaries that synthesize the sentiment expressed by the reviewers.

III. Related Work

In this Section, we review the related literature about two key topics: Sentiment Analysis (with emphasis on aspect clustering) and Multilingual Document Clustering.

A. Sentiment Analysis

A few years ago, Feldman [4] mentioned that over 7000 research papers had already been written about sentiment analysis. This demonstrates the significant interest in this area, which has the goal of labeling texts of different granularities (entire reviews, sentences, and aspects) and different levels of analysis.

The first approaches in this field attempted to classify sentiments at the document or sentence level, which means assigning a polarity (positive, negative, or neutral) to the entire review text/sentence. Solutions for this problem apply a wide range of methods from lexicon-based resources [5] to deep learning techniques [6]. Dashtipour *et al.* [7] point out that while most publications focus on the English language, few study techniques that address the multilingual sentiment analysis problem. Lo *et al.* [8] studied the existing approaches for multilingual sentiment analysis and classified them in three categories – lexicon, corpus, and translation-based. There are also approaches that use a combination of these techniques.

The aspect level appears as the most important mainly due to the rich information that it conveys [9]. In this level of analysis, often called Aspect-Based Sentiment Analysis (ABSA), there is the need to identify aspects and entities in natural language texts. The aspect-term extraction task can be classified into three main groups according to the underlying approach [2]: (i) based on language rules [10]–[12]; (ii) based on sequence labeling models [13], [14]; and (iii) based on topic models [15]. However, there are other works that do not fit in only one of these groups as they combine resources from more than one approach [16], [17]. Furthermore, state-of-the-art approaches rely on more sophisticated architectures like recurrent neural networks such as LSTM, Bi-LSTM, and Neural Attention Models [18]–[21].

Most of the existing work on aspect extraction is designed to deal with reviews written in English. However, in the last few years, researchers started to explore aspect extraction in other languages. In 2016, SemEval made available multilingual datasets for sentiment analysis at aspect level. Participants could use one or more languages in their solutions. The evaluation campaign received 245 submissions from 29 teams for that task [22]. Such datasets boosted the research in this area. For example, García-Pablos *et al.* [23] propose a topic modeling solution for multilingual aspect extraction and classification, which is almost unsupervised (requiring only a few seed words per language) and achieves competitive results. Al-Smadi *et al.* [24] improved the results on the Arabic dataset by exploring morphological, syntactic, and semantic features from the Arabic language.

After performing the aspect extraction task, the clustering of aspects is necessary to group together the representations of the same aspect (*e.g., price*, *cost*, and *charged amount* all refer to the same aspect of a given product). Next, we report on existing approaches for aspect clustering.

Dictionaries and Taxonomies: The first approaches for aspect clustering relied on pre-existing resources such as Synonym Dictionaries [25] or Taxonomies [26]. Dictionaries are usually not considered as good resources for this task because they cannot map contextual similarity between expressions. Also, many aspects such as brand names, places or domain-specific words do not typically appear in dictionaries. Taxonomies have the disadvantage of being domain dependent and are difficult to build and maintain. Therefore, these approaches are no longer used.

Topic-Modeling Techniques: The algorithms in this category employ techniques such as Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA), in order to group similar aspect phrases taking into account the semantic similarity between aspects. Guo *et al.* [27] proposed a multi-level LSA (mLSA) approach which builds two LDA models in order to group product aspect phrases. It requires semi-structured reviews (with marked positive and negative aspects of the product/service) to work. Zhai *et al.* [28] proposed a modification on the original LDA method to support soft constraints like must-links and cannot-links.

Semi-Supervisioned Algorithms: This technique was widely used for aspect clustering. It consists in labeling part of the input data with the cluster information to facilitate and improve cluster formation. In most cases, this information is obtained automatically from the data. The seminal work in this category is by Zhai *et al.* [29] who automatically obtains the labeled data by leveraging lexical similarity and contextual information. The aspect phrases are first grouped according to the words they share. For example, “cake", “chocolate cake", and “lava cake" are joined together in the same cluster in this phase. Next, the lexical similarity between the groups formed in the previous phase is measured in order to combine the n most similar groups. Also, the contextual information about the aspects is taken into account through the use of virtual documents, which consider a window of $[-t, t]$ words before and after every aspect phrase occurrence. Once part of the data is labeled, the EM algorithm based on a Naïve Bayes classifier is employed to cluster the aspect phrases. Subsequent works changed some characteristics of Zhai *et al.*'s proposal [29]. For example, there are variations of the pre-grouping heuristics, such as taxonomies automatically extracted from e-commerce pages [30], statistical distribution of data in positive and negative reviews [31], and co-occurrence of aspects and words in reviews [32]. Different algorithms were also used in this task, like k-means [33], [34], Multinomial Naïve Bayes [31], and Spectral Clustering [35].

Other Clustering Approaches: There are also a few proposed solutions to address the problem of clustering aspect phrases which do not fit into the categories discussed in the previous sections. For example, some works [36]–[38] use hierarchical clustering in order to produce multi-granular summaries, which can be customized according to the user's needs. Cao *et al.* [39] clustered aspect phrases and opinion words simultaneously by using a constrained hidden Markov random field model. Finally, Jiajia *et al.* [40] combined a feature-opinion relation matrix with two constraint matrices in their clustering model.

B. *Multilingual Document Clustering*

The task of Multilingual Document Clustering aims to group documents written in more than one language according to their subjects. To achieve this goal, one can use machine translation techniques in order to translate entire documents [41], [42] or just some document features [43] to create a monolingual space and then apply monolingual document clustering.

Other approaches in this area attempt to extract language-independent features from the multilingual documents. For example, Denicia-Carral *et al.* [44] cluster documents in linguistically related languages by orthographic and thematic similarity. Named Entity co-occurrences also helped clustering news in Spanish and English [45].

The works in this area were developed mostly for grouping news articles, which are typically longer documents. So far, the application of these techniques for clustering reviews (or aspects extracted from reviews) remains unexplored.

IV. Multilingual Aspect Clustering

In this work, we add techniques inspired by multilingual document clustering to (monolingual) aspect clustering in order to address multilingual aspect clustering. In order to group multilingual aspects, our proposed approach leverages the contextual information of aspect phrases and word embeddings in an unsupervised clustering algorithm. Figure 2 shows our proposed framework. An important difference between our work and existing proposals, especially in relation to Zhai *et al.*'s [29], is the use of unsupervised learning, which does not require a labeling step before clustering aspect phrases. This makes our method completely automatic.

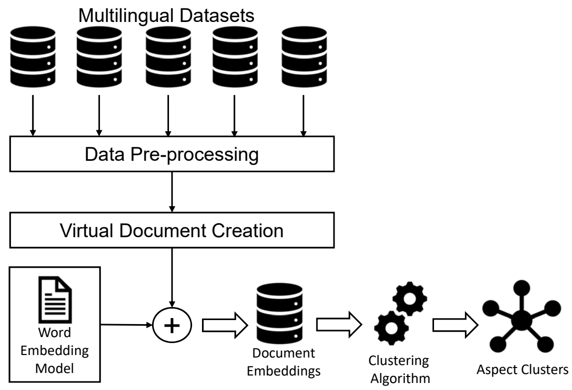


Figure 2: Proposed Multilingual Aspect Clustering framework

The input for our method is a set of multilingual reviews along with the extracted aspect phrases. In this paper, we assume that the aspects have already been extracted beforehand, so we have not focused on the aspect extraction task.

We start with a preprocessing phase that consists of three standard steps: (i) splitting of the review text into sentences; (ii) tokenization; and (iii) converting all words to lowercase.

Once the data is preprocessed, the virtual documents for each aspect phrase are built. This step follows the proposal by Zhai *et al.* [29], and consists in extracting the surrounding words of each occurrence of an aspect phrase. We extract words in a $[-t, t]$ window, removing stopwords and other aspect phrases that co-occur in the same sentence. The Virtual Document of an aspect phrase is the concatenation of the surrounding words of all occurrences of that aspect phrase in the dataset.

Once the Virtual Document Creation phase is concluded, we have l sets of aspect phrases and their respective documents, where l is the total number of languages present on the datasets. In order to group the aspect phrases together, we need to generate a common representation for the aspect phrases that is language independent. Therefore, we employ *multilingual word embeddings* to this task. Word Embeddings are a language model that represents words as low dimensional vectors, keeping the distributional similarities between words [46]. Since our documents will be formed by reviews in different languages, it is necessary that the word embeddings can handle this kind of data. For that, one can use embeddings trained with multilingual data, with parallel or comparable corpora, or employ techniques that can transform monolingual word embedding spaces into compatible multilingual ones [47].

Multilingual word embeddings have the advantage of sharing semantics between words in different languages and the capacity of knowledge transfer from resource-rich languages to the ones with few resources. These factors motivated the emergence of several approaches in recent years. Some approaches learn a matrix capable of performing the linear transformation of monolingual word embeddings in language x to the word embeddings of a language y. These approaches usually rely on bilingual dictionaries or on other translation tools to identify related words across languages. Many techniques also trained multilingual word embeddings from parallel or comparable multilingual sources [47].

The *Document Embeddings* are formed by the union of the word embedding representations of each word in the Virtual Document and the word embeddings of each word in the aspect phrase. This task can be categorized as unsupervised learning because we only use the Document Embeddings as input for our clustering algorithm. We do not use semi-supervised algorithms in our approach, because we work with multilingual data. Reviews in multiple languages do not fit into the existing labeling techniques proposed for semi-supervised approaches for monolingual aspect clustering, which basically focus on the lexical similarity between aspects. Another important fact that makes us use an unsupervised approach is that we do not want to rely on the translation of the reviews, neither on the manual labeling of the aspect phrases. The goal is to make our approach as simple as possible.

The last step in our approach is to cluster the Document Embeddings in order to group together aspect phrases with the same semantic context. We employ a centroid-based clustering algorithm to that task. The pseudocode is shown in Algorithm 1. The inputs for the algorithm are the Document Embeddings and the desired number of clusters (k). The output is a set of clusters of related Document Embeddings. It initially selects k documents as the centroids (lines 3-5). This selection can be done randomly, or by using some heuristic. We perform experiments selecting the centroids randomly and choosing the aspect phrases that appear more frequently in the dataset.

Then, for the remaining documents, their distance is measured in relation to the centroids, and the document is assigned to the nearest cluster (lines 7-12). When all documents have been assigned to a cluster, new centroids are chosen, and the documents are re-assigned (lines 13-15). This process converges when there are no changes for an entire iteration.

ALGORITHM 1: Multilingual Aspect Clustering (MAC)

Input : k – Number of Clusters
DE = $\{de_1, de_2, ..., de_i\}$ – Document Embeddings
Output: DC = $\{c_{de_1}, c_{de_2}, ..., c_{de_i}\}$ – Set of Document Clusters

1 DC $\leftarrow \emptyset$
2 centroids $\leftarrow \emptyset$
3 **for** $i \leftarrow 0$ ***to*** k **do**
4   centroids $\leftarrow$ `cent_selection`(*DE*)
5 **end**
6 **while** *convergence* **do**
7   **for** $i \leftarrow 0$ ***to*** $|DE|$ **do**
8   **for** $j \leftarrow 0$ ***to*** $|centroids|$ **do**
9   distance $\leftarrow$ `WMD` (*DE[i], centroids[j]*)
10   **end**
11         $c_{de}[i] \leftarrow$ cluster number with the lowerst distance
12   **end**
13   **for** $i \leftarrow 0$ ***to*** $|centroids|$ **do**
14   centroids[i] $\leftarrow$ The result of Eq. 1
15   **end**
16 **end**
17 **return** *DC*

The number of expected clusters (k) depends essentially on the characteristics of the dataset and on the goals of the analysis. Increasing k yields finer granularity which may be desirable in some settings. In Section V-B, we assess how different values of k impact on the quality metrics.

The distance measure we use in order to compare two Document Embeddings is the Word Mover's Distance (WMD) [48]. We chose this measure because it can capture the semantic dissimilarity of two context words from two different aspects and it also works with word embeddings. The goal of WMD is to measure the minimum traveling cost of the words in one document to words in the other document. The WMD between two documents is calculated by the summation of the smallest distance between each word in the first document and the words in the second one. It works even if the two documents have no words in common. For example, the sentences "*The wine list has interesting good values*" and "*They have a good beverage menu with reasonable prices*" has a small WMD score as they contain words that are semantically related *e.g., wine* and *beverage*, *list* and *menu*, and *values* and *prices*.

We designed a centroid selection method for our problem (instead of using the mean of document embeddings, for example) because the WMD distance requires two documents for its calculation, which forced us to always have a document embedding as a centroid of our cluster. The new centroid will be the one that has the lowest WMD average in relation to the other document embeddings belonging to that cluster. Eq. 1 shows how the centroid is chosen.

$$centroid(c) = \min_i \sum_{i=0}^{|c|} \frac{1}{|c|} \sum_{j=0}^{|c|} WMD(DE_i, DE_j) \qquad (1)$$

where $|c|$ corresponds to the number of elements in a cluster.

V. Experimental Evaluation

In this Section, we describe the evaluation of our proposed multilingual aspect clustering technique. Initially, the experimental setup is presented, and then the results are discussed.

A. Experimental Design

1) Datasets: We used the Restaurant dataset from SemEval 2016 - Task 5[1] in order to evaluate our approach. This is a multilingual dataset with reviews in five languages: English, Dutch, Russian, Spanish, and Turkish. It was originally designed for the aspect extraction task.

Every sentence in this dataset has annotations about the aspect phrases classified into six aspect clusters: Restaurant, Food, Drinks, Service, Ambience, and Location. We used this classification scheme as the gold standard in our evaluations. Some statistics of the datasets can be found in Table I. It is important to notice that we only used the explicit aspect phrases in the reviews, which means we did not consider targets marked as `NULL` in the dataset. Some aspect phrases are categorized in more than one of the above aspect clusters, so we chose as the category the one which has more assignments for that aspect phrase.

[1]Avaiable at http://alt.qcri.org/semeval2016/task5/

Table I: Statistics of the SemEval Datasets

Dataset	#Reviews	#Sentences	#Aspect Phrases
English	350	2,000	644
Dutch	300	1,722	508
Russian	312	3,655	1,024
Spanish	627	2,070	543
Turkish	300	1,232	831
Total	1,889	10,679	3,550

2) Baseline: In order to evaluate our technique, we implemented the algorithm proposed by Zhai *et al.* [29] as baseline. This approach was chosen because it is the most seminal paper in aspect clustering field, it is simple, well detailed, and requires few resources in its implementation (Wordnet and lists of stopwords). Also, it can be applied to our datasets because it does not require semi-structured data or extra manual annotations in the reviews.

Since this technique was originally designed for monolingual aspect clustering, we had to make some adaptations for it to work with multilingual data. First, we translated the reviews originally written in Dutch, Russian, Spanish, and Turkish into English. We also removed the stopwords from the virtual documents in two occasions, before and after translation. Finally, we considered words with the same translations as if they were the same aspect phrases. For example, the words 'nagerecht', 'десерт', 'postre', and 'tatlı' are all grouped together in the same virtual documents of the aspect phrase 'dessert'. The remaining configurations are the same as in the original article.

3) Evaluation Metrics: As proposed by Zhai *et al.* [29], we measured the performance of our clustering algorithms in terms of Entropy and Purity. The goal is to simultaneously minimize entropy and maximize purity. In our evaluation, we consider a dataset DS, clustered into k disjoint sets $\{DS_1, DS_2, ..., DS_K\}$ and its respective golden partitions $G = \{g_1, g_2, ..., g_K\}$.

Purity: Purity intends to measure the largest portion of a dataset that contains data from a single golden partition *i.e.,* the highest percentage of correctly clustered points. It can be calculated as in Eq. 2, where $P_i(g_i)$ is the proportion of g_i data points in D_i. The purity of entire clusters is calculated according to Eq. 3.

$$purity(DS_i) = \max_j \ P_i(g_j) \qquad (2)$$

$$purity_{total} = \sum_{i=1}^{k} \frac{|DS_i|}{|DS|} purity(DS_i) \qquad (3)$$

Entropy: The entropy of a cluster, calculates as in Eq. 4, is the proportion of each gold partition present in it. The entropy of set of clusters is obtained with Eq. 5.

$$entropy(DS_i) = -\sum_{j=1}^{k} P_i(g_j) \log_2 P_i(g_j) \qquad (4)$$

$$entropy_{total} = \sum_{i=1}^{k} \frac{|DS_i|}{|DS|} entropy(DS_i) \qquad (5)$$

4) Multilingual Aspect Clustering Setup: The configuration setup for our proposal is as follows. FastText[2] was used for word embeddings. Their authors have made available the pre-trained multilingual word vectors for 157 languages trained on Wikipedia. We employed their models in to treat out of vocabulary words, which enriched our review representations. In order to align the FastText vectors, we used the transformation matrices by Smith *et al.* [49][3]. Gensim[4] was used for the word vector representations and for computing WMD. The centroid-based clustering algorithm used was k-means. This choice was motivated by its efficiency (*i.e.,* it does not require pairwise comparisons among all data items) and because it allows selecting number of clusters that should be generated.

We present two versions of our multilingual aspect clustering algorithm, with different centroid selection techniques. In MAC-RAND, we choose the centroids randomly, while in MAC-CENT we select as centroids the aspect phrases that appear more frequently in the datasets. In both configurations, the value of k was set to six. The tests were run in each language separately, and with all languages together. In our experiments with k-means, we ran the algorithm ten times and calculated the average for purity and entropy to mitigate the effects of variability.

B. Results

Tables II and III present the results of our method (MAC) and the baseline (L-EM). The results show that our technique outperforms the entropy values of the baseline (where smaller values mean better results). As for purity, the values are very close and we even surpass the baseline in some cases. It should be noted that our method is unsupervised, while the baseline relies on two phases of pre-processing before clustering the aspect phrases. Also, the baseline requires all the reviews to be in the same language.

Table IV shows some excerpts of the clusters generated by our algorithm to illustrate some strengths and weaknesses of our approach. Cluster #1, for example, shows that our method is able to group together aspect phrases that are synonyms (*waitress*, *waitstaff*, *servers*), or whose translation has the same meaning among languages (управляющий, manager, propietario). We also noticed that our approach is able to detect similar words with different spellings. This phenomenon is frequent in the Russian language, as can be seen in cluster #2, where we notice many word groups where this property can be validated (ресторaном, ресторане and ресторан, for example). This property is an improvement over the heuristic used in the baseline, which just groups aspect phrases that share equal words. Our approach can also reproduce the effects of this heuristic, an example of that is cluster #3, that groups the aspect phrases with the word "menu", which have the same form in English, Dutch, Spanish, and Turkish. Interestingly,

[2]Avaiable at https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md
[3]Avaiable at https://github.com/Babylonpartners/fastText_multilingual
[4]Avaiable at https://radimrehurek.com/gensim/

Table II: Experimental Results – ENTROPY

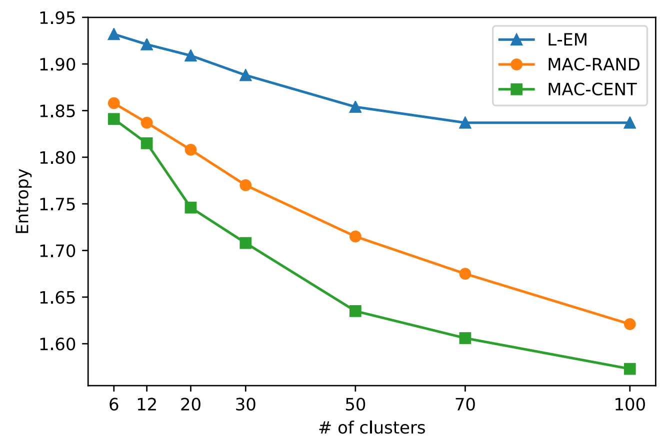
Method	English	Dutch	Russian	Spanish	Turkish	All
L-EM	1.748	1.753	1.974	1.591	2.076	1.932
MAC-RAND	1.654	1.801	1.814	**1.517**	**2.002**	1.858
MAC-CENT	**1.624**	**1.719**	**1.706**	1.540	2.039	**1.841**

Table III: Experimental Results – PURITY

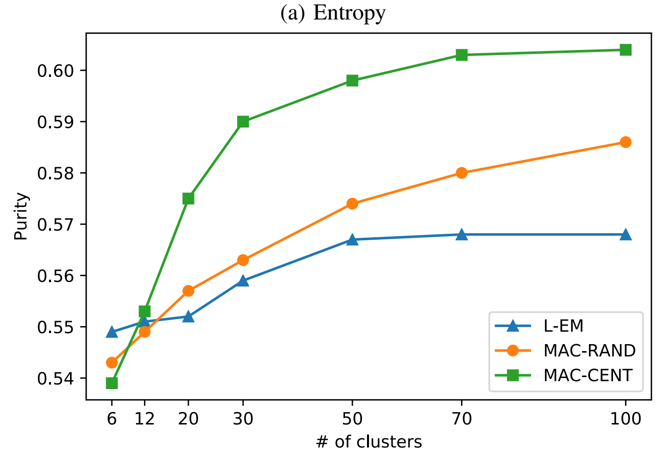
Method	English	Dutch	Russian	Spanish	Turkish	All
L-EM	0.598	**0.591**	0.503	**0.644**	0.484	**0.549**
MAC-RAND	0.605	0.559	0.540	**0.644**	0.487	0.543
MAC-CENT	**0.629**	0.576	**0.571**	0.581	**0.492**	0.539

our algorithm also includes in this cluster the aspect phrase меню, which is the translation of the word menu to Russian.

MAC is able to group semantically related aspect phrases. Two examples of this ability are shown in Table IV. Cluster #4 groups aspect phrases related to seafood (which can be seen as the aspect cluster of this group), while cluster #5 has aspect phrases related to artistic presentations. We achieved this thanks to an adequate representation of the virtual documents, allied to a similarity metric that is strong enough to capture the similarity between aspect phrases and their contexts.



(a) Entropy



(b) Purity

Figure 3: Experimental results with variation of k

Table IV: Excerpts of clusters generated by our algorithm

#	Centroid	Aspects
1	waitress	управляющий (manager) – waitstaff – hostess – manager – servers – gentleman – propietario (manager)
2	обслуживание (service)	сервис (service) – обслуживания (service) рестораном (restaurant) – ресторане (restaurant) – ресторан (restaurant) – место (place) – заведение (establishment) – заведению (establishment) – атмосфера (atmosphere) – атмосферу (atmosphere) – интерьера (interior) – интерьер (interior) официантов (waiters) – официанты (waiters) – официант (waiter) – официантка (waitress) – персонала (staff) – персонал (staff) музыка (music) – живая музыка (live music)
3	menu 'parels van india' (menu 'pearls of india')	menu kaart (menu card) – 3 gangen menu (3 course menu) – детское меню (children's menu) – блюд из меню (dishes from the menu) – sake menu – Menu de Primavera (Spring Menu) – menu fiyatları (menu prices)
4	scallops	scampi in de look – рыба в беконе (fish in bacon) – sea urchin – fried shrimp – lobster knuckles – oysters – stir fry blue crab – fried oysters and clams – pulpo con langostinos (octopus with prawns) – vieira con sopa (scallop with soup) – soya soslu somon (salmon with soy sauce)
5	грузинские танцевальный коллектив (Georgian dance group)	голос солистки (voice of the soloist) – концертик (concert) – песни (songs) – программа (program) – belly dancing show – müzik seçimleri (music selections) – las Posesas (theater play)

Despite the good results, our algorithm has some limitations – being unsupervised, it suffers from the drawbacks of this type of algorithm. Sometimes it is hard to guide the learning process in order to reach our clustering goal, which causes some aspect phrases to be misclassified. Another issue is to do with the integration of the multilingual datasets. We noticed that sometimes the clusters have only aspect phrases in one language. This is caused by the bias introduced in the normalization of the word embeddings phase. This can be seen when we make a comparison between the cosine similarity of a word and its translations into other languages. For example, the distance between the vector of the word 'dessert' and its translations десертам, postre, nagerecht and tatlı is 0.68, 0.59, 0.71, 0.48, respectively. This happens in cluster #2 in Table IV, which has only aspect phrases in Russian. Some of that aspect phrases are more related to other clusters instead of the cluster #2, for instance, музыка (music) and живая музыка (live music) are more related to cluster #5.

To assess how the method behaves with different number of clusters, we performed tests varying the parameter k. The results can be seen in Figure 3. When the number of clusters increases, our method showed a more pronounced drop in entropy and a gain in purity, compared to the baseline (Figure 3a). This happens because the heuristics of the baseline to label the data tends to get worse as the number of clusters increases. At the same time, our approach tends to select better initial centroids as the number of cluster increases. For a small number of clusters, our centroid selection technique did not work well, because the more frequent aspect phrases refer to the same aspect cluster. For example, it selects as centroids "service", "обслуживание музыка" and "servicio", which are the translation of service into Russian and Spanish. With a larger number of clusters, the centroid selection algorithm tends to select more diversified aspect phrases.

VI. Conclusion

In this work, we proposed an unsupervised approach to address the problem of Multilingual Aspect Clustering. Our solution combines unsupervised clustering, syntactic term similarity, and word embeddings.

We ran experiments on restaurant reviews written in English, Spanish, Russian, Dutch, and Turkish and compared our performance against an established baseline. The results show that our unsupervised clustering technique outperforms a semi-supervised baseline in many cases.

Future work will include a study of means to improve word embedding normalization and to prune irrelevant aspect phrases. Another important feature will be the development of heuristics that can be used with multilingual data and that will allow us to use semi-supervised approaches in multilingual aspect clustering. Finally, we wish to build a visualization tool that summarizes the results of aspect clustering.

Acknowledgment: This work was partially supported by CNPq/Brazil and by CAPES Finance Code 001.

References

[1] B. Liu, "Opinion mining and sentiment analysis," in *Web data mining: exploring hyperlinks, contents, and usage data*, 2nd ed. Springer Science & Business Media, 2011, ch. 11.

[2] L. Zhang and B. Liu, "Aspect and entity extraction for opinion mining," in *Data mining and knowledge discovery for big data*. Springer, 2014, pp. 1–40.

[3] C. Banea, R. Mihalcea, and J. Wiebe, "Multilingual subjectivity: are more languages better?" in *International Conference on Computational Linguistics*, 2010, pp. 28–36.

[4] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, Apr. 2013.

[5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[6] A. Radford, R. Józefowicz, and I. Sutskever, "Learning to generate reviews and discovering sentiment," *CoRR*, vol. abs/1704.01444, 2017. [Online]. Available: http://arxiv.org/abs/1704.01444

[7] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, and Q. Zhou, "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognitive Computation*, vol. 8, no. 4, pp. 757–771, Aug 2016.

[8] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, "Multilingual sentiment analysis: from formal to informal and scarce resource languages," *Artificial Intelligence Review*, vol. 48, no. 4, pp. 499–527, Dec 2017.

[9] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[10] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 168–177.

[11] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.

[12] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh, "A rule-based approach to aspect extraction from product reviews," in *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, 2014, pp. 28–37.

[13] W. Jin, H. H. Ho, and R. K. Srihari, "A novel lexicalized hmm-based learning framework for web opinion mining," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 465–472.

[14] N. Jakob and I. Gurevych, "Extracting opinion targets in a single-and cross-domain setting with conditional random fields," in *Proceedings of the 2010 conference on empirical methods in natural language processing*, 2010, pp. 1035–1045.

[15] S. Moghaddam and M. Ester, "Ilda: interdependent lda model for learning latent aspects and their ratings from online product reviews," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 665–674.

[16] Z. Toh and J. Su, "Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 496–501.

[17] ——, "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)*, 2016, pp. 282–288.

[18] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 616–626.

[19] ——, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms." in *AAAI*, 2017, pp. 3316–3322.

[20] A. Giannakopoulos, C. Musat, A. Hossmann, and M. Baeriswyl, "Unsupervised aspect term extraction with b-lstm & crf using automatically labelled datasets," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017, pp. 180–188.

[21] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "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)*, 2017, pp. 388–397.

[22] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, A.-S. Mohammad, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016, pp. 19–30.

[23] A. García-Pablos, M. Cuadros, and G. Rigau, "W2vlda: almost unsupervised system for aspect based sentiment analysis," *Expert Systems with Applications*, vol. 91, pp. 127–137, 2018.

[24] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing aspect-based sentiment analysis of arabic hotels' reviews using morphological, syntactic and semantic features," *Information Processing & Management*, 2018.

[25] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 342–351.

[26] G. Carenini, R. T. Ng, and E. Zwart, "Extracting knowledge from evaluative text," in *Proceedings of the 3rd international conference on Knowledge capture*, 2005, pp. 11–18.

[27] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, "Product feature categorization with multilevel latent semantic association," in *Proceedings of the 18th ACM conference on Information and knowledge management*, 2009, pp. 1087–1096.

[28] Z. Zhai, B. Liu, H. Xu, and P. Jia, "Constrained lda for grouping product features in opinion mining," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2011, pp. 448–459.

[29] ——, "Clustering product features for opinion mining," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, ser. WSDM '11. ACM, 2011, pp. 347–354.

[30] T. Wang, Y. Cai, G. Zhang, Y. Liu, J. Chen, and H. Min, "Product feature summarization by incorporating domain information," in *International Conference on Database Systems for Advanced Applications*, 2013, pp. 231–243.

[31] L. Zhao, M. Huang, H. Chen, J. Cheng, and X. Zhu, "Clustering aspect-related phrases by leveraging sentiment distribution consistency," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1614–1623.

[32] Y. Zhang, M. Liu, and H.-X. Xia, "Clustering context-dependent opinion target words in chinese product reviews," *Journal of Computer Science and Technology*, vol. 30, no. 5, pp. 1109–1119, 2015.

[33] L. Liu, Z. Lv, and H. Wang, "Opinion mining based on feature-level," in *Image and Signal Processing (CISP), 2012 5th International Congress on*, 2012, pp. 1596–1600.

[34] S. Xiong and D. Ji, "Exploiting flexible-constrained k-means clustering with word embedding for aspect-phrase grouping," *Information Sciences*, vol. 367-368, pp. 689 – 699, 2016.

[35] S. Huang, Z. Niu, and Y. Shi, "Product features categorization using constrained spectral clustering," in *Natural Language Processing and Information Systems*, E. Métais, F. Meziane, M. Saraee, V. Sugumaran, and S. Vadera, Eds., 2013, pp. 285–290.

[36] J. Pavlopoulos and I. Androutsopoulos, "Multi-granular aspect aggregation in aspect-based sentiment analysis," in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, pp. 78–87.

[37] Y. Zhao, B. Qin, and T. Liu, "Clustering product aspects using two effective aspect relations for opinion mining," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, 2014, pp. 120–130.

[38] Y. He, J. Song, Y. Nan, and G. Fu, "Clustering chinese product features with multilevel similarity," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, 2015, pp. 347–355.

[39] Y. Cao, M. Huang, and X. Zhu, "Clustering sentiment phrases in product reviews by constrained co-clustering," in *Natural Language Processing and Chinese Computing*, 2015, pp. 79–89.

[40] W. Jiajia, L. Yezheng, J. Yuanchun, S. Chunhua, S. Jianshan, and D. Yanan, "Clustering product features of online reviews based on nonnegative matrix tri-factorizations," in *Data Science in Cyberspace (DSC), IEEE International Conference on*, 2016, pp. 199–208.

[41] I. Flaounas, O. Ali, M. Turchi, T. Snowsill, F. Nicart, T. De Bie, and N. Cristianini, "Noam: news outlets analysis and monitoring system," in *Proceedings of the 2011 ACM SIGMOD international conference on Management of data*, 2011, pp. 1275–1278.

[42] J. Van Gael and X. Zhu, "Correlation clustering for crosslingual link detection." in *IJCAI*, 2007, pp. 1744–1749.

[43] A. Rauber, M. Dittenbach, and D. Merkl, "Towards automatic content-based organization of multilingual digital libraries: An english, french, and german view of the russian information agency novosti news," in *Third All-Russian Conference Digital Libraries: Advanced Methods and Technologies*, 2001.

[44] C. Denicia-Carral, M. Montes-Gómez, L. Villaseñor-Pineda, and R. M. Aceves-Pérez, "Bilingual document clustering using translation-independent features," in *Proceedings of CICLing*, vol. 10, 2010.

[45] S. Montalvo, V. Fresno, and R. Martínez, "Nesm: A named entity based proximity measure for multilingual news clustering," *Procesamiento del lenguaje natural*, no. 48, 2012.

[46] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., 2013, pp. 3111–3119.

[47] S. Ruder, "A survey of cross-lingual embedding models," *CoRR*, vol. abs/1706.04902, 2017. [Online]. Available: http://arxiv.org/abs/1706.04902

[48] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *International Conference on Machine Learning*, 2015, pp. 957–966.

[49] S. L. Smith, D. H. P. Turban, S. Hamblin, and N. Y. Hammerla, "Offline bilingual word vectors, orthogonal transformations and the inverted softmax," *CoRR*, vol. abs/1702.03859, 2017. [Online]. Available: http://arxiv.org/abs/1702.03859