Knowledge-Based Systems 192 (2020) 105339

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Multilingual aspect clustering for sentiment analysis ☆

Lucas Rafael Costella Pessutto, Danny Suarez Vargas, Viviane P. Moreira *

Institute of Informatics, Universidade Federal do Rio Grande do Sul, Caixa Postal 15.064, 91.501-970, Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 14 January 2019
Received in revised form 3 December 2019
Accepted 4 December 2019
Available online 9 December 2019

Keywords:
Aspect-based sentiment analysis
Multilingual aspect clustering
Unsupervised learning
Word embeddings

ABSTRACT

In the last few years, there has been growing interest in aspect-based sentiment analysis, which deals with extracting, clustering, and rating 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. Hence, aspect clustering becomes necessary. Current solutions for aspect clustering are monolingual, but in many practical situations, reviews for a given entity are available in several languages, calling for multilingual integration. In this article, we address the novel task of multilingual aspect clustering, which aims at grouping semantically related aspects extracted from reviews written in several languages. Our method is unsupervised and relies on the contextual information of the aspects, which is represented by word embeddings. This representation allied with a suitable similarity measure allows clustering related aspects. Our experiments on two datasets with five languages each showed that our unsupervised clustering technique achieves results that outperform monolingual baselines adapted to work with multilingual data. We also show the benefits of the multilingual approach compared to using languages in isolation.

1. Introduction

The dawn of the Web 2.0 changed the way users interact, 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 sold. This information becomes useful to other users as they can rely on previous shopping experiences from others as a basis for their own purchases. Companies can also take advantage of these 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. Sentiment Analysis is the field of study which aims at processing the information conveyed by unstructured texts, providing structured information that facilitates the understanding of opinions, attitudes, or emotions towards a particular entity [1,2].

Sentiment Analysis can be performed at different levels of granularity (entire review, sentence, or aspect). In spite of being largely explored in the literature [3–5], sentence and review levels are less informative than the aspect level. Aspect-Based Sentiment Analysis (ABSA) aims to identify and rate the features (or aspects) of the entity being evaluated.

Typically, ABSA involves the following phases: (*i*) identify and extract entities in reviews; (*ii*) identify and extract the aspects of an entity; (*iii*) cluster similar aspects; and (*iv*) determine the polarity of the sentiment over the entities and the aspects. The focus of our work is on the third phase, *i.e., aspect clustering*.

Aspect clustering is necessary because even state-of-the-art aspect extraction techniques 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. Aspect clustering groups the different words that refer to the same feature. This is a fundamental step to build 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 AirBnB, Amazon, and TripAdvisor (which receives almost 400 million visitors a month worldwide [6]). As a consequence, dealing with multilingual data becomes necessary. In this scenario, the distribution of reviews across languages is unbalanced — some languages may have too few reviews, lacking the required volume of data to allow for sentiment-analysis 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 [7].

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.knosys.2019.105339.

* Corresponding author.
E-mail addresses: lrcpessutto@inf.ufrgs.br (L.R.C. Pessutto), dsvargas@inf.ufrgs.br (D.S. Vargas), viviane@inf.ufrgs.br (V.P. Moreira).

https://doi.org/10.1016/j.knosys.2019.105339
0950-7051/

The focus of this work is on *multilingual aspect clustering*, which can be defined as *the task of grouping aspects referring to the same feature across multiple languages*. An example of how multilingual aspect clustering can be used for Sentiment Analysis is shown in Fig. 1. Initially, the set of reviews is composed of reviews in three languages (English, Spanish, and Dutch) with the aspects already extracted. The next step is to build representations for the aspects using contextual information. Then, a clustering algorithm is applied in order to group semantically similar aspects across languages. 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 (shown as a bar chart) that synthesize the sentiment expressed by the reviewers across many languages. If one language does not have reviews that cover a given aspect, the opinion can be extracted from the other languages. For example, in Fig. 1, there are no reviews for the Restaurant aspect in German, but this gap was filled by the availability of reviews in other languages.

Existing solutions for aspect clustering are monolingual, *i.e.,* even if they can be applied to different languages, they work with each language separately. Our solution combines unsupervised clustering, semantic term similarity, and word embeddings. We carried out experiments on two datasets. The first contains restaurant reviews written in English, Spanish, Russian, Dutch, and Turkish. And the second dataset contains reviews on digital cameras written in English, German, Italian, Spanish, and French. Since multilingual aspect clustering has not been addressed before, we had to adapt two baselines [8,9] using machine translation so that they were able to work with multilingual data. The results show that our proposed clustering technique outperforms both baselines. Furthermore, we show that multilingual clustering is advantageous compared to using each languages in isolation.

A summarized version of this article was presented in Pessutto et al. [10]. Here, we extend it by adding another clustering algorithm and similarity metric. These new explored alternatives yielded the best results in our experimental evaluation. Also, we added another baseline and a new dataset to our experiments. New experiments were done to assess the benefits of a multilingual approach as opposed to using languages in isolation. Finally, in this article, we present a broader coverage of the related literature.

The remainder of this article is structured as follows. Section 2 presents background knowledge that supports our multilingual aspect clustering algorithm. In Section 3, related work on aspect-based sentiment analysis, monolingual aspect clustering, and multilingual document clustering are discussed. In Section 4, our proposed approach to the Multilingual Aspect Clustering problem is presented. Section 5 describes the experimental evaluation, presents our results and a discussion. Finally, Section 7 concludes this article, summarizing our contributions and identifying possibilities for future work.

2. Background

This section covers the background concepts which served as the basis for this article. We address vector representation of words, medoid-based clustering algorithms, and distance measures.

2.1. Vector representation of words

The goal of representing words in a vector space is to map the semantic similarity between them. The techniques employed for this task are based on the hypothesis that words with the same meaning are used in similar contexts along with documents.

Research aiming at dimensionality reduction of word vectors resulted in the development of *word embeddings* — a language model which represents words as low dimensional vectors, keeping the distributional similarity between them [11]. Word embeddings are able to map linguistic regularities present in documents, *e.g.,* gender, verb tenses, and even the relationship between countries and capitals are maintained in the embedding space [12]. Multilingual word embeddings cater for situations which require that words in different languages share the same vector space. In order to achieve that, some approaches learn a matrix capable of performing the linear transformation of monolingual word embeddings in language l_1 into the word embeddings of language l_2. These approaches usually rely on bilingual dictionaries or other translation means to map pairs of words across different languages. It is also possible to train multilingual word embeddings from parallel or comparable sources [13].

In this work, we take advantage of the semantic power provided by word embeddings in order to represent the aspects and the contextual information. We employ multilingual word embeddings in that representation, which allows us to represent aspects in different languages in the same vector space.

2.2. Medoid-based clustering algorithms

Clustering algorithms group data points such that data within the same cluster have high similarity compared to data belonging to different clusters [14]. Medoid-Based Clustering Algorithms assume that data points are distributed in a Euclidean space, so each cluster can be represented by its medoid, which is usually the most representative data point in a cluster. This approach differs from centroid-based algorithms which use the mean rather than using the most central data point. One of the most widely used clustering algorithms in this category is k-medoids, in which k is the sole parameter that specifies the number of desired clusters.

The *Bisecting k-medoids* algorithm is a variation to work in cases in which the number of clusters (k) is unknown. Rather than giving k as input, the user informs a threshold (s) that represents the maximum number of elements allowed in each cluster.

2.3. Distance/similarity measures

The use of a suitable distance/similarity measure is key for the good performance of clustering algorithms. In addition to correctly expressing the distance between the data points, these measures need to guarantee the convergence of the k-medoids algorithm. Euclidean, Manhattan, Jaccard, and Cosine are widely used for this purpose. The first two measures are applied in Euclidean Spaces while Jaccard and Cosine are more suitable for text documents.

Our data points are documents represented as a set of word embeddings. Thus, in order to compute the cosine between sets of vectors, one can simply take the mean of each set of vectors and then compute their cosine. The cosine between two documents $A = \{a_1, a_2, \ldots, a_p\}$ and $B = \{b_1, b_2, \ldots, b_q\}$ (where each a in A and b in B are a m-dimensional word embedding representation of the words in the documents) and considering $\overline{A} = \frac{1}{|A|}\sum_{i=1}^{|A|} a_i$ and $\overline{B} = \frac{1}{|B|}\sum_{j=1}^{|B|} b_j$ the mean vectors of A and B, the cosine is given by Eq. (1).

$$cos(\overline{A}, \overline{B}) = \frac{\sum_{i=1}^{m} \overline{a}_i \cdot \overline{b}_i}{\sqrt{\sum_{i=1}^{m} \overline{a}_i^2 \cdot \sum_{i=1}^{m} \overline{b}_i^2}} \tag{1}$$

In this work, we also consider the *Word Mover's Distance* (WMD) [15], which was specifically designed to work with word

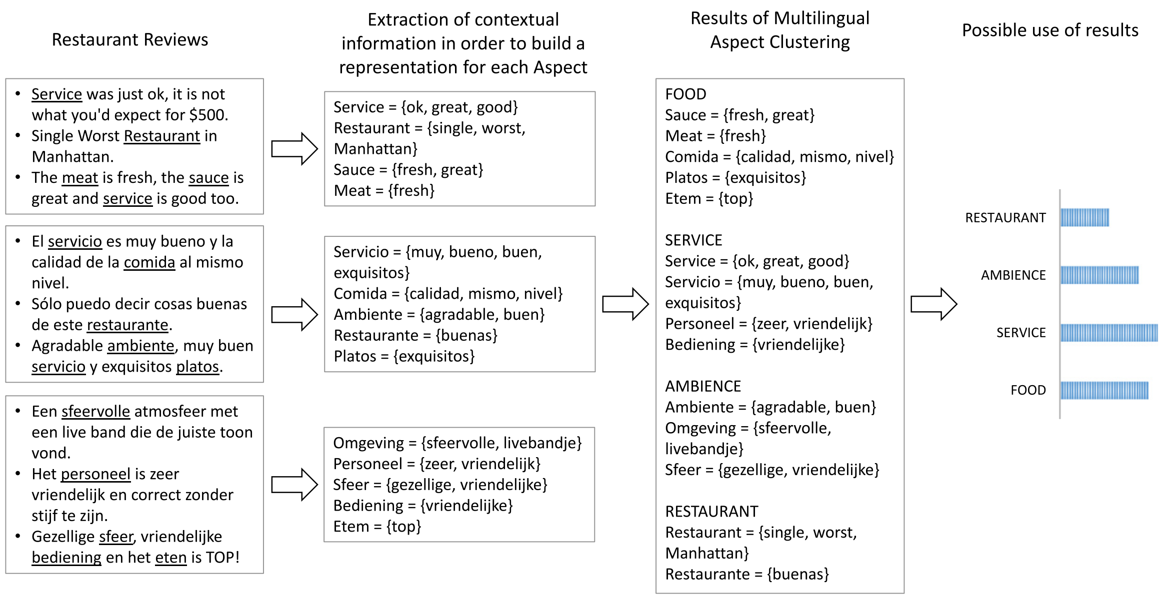


Fig. 1. An example of multilingual aspect clustering.

embeddings. WMD is a special case of the well known Earth's Mover Distance, which measures the distance between two probability distributions over a region. WMD takes advantage of the semantic relationships present in word embeddings. The goal of WMD is to measure the minimum traveling cost of the word embeddings in one document to 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 well 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*.

To compute WMD, it is necessary to represent the documents as normalized bag-of-words (nBoW), *i.e.*, a bag-of-words model with the addition of weights. Consider that a word i appears c_i times in a document. We can compute its weight (d_i) as in Eq. (2).

$$d_i = \frac{c_i}{\sum_{j=1}^{n} c_j} \tag{2}$$

The distance between documents d and d' is the minimum cumulative cost required to transform all words in d into the words in d'. To calculate it, one needs to build a flow matrix (T), where $T_{i,j}$ represents how much of a word i in d travels to word j in d'. Eq. (3) shows how such a cost is calculated.

$$WMD(d, d') = \sum_{i,j} T_{i,j} c(i, j) \tag{3}$$

3. Related work

In this section, we review the related literature about three key topics: aspect-based sentiment analysis, (monolingual) aspect clustering, and multilingual document clustering. Then we discuss the research gap that we address in this article.

3.1. Aspect-based sentiment analysis

The aspect level appears as the most important in Sentiment Analysis, mainly due to the relevant information that it conveys [16]. In this level of analysis, often called *Aspect-Based Sentiment Analysis*, the aspects and entities are identified in natural language texts. Solutions for aspect extraction can be classified into three main groups, according to the underlying approach [17]: (*i*) based on language rules [18–20], (*ii*) based on sequence labeling models [21,22], and (*iii*) based on topic models [23]. However, other works do not fit in only one of these categories as they combine resources from more than one approach [24,25]. Furthermore, state-of-the-art approaches rely on more sophisticated architectures like recurrent neural networks such as LSTM, Bi-LSTM, Neural Attention Models, and Convolutional Neural Networks [9,26–29].

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 released a multilingual dataset for ABSA. However, none of the participating teams used the dataset in a multilingual way – *i.e.*, by combining all languages together.

3.2. Clustering monolingual aspects

After performing aspect extraction, clustering aspects is necessary to group the different 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 Dictionaries [30] or Taxonomies [31]. Dictionaries are not usually 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 terms do not typically appear in dictionaries. Taxonomies have the disadvantage of being domain-dependent, difficult to build and maintain. Therefore, these approaches have fallen out of favor.

Topic modeling techniques: Solutions in this category employ techniques such as Latent Semantic Analysis and Latent Dirichlet Allocation, in order to group similar aspects taking into account the semantic similarity between aspects [32,33]. This set of techniques tends to perform poorly because they depend solely on probabilistic models based on co-occurrence frequency, which is not enough to identify the semantic similarity between the aspects.

Semi-supervised algorithms are widely used for aspect clustering. They consist in labeling part of the input data with the cluster information, to enable the formation of new (improved) clusters. In most cases, this information is obtained automatically from the data. The seminal work in this category is by Zhai et al. [8], which works as follows: (*i*) the aspects are grouped according to the words they share; (*ii*) the lexical similarity between the groups is measured, and the n most similar groups are clustered — the contextual information about the aspects is taken into account by considering a window of $[-t, t]$ words around the aspect phrases; and (*iii*) once part of the data is labeled, the EM algorithm based on a Naïve Bayes classifier clusters the aspects. This approach has been tested on a number of datasets and is still used as a baseline by new techniques.

Unsupervised algorithms are desirable as they do not depend on annotated data. The work by He et al. [9], known as ABAE, falls into this category. It consists of a neural network with an attention mechanism to quantify the importance of each word in the review sentences. The importance is useful to determine whether the word is an aspect. ABAE outperformed methods based on topic modeling and on k-means clustering.

Other clustering approaches: There are also a few proposed solutions which do not fit into the categories discussed above. For example, some works [34–36] use hierarchical clustering to produce multi-granular summaries, which can be customized according to the user's needs. Aspects and opinion words were clustered simultaneously by using a constrained hidden Markov random field model [37]. Factorization of feature-opinion matrices has also been applied for aspect clustering [38]. Finally, linguistic resources were used to provide relations between aspects [39].

3.3. Multilingual document clustering

The task of Multilingual Document Clustering has the goal of grouping documents written in more than one language according to their subjects. Multilingual Document Clustering is done in two steps. First, the documents in the collection are represented in a language-independent manner, and then the groups are formed based on document representations [40].

Most techniques developed for monolingual document clustering rely on the co-occurrence of words in documents. Thus, they are not suitable for the multilingual setting since the documents in different languages do not share the same vocabulary.

We can divide the approaches used to perform multilingual document clustering into two groups, based on how they represent documents in a feature space. (*i*) *Monolingual Feature Space Techniques* aim to create a monolingual feature space of documents in order to cluster multilingual documents. Some use machine translation systems to translate entire documents [41] or just chosen document features [42,43], while others can rely on multilingual resources like dictionaries [44,45] or thesauri [46] to create the monolingual feature space. Once this feature space is created, monolingual document clustering techniques can be applied. (*ii*) *Multilingual Feature Space Techniques* attempt to map all the documents in a shared language-independent space or to extract language-independent features from the multilingual documents. They can also combine both strategies to represent the document collection [47]. Some solutions have employed syntactic similarity [48,49]. However, these are inapplicable if the languages have completely different alphabets. Alternatives include the use of a comparable corpus [50] or topic modeling techniques [51].

3.4. Research gap

Table 1 shows the 24 surveyed related works in ABSA describing the task they address (aspect extraction or aspect clustering) and the languages they were experimented on (monolingual or multilingual). We can see that *all* existing solutions, both for aspect extraction and for clustering, are *monolingual* – *i.e.,* even if they were tested on different languages, they work with each language separately. Thus, our work is the first to address these tasks in a multilingual setting.

As expected, English (EN) is the language in which most approaches were tested, followed by Chinese (CN). Techniques that rely on the co-occurrence of context words cannot be directly applied to multilingual aspect clustering because when the reviews are in many languages, the intersection between vocabularies is (almost) empty. As discussed in Section 3.2, dictionaries have limitations as they typically do not contain expressions and lack coverage for domain-specific data.

We also investigated the existing techniques that aim to cluster multilingual documents (in Section 3.3). However, the focus of multilingual document clustering has been to group news articles. Review texts differ from news in many ways, especially regarding the size of the documents and the features considered in the clustering process. To identify groups of news, the most important features tend to be nouns, noun phrases, and named entities. Techniques for multilingual document clustering that extract multilingual features usually rely on dictionaries or comparable corpora in order to represent documents. This kind of resource is not available for most language-pairs and domains, limiting its applicability.

Despite Multilingual Sentiment Analysis having been an established research topic for a few years, the specific topic that we address in this article is novel. Existing solutions for Multilingual Sentiment Analysis have focused on extending sentiment analysis to a wider number of languages, either by generating language resources [52], or by showing how labels can be mapped across languages [53–55]. The most widely used approach is to apply machine translation [7,56]. Notice that these works addressed polarity classification, and not aspect extraction or clustering.

In summary, we can see that multilingual aspect clustering remained an unexplored research topic. Thus, to the best of our knowledge, we are the first to propose a solution for it.

4. Multilingual aspect clustering

This section presents our approach to addressing the problem of Multilingual Aspect Clustering. We combine techniques inspired by multilingual document clustering (Section 3.3) and monolingual aspect clustering (Section 3.2) to address multilingual aspect clustering. Our proposed approach leverages the contextual information of aspect phrases and word embeddings in an unsupervised clustering algorithm which groups multilingual aspect phrases.

4.1. Problem definition and solution overview

In the context of multilingual aspect clustering, the set of input reviews can be defined as $R = \{R_1, R_2, \ldots, R_l\}$, where R_l corresponds to the subset of reviews in language l, with $l \geq 2$. 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 reviews $R_l = \{r_l^1, r_l^2, \ldots, r_l^m\}$, where r_l^m denotes the mth review in language l.

For each subset of reviews (R_l), the aspects need to be extracted. The explicit representation of an aspect that occurs in a set of reviews is called *aspect phrase* (also called *product feature*

Table 1
Comparison of the characteristics of the works on ABSA.

Work	ABSA task		Language support	
	Extraction	Clustering	Monolingual	Multilingual
Hu & Liu [18]	✓		EN	
Qiu et al. [19]	✓		EN	
Poria et al. [20]	✓		EN	
Jin et al. [21]	✓		EN	
Jakob et al. [22]	✓		EN	
Moghaddam & Ester [23]	✓	✓	EN	
Toh & Su [24,25]	✓	✓	EN	
Wang et al. [26,27]	✓		EN	
Giannakopoulos et al. [28]	✓		EN	
He et al. [9]	✓		EN	
Poria et al. [29]	✓		EN	
Liu et al. [30]	✓	✓	EN	
Carenini et al. [31]	✓	✓	EN	
Guo et al. [32]	✓	✓	EN, CN	
Zhai et al. [8,33]		✓	EN	
Pavlopoulos & Androutsopoulos [34]		✓	EN	
Zhao et al. [35]		✓	CN	
He et al. [36]		✓	CN	
Cao et al. [37]		✓	CN	
Jiajia et al. [38]		✓	EN	
Vargas & Pardo [39]		✓	PT	
MAC/BMAC		✓		EN, NL, ES, TR, RU EN, DE, ES, IT, FR

or *surface form* in the literature). An aspect phrase is composed of one or more terms (*e.g., battery, battery life*), hence the need to use the term *phrase*. The set of aspect phrases (AP) is the union of all q sets of aspects in the reviews. More formally, $AP = \{AP_1, AP_2, \ldots, AP_q\}$. A set of aspect phrases (AP) in multiple languages is called an *aspect cluster* (AC).

The problem of multilingual aspect clustering can be defined as the task of mapping the aspects in AP into the AC set, where $AC = \{c_1, c_2, \ldots, c_k\}$. Each subset $c_i \in AC$ contains several aspect phrases referring to the same aspect cluster, where k is the total number of aspect groups. Essentially, multilingual aspect clustering is a function $f : AP \rightarrow AC$. Mapping the set AP into the set AC involves creating a representation for each aspect phrase in the dataset (which may be language dependent). Next, it is necessary to normalize the representation of the aspect phrases into a language independent representation, in order to deal with multilinguism. Thus, the function f can be decomposed into three other functions – $f_1 : AP \rightarrow AR$, $f_2 : AR \rightarrow LIR$, and $f_3 : LIR \rightarrow AC$, where AR is the language dependent representation of an aspect phrase, and LIR consists of a language independent representation of the aspects.

Two important properties of the AC set should be highlighted: (i) the union of all subsets c_i in AC results in the set of aspect phrases AP, meaning that all aspects have to be assigned to a group; and (ii) the intersection of the subsets $c_i \in AC$ is an empty set since every aspect phrase belongs to just one aspect group.

Our proposed solution for multilingual aspect clustering relies on a set of steps which are summarized in Fig. 2. The first step is to extract the aspect phrases from the reviews. In this work, we do not propose new approaches for aspect extraction since our focus is on aspect clustering. As seen in Section 3.1, there are a number of techniques for aspect extraction that could be used to accomplish this task.

Next, our solution needs to represent the context of the aspect phrases, *i.e.,* function f_1. Context information is crucial since aspect phrases which appear in similar contexts are likely to be semantically related. We refer to this representation as *virtual document* [8]. The context is represented by the surrounding 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. At the end of the virtual document creation phase, there are l sets of aspect phrases and their respective documents, where l is the total number of languages in the dataset. For example, in the sentence *"The service is amazing and the ambience is good for a date"*, with a window size of $t = 5$, the virtual document for the aspect phrase *service* is composed of the word *{amazing}*, and the virtual document for the aspect phrase *ambience* is *{amazing, good, date}*. Note that we removed the stopwords *the, is, and, for*, and *a* after the virtual documents are constructed. Any aspect phrases that co-occur in the same sentence are also removed from the virtual document, *i.e., service* will not be in the virtual document for *ambience* and vice-versa. In this work, we set the value of t as ten.

Then, in order to group the aspect phrases, we need a common, language-independent representation, *i.e.,* function f_2. We employ *multilingual word embeddings* to this task. Since our documents consist of 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 [13]. The goal is to obtain a shared representation of the semantics of words in different languages. Furthermore, multilingual word embeddings enable knowledge transfer from languages with rich resources to the ones with scarce resources.

In this work, monolingual word embeddings were normalized so that all languages could share the same vector space. We used normalization matrices capable of performing a linear transformation of monolingual word embeddings in language l_1 in the word embeddings of a language l_2 [57]. We can see the multilingual word embeddings as function that transforms a text into a vector space, there is a function for each language, and the vector space is common for all languages.

The *Document Embedding* is a set formed by the concatenation of the word embedding representations of each word in the virtual document and the word embeddings of each word in the aspect phrase. In the previous example, the Document Embedding of the aspect phrase *service* is the word embedding representation of the words *amazing* and *service*. For the aspect phrase *ambience*, its Document Embedding is *{amazing, good, date, ambience}*.

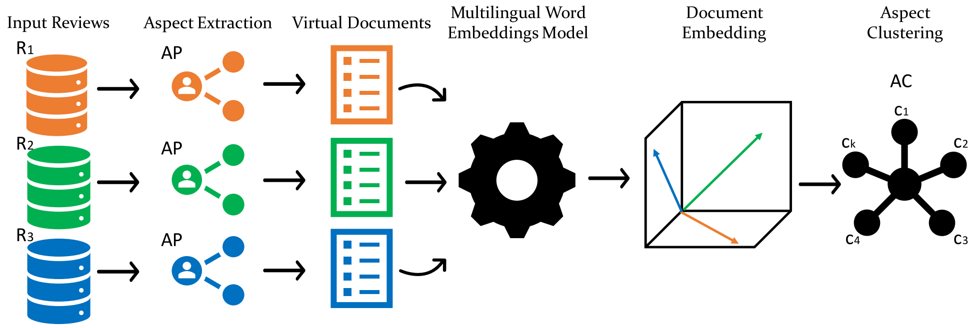


Fig. 2. The proposed approach for multilingual aspect clustering.

Finally, when all the aspect phrases in the dataset are represented in the same space, clustering algorithms are used to group the aspect phrases that belong to the same category, *i.e.,* function f_3. We do not rely on semi-supervised algorithms because we work with multilingual data. Multilingual reviews cannot be handled by labeling techniques proposed for semi-supervised approaches for monolingual aspect clustering as they focus on the lexical similarity between aspects. Another important fact that makes us opt for an unsupervised approach is that we do not want to rely on translating the reviews, neither on the manual labeling of the aspect phrases (which is a difficult task to perform with multilingual data). The goal is to make our approach as simple as possible. Unsupervised approaches are a good choice as they allow building a domain-independent solution for the multilingual aspect clustering problem. This task can be categorized as unsupervised learning because we only use the document embeddings as input for our clustering algorithm.

4.2. Clustering algorithms

This section shows how we employ medoid-based clustering algorithms in order to address the multilingual aspect clustering task. We propose two alternatives of clustering algorithms — one using the standard k-medoids and another using the bisecting k-medoids.

`MAC` **– Multilingual Aspect Clustering with k-medoids** is detailed in Algorithm 1. This choice for k-medoids was motivated by its efficiency (*i.e.,* it does not require pairwise comparisons among all data items). The inputs 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 medoids (lines 3–5). This selection can be made randomly, or by using a heuristic as discussed in Section 4.2.2.

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

The number of expected clusters (k) depends essentially on the characteristics of the dataset and the goals of the analysis. By increasing k, we obtain a finer granularity which may be desirable in some settings. In Section 6, we assess how different values of k impact our clustering quality metrics.

`BMAC` **– Multilingual Aspect Clustering with Bisecting k-medoids** aims at addressing the limitation of standard k-medoids which requires the number of desired clusters as input. In many real-life situations, this parameter may be unknown. Thus, in this Section, we propose an alternative approach which uses the *Bisecting k-medoids.*

ALGORITHM 1: Multilingual Aspect Clustering (`MAC`)

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

1 $AC \leftarrow \emptyset$
2 $medoids \leftarrow \emptyset$
3 **for** $i \leftarrow 0$ **to** k **do**
4 | $medoids \leftarrow$ med_selection(DE)
5 **end**
6 **while** *not convergence* **do**
7 | **for** $i \leftarrow 0$ **to** $|DE|$ **do**
8 | | **for** $j \leftarrow 0$ **to** $|medoids|$ **do**
9 | | | $sim \leftarrow$ similarity(DE[i], medoids[j])
10 | | **end**
11 | | $c_{de}[i] \leftarrow$ cluster with the highest similarity
12 | **end**
13 | **for** $j \leftarrow 0$ **to** $|medoids|$ **do**
14 | | medoids[i] $\leftarrow$ Result of Eq. (4)
15 | **end**
16 **end**
17 **return** AC

With the bisecting k-medoids, instead of informing the desired number of clusters, one informs a threshold value (s), which is the maximum number of aspect phrases that are allowed in each cluster. Algorithm 2 shows the pseudocode for Bisecting Multilingual Aspect Clustering (`BMAC`).

The `BMAC` alternative starts applying our `MAC` algorithm into all document embeddings, with $k = 2$. Once the data is clustered, our algorithm chooses the largest cluster and splits it into two other clusters. We repeat this division (*i.e.,* bisection) process until the size of the clusters is smaller than the threshold parameter.

ALGORITHM 2: Bisecting Multilingual Aspect Clustering (`BMAC`)

Input : s – threshold that specifies the maximum number of elements in a cluster
$DE = \{de_1, de_2, \ldots, de_i\}$ – Document Embeddings
Output: $AC = \{c_{de_1}, c_{de_2}, \ldots, c_{de_i}\}$ – Set of Aspect Clusters

$AC \leftarrow \emptyset$
$data \leftarrow DE$
repeat
| $AC \leftarrow$ MAC(k = 2, DE = $data$)
| $c \leftarrow$ cluster with maximum number of data points
| $max \leftarrow$ length of cluster c
| $data \leftarrow$ data points of cluster c
until $max \leq s$
return AC

4.2.1. Distance measures

We used two distance measures to compare document embeddings that were discussed in Section 2.3 – cosine (`COS`) and

Table 2
Statistics of the datasets

SemEval dataset			
Language	#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
Camera dataset			
Language	#Reviews	#Sentences	#Aspect phrases
English	5,720	17,717	115
German	2,310	7,029	112
Italian	2,760	6,605	89
Spanish	2,060	5,183	106
French	1,531	3,397	111
Total	14,381	39,931	533

Word Mover's Distance (`WMD`). The goal was to assess which was the most suitable to generate good clusters, capturing the semantic similarity of the document embeddings.

4.2.2. Medoid selection

We propose two different medoid selection techniques. In the first alternative, `RAND`, the medoids are randomly chosen. In the second, `CENT`, we selected as medoids the aspect phrases that appear more frequently in the dataset. At each iteration, the new medoid is the one that has the highest similarity (or smallest distance) in relation to the other document embeddings belonging to that cluster. Eq. (4) shows how the medoid of a cluster c is chosen.

$$medoid(c) = \max_i \sum_{i=0}^{|c|} \frac{1}{|c|} \sum_{j=0}^{|c|} similarity(DE_i, DE_j) \tag{4}$$

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

5. Experimental design

In this section, we describe the evaluation of our proposed multilingual aspect clustering technique. Initially, the experimental design is presented, and then the results are discussed. The experimental design includes the description of the datasets, baselines, evaluation metrics, and the parameters used in the configuration of all systems.

5.1. Datasets

In order to evaluate our proposed solutions, we required multilingual datasets of reviews in which the aspects had already been annotated. Since we are addressing a novel task, there were no ready datasets tailored to the evaluation of multilingual aspect clustering.

Our first dataset was originally designed to assess monolingual aspect extraction (each language being treated separately). It was created for SemEval 2016 - Task 5.[1] This is a multilingual dataset with *restaurant* reviews in five languages: English, Dutch, Russian, Spanish, and Turkish. Reviews are divided into sentences. Every sentence has its 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.

The annotations in this dataset include explicit and implicit aspects. Here, we are only considering the explicit aspect phrases in the reviews because we cannot extract contextual information of the implicit aspect phrases, since their opinion target is annotated as `NULL`. In addition, some aspect phrases are categorized in more than one aspect cluster, so we chose as the category the one with the most assignments for that aspect phrase. We made this decision because we noticed that just a few aspect phrases have occurrences in two or more categories — 129 in 3550 aspect phrases. Most cases in which an aspect phrase was annotated in more than one aspect group the aspect groups were `Restaurant`, `Ambience`, and `Location`. These three categories are quite similar, and people tend to use the same aspect phrases in order to describe them. For example, the aspect phrase *restaurant* can be used to describe the restaurant itself (*e.g., Best restaurant in Brooklyn*), the ambience (*e.g., It is a small cute restaurant*), and the restaurant location (*e.g., The restaurant looks out over beautiful green lawns to the Hudson River and the Statue of Liberty*). Looking at the frequency of the *restaurant* aspect in the English dataset we see that out of 43 occurrences, 33 refer to the aspect group `Restaurant`, nine refer to `Ambience`, and in just one it refers to `Location`. Since the reviews are from multiple restaurants, very different aspect phrases are expected to be grouped in the same category. For example, in the `Food`, category, we have aspect phrases related to Italian restaurants (ravioli, pizza), Japanese Restaurants (sushi, temaki), French Restaurants (foie gras, gâteau), among many other kinds of restaurants. This excessive variation of aspects poses a challenge for the clustering task.

The second dataset contains a total of 14,381 reviews from ten digital cameras collected from Amazon. The reviews are in English, German, Italian, Spanish, and French. The dataset was annotated with the opinion target and the gold standard clusters. The following protocol was adopted in the annotation process. First, we extracted the aspect phrases from the reviews using Poria et al.'s algorithm [20]. This algorithm was adapted to work with the languages in the dataset (*i.e.,* the appropriate language resources such as dependency parsers and lemmatizers were used). Next, we ranked the twenty most frequent aspects in these reviews for each camera, pruning aspect phrases which were incorrectly labeled by the aspect extraction algorithm. Finally, the most frequent aspects were assigned to nine different aspect categories – `General`, `Dimension`, `Display`, `Exposure Control`, `Lens`, `Imaging`, `Price`, `Zoom`, and `Video`. This annotation process was performed by two different annotators and the results were compared. In case of disagreement, the annotators discussed their perceptions until they reached a consensus. We have made the annotated dataset of digital cameras freely available.[2] Statistics of the datasets can be found in Table 2.

In a real use scenario, the reviews are made for a specific entity (a restaurant, a hotel, a camera, *etc.*.). While this individualization is not possible in SemEval dataset (since the Restaurant for which the review was made is not available), in the Camera dataset, this can be explored.

Table 3 shows the distribution of the aspects in the clusters. In the SemEval dataset, over half of the aspect clusters belong to the `Food` cluster. Also, none of the other aspect clusters has more than 20% of the aspect phrases in the review set *R*. This is a peculiarity of restaurant reviews, in which the food aspect has a huge importance compared to the other aspects of this domain. The Camera dataset presents a more homogeneous distribution with the biggest cluster containing 44% of the aspect phrases. While the number of reviews is much larger on the Camera dataset, the number of aspect phrases is larger on the SemEval dataset.

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

[2] https://github.com/lucasrafaelc/Multilingual-Aspect-Clustering.

Table 3
Distribution of the aspect clusters in the reviews.

SemEval dataset		
Aspect cluster	#Aspect phrases	%
Restaurant	416	11.72
Food	1,828	51.49
Drinks	256	7.21
Service	497	14.00
Ambience	497	14.00
Location	56	1.58
Total	3,550	100
Camera dataset		
Aspect cluster	#Aspect phrases	%
General	7,844	44.03
Battery	138	0.77
Dimension	381	2.14
Display	101	0.57
Exposure control	502	2.82
Imaging	4,397	24.68
Lens	306	1.72
Memory	99	0.56
Other	2,675	15.01
Performance	124	0.70
Price	727	4.08
Video	337	1.89
Zoom	184	1.03
Total	17,815	100

This happens because the number of restaurants in the reviews is much greater than the number of cameras. The differences between the datasets allow us to test how our technique behaves under different settings and domains.

5.2. Baselines

We used two baselines already mentioned in Section 3.2. The first is the algorithm proposed by Zhai et al. [8] (L-EM). This method was chosen because it is the most seminal work in aspect clustering, and because it can be applied to our datasets since it does not require semi-structured data or extra manual annotations in the reviews. The configurations are the same as in the original article, described in Section 3.2.

The second baseline is ABAE, a neural attention-based method for aspect extraction proposed by He et al. [9]. We used the same parameters suggested by the authors (word-embeddings dimension = 200, batch-size = 50, vocabulary size = 9000, training epochs = 15). The only change was on the number of desired aspect clusters which was set to six for SemEval and to thirteen in the Camera dataset. In order to train ABAE, we needed a sample of reviews from the domain. Thus, for the SemEval dataset, we crawled restaurant reviews from TripAdvisor in all five languages. As for the Camera dataset, we collected reviews from Amazon also in five languages. In both datasets, we gathered about 20 Mb of text per language as this is the amount of data used in the original paper [9].

Since these techniques were originally designed for monolingual aspect clustering, we had to make some adaptations for them to work with multilingual data. First, we used machine translation to map all reviews to English. 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 document of the aspect phrase 'dessert'.

5.3. Evaluation metrics

As proposed by Zhai et al. [8], we measured the performance of our clustering algorithms in terms of Entropy and Purity. In our evaluation, we consider the set AC, clustered into k disjoint sets $\{c_1, c_2, \ldots, c_k\}$ and its respective golden partitions $G = \{g_1, g_2, \ldots, g_k\}$. The goal of the clustering algorithm is to *minimize* entropy and *maximize* purity.

Purity. Purity intends to measure the largest portion of a cluster that contains data from a single golden partition, *i.e.,* the highest percentage of correctly clustered points. It can be calculated as in Eq. (5), where $P_i(g_i)$ is the proportion of g_i data points in c_i. The purity of entire clusters is calculated according to Eq. (6). Multilingual clustering is evaluated in the same way as monolingual clustering. The only difference is that the golden partitions have aspect phrases in many languages.

$$purity(c_i) = \max_j P_i(g_j) \tag{5}$$

$$purity_{total} = \sum_{i=1}^{k} \frac{|c_i|}{|AC|} purity(c_i) \tag{6}$$

Entropy. The entropy of a cluster is measured by the proportion of each gold partition present in it. It is calculated as in Eq. (7). The entropy of a cluster is obtained following Eq. (8).

$$entropy(DS_i) = -\sum_{j=1}^{k} P_i(g_j) \log_2 P_i(g_j) \tag{7}$$

$$entropy_{total} = \sum_{i=1}^{k} \frac{|c_i|}{|AC|} entropy(c_i) \tag{8}$$

Coverage. The coverage for a set of reviews R (in one or more languages) is the percentage of the expected AC that can be found in the reviews. The idea is to assess how many of the product features are commented in the set of reviews. Thus, if a product has ten features, but only seven are rated in the set of reviews, then its coverage if 70%.

$$coverage_{R_l} = \frac{|AC_{R_l}|}{|G|} \tag{9}$$

5.4. Setup for MAC and BMAC

The configuration setup for our proposal is as follows. First, a pre-processing phase are employed, consisting on three standard steps: (*i*) splitting of the review text into sentences; (*ii*) tokenization; and (*iii*) converting all words to lowercase.

FastText[3] 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 to treat out of vocabulary words, which enriched our review representations. In order to align the FastText vectors, we use the transformation matrices by Smith et al. [57].[4] We chose not to train our own word embedding representation because that would require a large amount of data to achieve a minimally satisfactory word embedding model. Since we are working with reviews, which tend to be short texts, the task is harder — especially for some languages with few reviews available online.

We used Gensim[5] to get the word vector representations and for computing the similarity measures (WMD and cosine). We implemented our own versions of the clustering algorithms. This was necessary because our input consists of sets of word embeddings and existing implementations were not able to readily

3 Available at https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md.
4 Available at https://github.com/Babylonpartners/fastText_multilingual.
5 Available at https://radimrehurek.com/gensim/.

deal with them. We experimented with two algorithms: the traditional *k*-medoids (MAC) and the bisecting *k*-medoids algorithm (BMAC). In MAC runs, we set the value of *k* to six in SemEval and to thirteen in the Camera dataset, as these were the number of gold partitions in our datasets. For BMAC, we set the threshold *s* to 100 for the SemEval dataset and to eight for the Camera dataset, which means that our clusters will have a maximum of a hundred (or eight) aspect phrases each. This threshold is set empirically and depends on the size of the dataset. Please, refer to the results in Section 6.5 for an analysis of how this parameter affects the quality metrics.

We also tested two different medoid selection techniques: random selection (RAND) and choosing the most frequent aspect phrase (CENT). For the RAND runs, we ran the algorithm ten times and calculated the averages for purity and entropy to mitigate the effects of variability. This was not required for the CENT configuration, as the results do not change across different runs since the chosen medoids are always the same.

6. Results

In this Section, we evaluate our proposed algorithms in a number of ways. First, in Section 6.1, we assess how MAC and BMAC rate in terms of purity and entropy compared to the baselines in monolingual and multilingual aspect clustering. Then, in Section 6.2, we report on two experiments that show the benefits of using a combination of languages as opposed to having languages in isolation. In Section 6.3, we perform a visual inspection of the generated clusters to evaluate their coherence and again confirm that multilingual clustering can produce more coherent groups. Finally, in Sections 6.4 and 6.5, we show how variations of the parameters *k* in MAC and *s* in BMAC affect the results of these algorithms.

6.1. Analysis of entropy and purity

Here we show the results of the tests performed with our Multilingual Aspect Clustering solution under different configurations regarding the clustering algorithm (MAC or BMAC), medoid selection (CENT or RAND), and similarity metric (WMD or COS) – resulting in eight possible combinations. We also show a comparison against two monolingual baselines (L-EM and ABAE), which were adapted to work with multilingual data. Tables 4 and 5 present the scores for purity and entropy, respectively. The best results are in bold. For the Camera dataset, the results are the average of all ten camera models in the dataset.

Table 4
Purity results — the higher, the better.

SemEval dataset						
Method	Monolingual					Multilingual
	English	Dutch	Russian	Spanish	Turkish	
L-EM	0.60	0.59	0.50	0.64	0.48	0.55
ABAE	**0.69**	0.56	0.56	**0.78**	0.50	0.60
MAC-RAND-WMD	0.61	0.56	0.54	0.64	0.49	0.54
MAC-RAND-COS	0.61	0.56	0.55	0.65	0.48	0.54
MAC-CENT-WMD	0.63	0.58	0.57	0.58	0.49	0.54
MAC-CENT-COS	0.62	0.55	0.52	0.66	0.48	0.56
BMAC-RAND-WMD	0.61	0.56	0.61	0.65	0.50	0.60
BMAC-RAND-COS	0.67	**0.61**	**0.63**	0.67	**0.51**	0.62
BMAC-CENT-WMD	0.61	0.56	0.58	0.65	0.49	0.61
BMAC-CENT-COS	0.65	0.60	0.59	0.67	**0.51**	**0.63**
Camera dataset						
Method	Monolingual					Multilingual
	English	German	Italian	Spanish	French	
L-EM	0.57	0.59	0.61	0.58	0.56	0.47
ABAE	0.61	**0.67**	**0.74**	**0.68**	**0.62**	0.57
MAC-RAND-WMD	0.60	0.57	0.59	0.62	0.58	0.39
MAC-RAND-COS	0.61	0.58	0.60	0.59	0.56	0.39
MAC-CENT-WMD	0.61	0.61	0.55	0.59	0.57	0.38
MAC-CENT-COS	**0.63**	0.61	0.61	0.58	0.55	0.40
BMAC-RAND-WMD	0.51	0.42	0.57	0.47	0.49	0.48
BMAC-RAND-COS	0.55	0.63	0.62	0.57	0.58	**0.63**
BMAC-CENT-WMD	0.47	0.42	0.55	0.48	0.45	0.50
BMAC-CENT-COS	0.50	0.51	0.57	0.49	0.49	0.54

Table 5
Entropy results — the lower the better.

SemEval dataset						
Method	Monolingual					Multilingual
	English	Dutch	Russian	Spanish	Turkish	
L-EM	1.75	1.75	1.97	1.59	2.08	1.93
ABAE	**1.21**	1.56	1.53	**0.97**	1.75	1.52
MAC-RAND-WMD	1.65	1.80	1.81	1.52	2.00	1.86
MAC-RAND-COS	1.58	1.70	1.71	1.44	2.04	1.83
MAC-CENT-WMD	1.62	1.72	1.71	1.54	2.04	1.84
MAC-CENT-COS	1.49	1.65	1.72	1.34	2.02	1.75
BMAC-RAND-WMD	1.59	1.71	1.55	1.43	1.96	1.53
BMAC-RAND-COS	1.24	**1.40**	**1.42**	1.29	1.85	1.44
BMAC-CENT-WMD	1.55	1.72	1.66	1.44	1.94	1.55
BMAC-CENT-COS	1.34	1.57	1.44	1.29	**1.83**	**1.38**
Camera dataset						
Method	Monolingual					Multilingual
	English	German	Italian	Spanish	French	
L-EM	1.30	1.15	1.11	1.25	1.22	1.91
ABAE	1.02	**0.82**	**0.70**	**0.81**	**1.00**	1.39
MAC-RAND-WMD	1.09	1.17	1.17	1.06	1.19	2.11
MAC-RAND-COS	1.16	1.29	1.22	1.23	1.37	2.20
MAC-CENT-WMD	1.03	1.06	1.22	1.12	1.22	2.08
MAC-CENT-COS	**1.01**	1.09	1.18	1.16	1.33	2.09
BMAC-RAND-WMD	1.52	1.69	1.29	1.54	1.57	1.56
BMAC-RAND-COS	1.38	1.16	1.10	1.29	1.28	**1.11**
BMAC-CENT-WMD	1.61	1.71	1.30	1.58	1.59	1.51
BMAC-CENT-COS	1.49	1.43	1.35	1.51	1.57	1.39

Multilingual results. The winning combination in terms of purity and entropy was BMAC-CENT-COS for the SemEval dataset and BMAC-RAND-COS in the Camera dataset, outperforming both baselines. Comparing the results across datasets, we observe a similar range of scores. This implies that the greater heterogeneity of the SemEval dataset did not have a negative impact on the performance of the systems. In comparison to L-EM, MAC and BMAC are unsupervised and do not require the two phases of preprocessing before clustering the aspect phrases. In contrast with ABAE, MAC and BMAC do not require training to cluster aspects and are able to deal with aspect phrases (and not just single words). However, ABAE is more complete in the sense that it can also perform aspect extraction, while we focus on clustering. Finally, recall that both baselines required translating the reviews, which represents a significant overhead.

Monolingual results. Looking at the monolingual results in the SemEval dataset, we can see that our solutions win in three out of five languages in terms of both entropy and purity (Dutch, Russian, and Turkish). MAC and BMAC were outperformed by ABAE in English and Spanish. In the Camera dataset, MAC-CENT-COS outperformed the baselines for English, but ABAE was better in the other languages. Comparing the relative performance for the different languages, for SemEval, Spanish had the best results (*i.e.,* higher purity and lower entropy), followed by English. Turkish, however, had the worst results. The same pattern was observed in all systems. In the Camera Dataset, Italian had the highest purity and the lowest entropy.

MAC versus BMAC. BMAC consistently outperformed MAC both on monolingual and multilingual runs across the two datasets, in terms of Purity. In terms of entropy, while MAC has better scores in the monolingual setting, BMAC was superior in the multilingual runs We attribute the superiority of BMAC to the fact that specifying the number of items per cluster allows generating groups of aspects that adapt better to the characteristics of the dataset. We also noticed that BMAC tends to be better for larger datasets.

RAND **versus** CENT. Although most of the winning combinations employed random medoid selection, the difference between the two strategies was very small. By averaging all the runs in which each of the two alternatives was used, we can see that the difference is only 0.02 in favor of random selection. This shows that both could be employed as viable alternatives.

COS **versus** WMD. Cosine was a better similarity metric than WMD. This happens because WMD is able to assign a low distance to aspect phrases that are similar even though they do not share words. While this is a desirable property in some cases, in other cases, it generated clusters with a higher entropy.

6.2. Monolingual versus multilingual clustering

Analysis of purity and entropy. Looking at the results in Tables 4 and 5 we can see that for all systems, the multilingual run is outperformed by at least one of the monolingual runs. Thus, in order to quantify the gain obtained by the multilingual approach, we calculated the proportional increase/decrease in purity and entropy incurred by each system/language combination. In this analysis, we used our best performing multilingual configurations (*i.e.,* BMAC-CENT-COS for SemEval and BMAC-RAND-COS for the Camera dataset), and our baselines ABAE and L-EM.

Fig. 3 shows a bar chart that quantifies the gains and losses of multilingual aspect clustering in comparison to the monolingual setting. The positive bars (*i.e.,* that are above zero) indicate that the corresponding language has benefited from the multilingual setting – *i.e.,* the multilingual performance was better than the monolingual performance in the given language, according to the results in Tables 4 and 5. For example, consider the purity scores for BMAC-CENT-COS in the SemEval dataset — in a monolingual setting, Turkish scored 0.51 but when it was combined with the other four languages, the score increased to 0.63, representing a proportional improvement of 24%. We can see that, in SemEval, Turkish, and Russian have benefited from multilingual clustering in all three systems. On the other hand, English and Spanish were better off in the monolingual setting. Dutch also benefited from the multilingual approach, in both BMAC and ABAE. In the Camera dataset, the multilingual runs in the baselines had worse performances than the monolingual ones. However, with BMAC, the results had a significant increase in English, Spanish, and French, both purity and entropy. Finally, we can also see in Fig. 3 that, for BMAC, the gains of the multilingual approach were as high as 33% while the biggest loss was 6%. If we take the mean of all gains and losses, we get a positive balance of about 10% in both datasets. Languages with a better monolingual result end up leveraging the quality of the clustering for languages with lower monolingual results. It is important to point out that the multilingual results are better than the average of the monolingual results, which means that the gains outweigh the losses. These results confirm the benefits of multilingual aspect clustering.

Coverage of individual and combined languages. Going back to our motivating example shown in Fig. 1, we saw that some languages do not have reviews comprising all aspects of a given product. This represents a limitation on the coverage that can be achieved by ABSA systems. In order to quantify the coverage of individual and combined languages, we analyzed the annotations of the Camera dataset (in which we can identify the target entity since the reviews are individualized). Recall that the Camera dataset has 13 aspect clusters (AC) for each camera model. Considering the ten camera models in the dataset, full coverage requires having ratings for all 13 product features of all ten models. The analysis consisted in calculating the number of AC that were present in any single language, and all possible combinations of two, three, four, and five languages. The results in Table 6 show the average, minimum and maximum coverage for these combinations. When a single language is used, we have an average coverage of 60%, meaning that only 8 out of 13 aspects tend to be present for any given camera model. The English version has the highest coverage, and still it reaches only 63%. The gain from adding a second language is quite significant, as the average coverage jumps to 71%. Adding more languages keeps improving coverage, with a combination of English, Spanish, French, and German being able to account for 11 AC. Even considering all five languages, not all camera models had all their features rated. These findings are in line with the work by Banea et al. [7] which quantified the gain in subjectivity classification performance when more languages are used. The results of this analysis advocate for the use of multilingual approaches.

Table 6
Coverage of individual and combined languages.

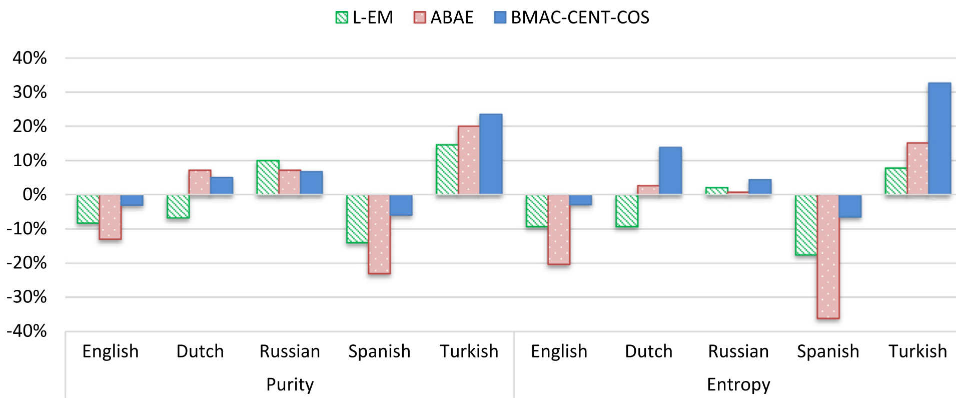
# of languages	Avg coverage	Min coverage	Max coverage
1	60%	54%	63%
2	71%	66%	75%
3	78%	73%	85%
4	82%	80%	85%
5	85%	85%	85%

6.3. Analysis of the resulting clusters

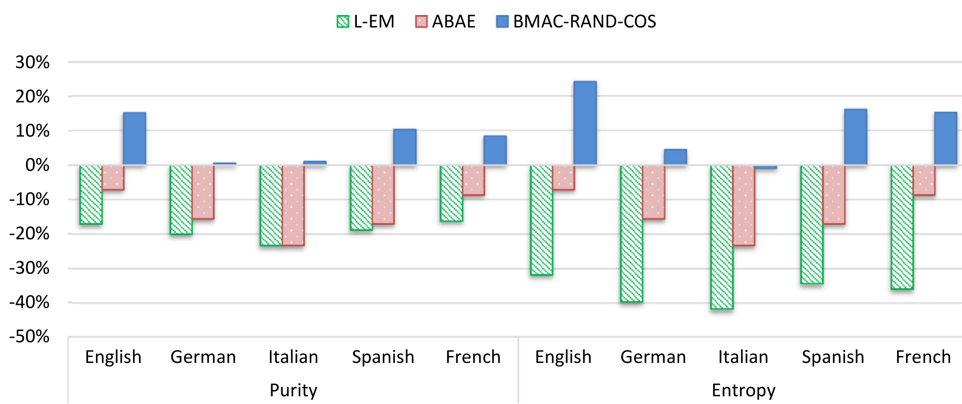
SemEval clusters. Table 7 shows some excerpts of the aspect clusters AC generated by BMAC for the SemEval dataset. Based on these results, we discuss some strengths and weaknesses of our approach. Cluster number one, for example, shows that our method is able to group aspect phrases that are synonyms in the same language (*waitress*, *waiters*, *waitstaff*), or across languages (staff, персонал and çalışanlar). 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 two, where we see many word groups where this property can be validated (рестораном, ресторане, and ресторан, for example). This property can be considered as an improvement over the heuristic used by L-EM, 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 three, that groups the aspect phrases with the word *"menu"*, which has the same form in English, Dutch, Spanish, and Turkish. It is interesting to note that our algorithm also includes in this cluster the aspect phrase меню, which is the translation of the word menu to Russian.

MAC and BMAC are able to group semantically related aspect phrases. We present two examples of such ability in Table 7. Cluster four groups aspect phrases related to seafood dishes. At the same time, cluster five has aspect phrases related to the decoration of the restaurants. We obtained this result thanks to an adequate representation of the virtual documents, allied to a measure of similarity that is strong enough to capture the contextual similarity between aspect phrases.

Camera clusters. To further illustrate the advantages of the multilingual approach, Fig. 4 shows excerpts of clusters generated in a monolingual and in a multilingual setting for a single digital camera. We can see that in the monolingual setting, attributes relating to different aspect categories were mixed in the same cluster in German and Italian. In Spanish, aspects from one category were allocated to different clusters. In the multilingual



(a) SemEval



(b) Cameras

Fig. 3. Proportional Gain/Loss of using multilingual versus monolingual aspect clustering.

approach, the clusters are more coherent. There is one cluster for aspects related to `Price` (highlighted in green) and aspects for the `General` category that were separated in Italian, appear together in cluster #9 (highlighted in orange). Despite being better, multilingual clustering is not perfect. The example shows that one aspect phrase relating to `Price` was assigned the wrong cluster. The aspects in gray were allocated to other clusters which are not shown in Fig. 4 due to space limitations (as there were 14 clusters in total).

Limitations: Despite the good results, our algorithm has some limitations. Because it is unsupervised, it suffers from the drawbacks of this type of approach. 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 compare the cosine similarity of a word and its translations into other languages. For example, the distance between the vector for the word 'dessert' and its translations десертам, postre, nagerecht and tatlı is 0.68, 0.59, 0.71, 0.48 respectively, while the most similar words in English are desserts (0.91), pastries (0.81), cakes (0.76), pancakes (0.74), and salads (0.74). This happens in cluster two of Table 7, which has only aspect phrases in Russian. Some of those aspect phrases are more related to other clusters instead of cluster two, for instance, музыка (music) and живая музыка (live music) are more related to cluster five.

6.4. Variation of parameter *k* in MAC

To assess how the method behaves with different numbers of clusters, we performed tests varying the parameter k in the SemEval dataset. The results can be seen in Fig. 5. When the number of clusters increases, our methods showed a more pronounced drop in entropy and a gain in purity. At the same time, our approach tends to select better initial medoids as the number of cluster increases. For a small number of clusters, our medoid selection technique did not work well, because the most frequent aspect phrases referred to the same aspect cluster. For example, it selected as medoids the aspects service, обслуживание музыка, and servicio, which are the translation of service in Russian and Spanish. With a larger number of clusters, the medoid selection algorithm tends to select more diversified aspect phrases.

6.5. Variation of parameter *s* in BMAC

In order to understand the behavior of the parameter s in BMAC, we performed tests with values ranging from 10 to 500 in the SemEval dataset. Fig. 6 shows the number of clusters generated for each execution of BMAC on the 3550 aspect phrases. These measures tend to grow/decrease exponentially as the number of aspects per cluster increase. Higher values of s tend to yield lower purity and entropy values and a small number of clusters. On the other hand, a small s value produces a large number of clusters, which is not desirable for most applications. The charts also show that the use of WMV as similarity measure tends to yield worse results — especially for small values of s. The medoid selection technique tends to produce more consistent results, considering

Table 7
Excerpts of clusters generated by BMAC for the SemEval dataset.

#	Medoid	Aspects
1	waitress	waiters – waitstaffs – staff – people – trato personal (personal care) – eigenaars (owners) – персонал (staff) – человек (person) – çalışanları (employees)
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) кухню (kitchen) – кухней (kitchen) – кухня (kitchen) – Качество кухни (quality of kitchen) – кухни (kitchens) 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	crab salad	grilled black cod, lobster bisque, bbq salmon, mejillones vapor (Steamed Mussels), ración gambas (prawns ration), garnaalkroketten (shrimp croquettes), langoest (lobster), stukje zalm (piece of salmon), soya soslu somon (salmon with soy sauce)
5	atmosphere	interior, downstairs lounge, decor, ambience, vibe, outdoor seating, furnishings, entorno (environment), limpio (cleansed), decoración (decor), ambiente (ambience), setting (surroundings), sfeer muziek (background music), музыкальная программа (music program), dekorasyonda (decoration), müzikleri (music)

German

CL	Aspect Phrase	Category
2	bilder	IMAGING
2	schnappschüsse	IMAGING
2	kamera	GENERAL
2	filme	VIDEO
2	fotos	IMAGING
2	sofortbildkamera	GENERAL
2	produkt	GENERAL
2	preis	PRICE

Italian

CL	Aspect Phrase	Category
1	prodotto	GENERAL
1	articolo	GENERAL
1	rapporto qualità prezzo	PRICE
1	acquisto	GENERAL
1	prezzo	PRICE

Spanish

CL	Aspect Phrase	Category
1	compra	GENERAL
3	producto	GENERAL
3	forma	DIMENSION
3	cámara	GENERAL
3	calidad	GENERAL
4	fotos	IMAGING
4	precio	PRICE
4	tamaño	DIMENSION
4	luz	EXPOSURE CONTROL
4	objetivo	LENS

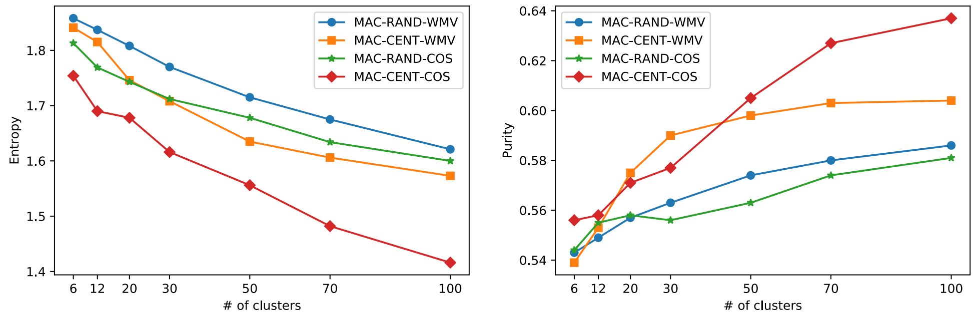
(a) Monolingual

CL	Aspect Phrase	Category
9	produkt	GENERAL
9	prodotto	GENERAL
9	acquisto	GENERAL
9	rapporto qualità prezzo	PRICE
9	compra	GENERAL
9	producto	GENERAL

CL	Aspect Phrase	Category
11	preis	PRICE
11	prezzo	PRICE
11	precio	PRICE

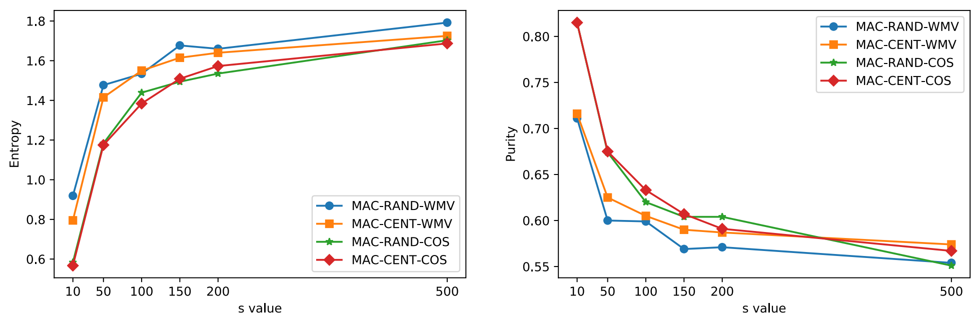
(b) Multilingual

Fig. 4. Excerpts of clusters created in a monolingual and in a multilingual setting. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Entropy (b) Purity

Fig. 5. Experimental results with the variation of *k*.



(a) Entropy (b) Purity

Fig. 6. Experimental results with the variation of threshold *s*.

that its values oscillate less than those produced by random selection. We also noticed that the cosine similarity measure tends to generate more clusters, especially for small values of *s*. In this work, we chose this parameter empirically, by estimating results of BMAC produced by the different *s* values and the clusters generated by each execution. The value of *s* is directly related to the size of the dataset. An estimation can be obtained by dividing the number of aspect phrases in the dataset by *s* which, yields the minimum number of clusters that will be generated by BMAC.

7. Conclusion

In this article we addressed the task of multilingual aspect clustering. Our main contributions include the formulation of the problem and the proposal of an unsupervised technique to solve it. The proposed technique combines multilingual word embeddings and a similarity measure in order to group related aspect phrases in multiple languages. Because it is unsupervised, our approach can be applied across domains, requiring only word embeddings for each language in the set of reviews.

We carried out experiments on a publicly available dataset containing restaurant reviews and we also crawled and annotated a dataset with reviews on digital cameras. Two systems that perform monolingual aspect clustering were adapted to work with multilingual data so that they could be used as baselines. The results showed that our unsupervised multilingual clustering technique achieves results that outperform the adapted baselines.

Furthermore, our experimental results demonstrate that multilingual aspect clustering is advantageous compared to using languages in isolation. This finding was observed in three different situations. (*i*) In a multilingual setting, one language can leverage the performance of the other. Overall, the gains for languages with poorer monolingual performances are greater than the losses incurred by the languages with better monolingual performances, resulting in a positive balance. (*ii*) The coverage of the aspects that are present in the reviews increases significantly as more languages are used. (*iii*) An inspection of the resulting clusters showed that multilingual clustering is able to create more coherent groups of aspects.

To the best of our knowledge, we are the first work to address multilingual aspect clustering. Our contribution opens up possibilities for future work. We envisage the following extensions: (*i*) Investigating a more accurate technique for word embedding normalization and pruning irrelevant aspect phrases to improve results; (*ii*) Allowing multi-granular clusters through the use of hierarchical clustering algorithms; (*iii*) Dealing with implicit aspects; (*iv*) Investigating aspect ambiguity, which occurs when an aspect phrase is assigned to different aspect clusters. A better understanding of this problem may bring new insights and improve the results achieved by this work; (*v*) Automatically deciding on the best parameter configuration based on the characteristics of the dataset; and (*vi*) Building a tool that caters for the entire framework shown in Fig. 1 supporting multilingual aspect extraction and visualization to summarize the results emphasizing the aspects in which the user is interested.

CRediT authorship contribution statement

Lucas Rafael Costella Pessutto: Conceptualization, Methodology, Writing - original draft, Software, Investigation, Validation, Visualization. **Danny Suarez Vargas:** Software, Resources, Writing - original draft. **Viviane P. Moreira:** Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision.

Acknowledgments

This work was partially supported by CNPq/Brazil and by CAPES Finance Code 001. The authors thank Tiago Melo for his work in generating the multilingual dataset on digital cameras. Finally, we thank the anonymous reviewers for their helpful suggestions which significantly improved our work.

References

[1] B. Liu, Opinion mining and sentiment analysis, in: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, second ed., Springer Science & Business Media, 2011, pp. 459–526, Ch. 11.

[2] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowl.-Based Syst. 89 (2015) 14–46.

[3] O. Appel, F. Chiclana, J. Carter, H. Fujita, Cross-ratio uninorms as an effective aggregation mechanism in sentiment analysis, Knowl.-Based Syst. 124 (2017) 16–22.

[4] O. Appel, F. Chiclana, J. Carter, H. Fujita, A consensus approach to the sentiment analysis problem driven by support-based iowa majority, Int. J. Intell. Syst. 32 (9) (2017) 947–965.

[5] O. Appel, F. Chiclana, J. Carter, H. Fujita, Successes and challenges in developing a hybrid approach to sentiment analysis, Appl. Intell. 48 (5) (2018) 1176–1188.

[6] A. Valdivia, M.V. Luzón, F. Herrera, Sentiment analysis in TripAdvisor, IEEE Intell. Syst. 32 (4) (2017) 72–77.

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

[8] Z. Zhai, B. Liu, H. Xu, P. Jia, Clustering product features for opinion mining, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, 2011, pp. 347–354.

[9] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, An unsupervised neural attention model for aspect extraction, in: Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017, pp. 388–397.

[10] L.R.C. Pessutto, D.S. Vargas, V.P. Moreira, Clustering multilingual aspect phrases for sentiment analysis, in: IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 182–189.

[11] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Adv. Neural Inf. Process. Syst. 26 (2013) 3111–3119.

[12] T. Mikolov, W.-t. Yih, G. Zweig, Linguistic regularities in continuous space word representations, in: NAACL: HLT, 2013, pp. 746–751.

[13] S. Ruder, I. Vulić, A. Søgaard, A survey of cross-lingual word embedding models, J. Artificial Intelligence Res. 65 (2019) 569–631.

[14] P. Tan, M. Steinbach, A. Karpatne, V. Kumar, Introduction to data mining, What's New in Computer Science Series, Pearson Education, 2013.

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

[16] B. Liu, Sentiment analysis and opinion mining, Synth. Lect. Hum. Lang. Technol. 5 (1) (2012) 1–167.

[17] L. Zhang, B. Liu, Aspect and entity extraction for opinion mining, in: Data Mining and Knowledge Discovery for Big Data, Springer, 2014, pp. 1–40.

[18] M. Hu, B. Liu, Mining and summarizing customer reviews, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.

[19] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27.

[20] S. Poria, E. Cambria, L.-W. Ku, C. Gui, A. Gelbukh, A rule-based approach to aspect extraction from product reviews, in: Workshop on Natural Language Processing for Social Media (SocialNLP), 2014, pp. 28–37.

[21] W. Jin, H.H. Ho, 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, Citeseer, Montreal, Quebec, 2009, pp. 465–472.

[22] N. Jakob, I. Gurevych, Extracting opinion targets in a single-and cross-domain setting with conditional random fields, in: Conference on Empirical Methods in Natural Language Processing, 2010, pp. 1035–1045.

[23] S. Moghaddam, M. Ester, Ilda: interdependent lda model for learning latent aspects and their ratings from online product reviews, in: ACM SIGIR Conference on Research and Development in Information Retrieval, 2011, pp. 665–674.

[24] Z. Toh, J. Su, NLANGP: Supervised machine learning system for aspect category classification and opinion target extraction, in: International Workshop on Semantic Evaluation (SemEval), 2015, pp. 496–501.

[25] Z. Toh, J. Su, 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), Association for Computational Linguistics, San Diego, California, 2016, pp. 282–288.

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

[27] W. Wang, S.J. Pan, D. Dahlmeier, X. Xiao, Coupled multi-layer attentions for co-extraction of aspect and opinion terms, in: AAAI, 2017, pp. 3316–3322.

[28] A. Giannakopoulos, C. Musat, A. Hossmann, M. Baeriswyl, Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets, in: Workshop on Computational Approaches To Subjectivity, Sentiment and Social Media Analysis, 2017, pp. 180–188.

[29] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49, New Avenues in Knowledge Bases for Natural Language Processing.

[30] B. Liu, M. Hu, J. Cheng, Opinion observer: analyzing and comparing opinions on the web, in: International Conference on World Wide Web, 2005, pp. 342–351.

[31] G. Carenini, R.T. Ng, E. Zwart, Extracting knowledge from evaluative text, in: Proceedings of the 3rd International Conference on Knowledge Capture, ACM, 2005, pp. 11–18.

[32] H. Guo, H. Zhu, Z. Guo, X. Zhang, Z. Su, Product feature categorization with multilevel latent semantic association, in: Conference on Information and Knowledge Management, 2009, pp. 1087–1096.

[33] Z. Zhai, B. Liu, H. Xu, 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.

[34] J. Pavlopoulos, I. Androutsopoulos, Multi-granular aspect aggregation in aspect-based sentiment analysis, in: Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 78–87.

[35] Y. Zhao, B. Qin, 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.

[36] Y. He, J. Song, Y. Nan, 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.

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

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

[39] F.A. Vargas, T.A.S. Pardo, Aspect clustering methods for sentiment analysis, in: International Conference on Computational Processing of the Portuguese Language, 2018, pp. 365–374.

[40] S. Ma, C. Zhang, D. He, Document representation methods for clustering bilingual documents, in: ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives Through Information & Technology, ASIST '16, 2016, pp. 65:1–65:10.

[41] I. Flaounas, O. Ali, M. Turchi, T. Snowsill, F. Nicart, T. De Bie, N. Cristianini, Noam: news outlets analysis and monitoring system, in: SIGMOD International Conference on Management of Data, 2011, pp. 1275–1278.

[42] A. Rauber, M. Dittenbach, 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: All-Russian Conference Digital Libraries: Advanced Methods and Technologies, 2001.

[43] H.-H. Chen, C.-J. Lin, A multilingual news summarizer, in: Conference on Computational Linguistics-Volume 1, 2000, pp. 159–165.

[44] B. Mathieu, R. Besançon, C. Fluhr, Multilingual document clusters discovery, in: Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval, 2004, pp. 116–125.

[45] X. Hong, Z. Yu, M. Tang, Y. Xian, Cross-lingual event-centered news clustering based on elements semantic correlations of different news, Multimedia Tools Appl. 76 (23) (2017) 25129–25143.

[46] B. Pouliquen, R. Steinberger, C. Ignat, E. Käsper, I. Temnikova, Multilingual and cross-lingual news topic tracking, in: International Conference on Computational Linguistics, 2004, p. 959.

[47] S. Montalvo, R. Martínez, A. Casillas, V. Fresno, Multilingual news document clustering: two algorithms based on cognate named entities, in: International Conference on Text, Speech and Dialogue, 2006, pp. 165–172.

[48] S. Montalvo, R. Martinez, A. Casillas, V. Fresno, Multilingual document clustering: an heuristic approach based on cognate named entities, in: International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics, 2006, pp. 1145–1152.

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

[50] D. Yogatama, K. Tanaka-Ishii, Multilingual spectral clustering using document similarity propagation, in: Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2, 2009, pp. 871–879.

[51] C.-P. Wei, C.C. Yang, C.-M. Lin, A latent semantic indexing-based approach to multilingual document clustering, Decis. Support Syst. 45 (3) (2008) 606–620.

[52] K. Denecke, Using sentiwordnet for multilingual sentiment analysis, in: 2008 IEEE 24th International Conference on Data Engineering Workshop, 2008, pp. 507–512.
[53] Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D.K. Honbo, A. Agrawal, A.N. Choudhary, Muses: Multilingual sentiment elicitation system for social media data, IEEE Intell. Syst. 29 (4) (2014) 34–42.
[54] J. Steinberger, M. Ebrahim, M. Ehrmann, A. Hurriyetoglu, M. Kabadjov, P. Lenkova, R. Steinberger, H. Tanev, S. Vazquez, V. Zavarella, Creating sentiment dictionaries via triangulation, Decis. Support Syst. 53 (4) (2012) 689–694.
[55] M.S. Hajmohammadi, R. Ibrahim, A. Selamat, H. Fujita, Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples, Inform. Sci. 317 (2015) 67–77.
[56] A. Balahur, M. Turchi, R. Steinberger, J.-M. Perea-Ortega, G. Jacquet, D. Küçük, V. Zavarella, A. El Ghali, Resource creation and evaluation for multilingual sentiment analysis in social media texts, in: Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC), 2014.
[57] S.L. Smith, D.H.P. Turban, S. Hamblin, N.Y. Hammerla, Offline bilingual word vectors, orthogonal transformations and the inverted softmax, 2017, arXiv:1702.03859.