WS4ABSA: An NMF-Based Weakly-Supervised Approach for Aspect-Based Sentiment Analysis with Application to Online Reviews

Alberto Purpura(✉), Chiara Masiero(✉), and Gian Antonio Susto(✉)

University of Padova, Padova 35122, Italy
{purpuraa,masieroc,gianantonio.susto}@dei.unipd.it

Abstract. The goal of Aspect-Based Sentiment Analysis is to identify opinions regarding specific targets and the corresponding sentiment polarity in a document. The proposed approach is designed for real-world scenarios, where the amount of available information and annotated data is often too limited to train supervised models. We focus on the two core tasks of Aspect-Based Sentiment Analysis: aspect and sentiment polarity classification. The first task – which consists in the identification of the opinion targets in a document – is tackled by means of a weakly-supervised technique based on Non-negative Matrix Factorization. This strategy allows users to easily embed some a priori domain knowledge by means of short seed terms lists. Experimental results on publicly available data sets related to online reviews suggest that the proposed approach is very flexible and can be easily adapted to different languages and domains.

Keywords: Aspect-based sentiment analysis
Non-negative matrix factorization · Text mining
Weakly-supervised learning

1 Introduction

Sentiment Analysis (SA) [11] is a growing area of research in Natural Language Processing. While SA aims at inferring the overall opinion of the writer in a document, Aspect-Based Sentiment Analysis (ABSA) is concerned with fine-grained polarity analysis, and its purpose is two-fold:

1. extracting relevant aspects - For instance, in a context of on-line reviews on restaurants, relevant aspects could be food, service, location, etc;
2. evaluating the sentiment polarity of each aspect separately.

In the context of real-world applications, there is a clear need for ABSA solutions that are interpretable and flexible, ie. that can be adapted to different

L. Soldatova et al. (Eds.): DS 2018, LNAI 11198, pp. 386–401, 2018.
https://doi.org/10.1007/978-3-030-01771-2_25

languages and domains. For example, ABSA is particularly suitable for Business to Consumer (B2C) companies to improve and develop their products, services or marketing strategies based on the feedback provided by customers in the form of online reviews; however, such online reviews may come from different countries and may be related to different product categories (e.g. laptops and smartphones), making ABSA a difficult task. Moreover, we remark that data annotation in this situation is time consuming and expensive, also because the subjectivity of this task is generally tackled by employing a panel of human annotators; thus, unsupervised or weakly-supervised approaches are often preferred to supervised ones. Furthermore, some domain knowledge is generally available, even if limited and partial. For instance some keywords used to describe aspects are generally known in advance and the same holds true for opinion words. We remark that the sentiment associated with opinion words is aspect-specific; for example, the opinion word 'cheap' conveys a positive sentiment with respect to a 'value for money' aspect in online reviews, while it conveys a negative sentiment on a 'quality' and an 'appearance' in case these aspects are considered in the ABSA task. With this scenario in mind, we propose Weakly-Supervised Approach for ABSA (WS4ABSA), a technique that is able to accomplish the core tasks of ABSA, and can be easily adapted to deal with different domains and languages. WS4ABSA tackles ABSA in two steps: (i) aspect classification and (ii) sentiment polarity classification. As for (i), we present a novel approach based on a well-known Topic Modeling technique, called Non-negative Matrix Factorization (NMF). The proposed approach allows the user to include some domain knowledge – in a weakly-supervised way – in order to link each topic discovered by NMF to the aspects which are referred to in a document. Indeed, WS4ABSA allows the user to embed a list of seed words to guide the algorithm towards more significant topic definitions. Regarding (ii), WS4ABSA employs another weakly-supervised framework based on the definition of a positive and a negative seed list with a few sentiment terms for each topic. These lists are then extended using Word2Vec [13] and used to assign a polarity to each aspect identified in step (i). Our system distinguishes itself from the ones in literature (detailed in Sect. 2) because it does not rely on any auxiliary resources or annotated data sets, but only on the aforementioned lists and some simple grammar rules to deal with negations. Therefore, WS4ABSA can be applied easily for the analysis of documents from any domain and in any language with the advantages of being easily interpretable and implementable. Moreover, WS4ABSA allows the user to include his prior knowledge on the problem and to iteratively improve the results thanks to the reformulation of the NMF problem objective function proposed here. This additional information can be used to steer the classification in a precise and predictable manner. The rest of the paper is organized as follows: WS4ABSA is presented in Sects. 3.1 and 3.2, by illustrating the procedure for aspect extraction and sentiment polarity classification, respectively. In Sect. 4, we test this approach on publicly available data sets, while final remarks and future research directions are reported in Sect. 5.

2 Related Work

Available approaches to ABSA can be divided into supervised [18], semi-supervised [23], weakly-supervised [5] or unsupervised [2] techniques. Support Vector Machines [12], Naive Bayes classifiers [20] and Maximum Entropy classifiers [26] are the most common approaches among the supervised machine learning methods to detect the aspects in a sentence or for sentiment classification. In the realm of supervised approaches, Neural Networks (NN) have also received an increasing interest in recent years. Convolutional NN, for example, have been successfully applied to ABSA such as in [18]. Unfortunately, supervised methods – particularly neural networks – require large annotated corpora to perform well. As said, this is an issue, especially for low-resource languages and specific application domains. Thus, unsupervised approaches are frequently adopted. In general, this type of techniques need labeled data only to test and validate the model. Most of the approaches that fall under this category use topic models to extract aspect and sentiment terms. The most adopted topic modeling techniques are Latent Dirichlet Allocation (LDA) [1] and NMF [14], mainly because their results are easy to interpret thanks to positivity constraints. NMF has two main advantages when compared to LDA: first, it allows for an easier tuning and manipulation of its internal parameters [22]; second, there are efficient and completely deterministic algorithms for computing a reliable approximate solution. W2VLDA [5] uses LDA to detect aspects and extracts the corresponding polarity based on very simple lists of seed words in input. The main drawback of W2VLDA is that it requires a language model trained on a large specific domain corpus to embed domain knowledge in aspect classification. Another topic modeling-based approach is UTOPIAN [3] that provides an interactive topic modeling system based on NMF that allows users to steer the results by embedding their domain knowledge. Finally, also [10] uses NMF to identify general sentiment linguistic indicators from one domain and then gauge sentiment around documents in a new target domain. Another example of the versatility of NMF can be seen in [24], where the authors attempt to learn topics from short documents using term correlation data rather than the usual high-dimensional and sparse term occurrence information in documents.

3 Weakly-Supervised Approach for ABSA (WS4ABSA)

Given a collection of documents, WS4ABSA tackles ABSA in two steps:

(i) based on a list of seed words for each aspect, WS4ABSA performs aspect extraction by means of NMF;
(ii) using a set of sentiment seed words for each aspect, for each document WS4ABSA assigns a sentiment polarity to each of the detected aspects.

3.1 Aspect Classification

In WS4ABSA, aspect classification is achieved by means of NMF. This technique aims at solving the following problem: given a non-negative $m \times n$ matrix

A (i.e. a matrix where each element $A_{ij} \geq 0, \forall i, j$), find non-negative matrix factors $W \in \mathbb{R}_{+}^{m \times k}$ (*term-topic matrix*) and $H \in \mathbb{R}_{+}^{k \times n}$ (*topic-document matrix*), for a given number of aspects $k \in \mathbb{N}_{+}$, such that $A \approx WH$. In our formulation of the problem, A represents the collection of n documents we want to analyze, for example using the Term Frequency-Inverse Document Frequency (TF-IDF) weighting scheme [17] with respect to the m distinct terms contained in the collection. W represents the associations between the terms contained in the collection and the k considered aspects, and H represents the associations between the aspects and each document in the indexed collection. Among the different problem formulations for NMF [8], here we consider the factorization problem based on the Frobenius Norm:

$$\min_{W \geq 0, H \geq 0} f(W, H) = ||A - WH||_F^2, \tag{1}$$

where, with $W, H \geq 0$ we impose the constraint on each element of the matrices to be non-negative, and with $|| \cdot ||_F$ we indicate the Frobenius norm. Although NMF is an NP-hard problem [21], one can still hope to find a local minimum as an approximation. In this work we will focus on the Block Coordinate Descent (BCD) method that is an algorithmic framework to optimize the above objective function. BCD divides variables into several disjoint subgroups and iteratively minimizes the objective function with respect to the variables of each subgroup at a time. Under mild assumptions, it is possible to prove that BCD converges to stationary points [7]. Multiplicative Updating (MU) is another popular framework for solving NMF [9], however it has slow convergence and may lead to inferior quality solutions [7]. In Sect. 3.1 we introduce a novel NMF problem formulation that is particularly suitable to embed domain knowledge, while Sects. 3.1 and 3.1 provide additional implementation details.

Proposed NMF Resolution Method. To solve the NMF problem with the BCD approach, we referred to a method called Hierarchical Alternating Least Squares (HALS) [4]. Let us partition the matrices W and H into $2k$ blocks (k blocks each that are respectively the columns of W and the rows of H), in this case we can see the problem in the objective function in Eq. (1) as

$$||A - WH||_F^2 = ||A - \sum\nolimits_{i=1}^{k} w_{\cdot i} h_{i \cdot}||_F^2. \tag{2}$$

To minimize each block of the matrices we solve

$$\min_{w_{\cdot i} \geq 0} \|h_{i \cdot}^T w_{\cdot i}^T - R_i^T\|_F^2, \quad \min_{h_{i \cdot} \geq 0} \|w_{\cdot i} h_{i \cdot} - R_i\|_F^2, \tag{3}$$

where $R_i = A - \sum_{\tilde{i}=1,\ \tilde{i} \neq i}^{k} w_{\cdot \tilde{i}} h_{\tilde{i} \cdot}$. The promising aspect of this $2k$-block partitioning is that each subproblem in (3) has a closed-form solution using Theorem 2 from [7]. The convergence of the algorithm is guaranteed if the blocks of W and H remain nonzero throughout all the iterations and the minima of (3) are attained at each step [7]. Finally, we include in Eq. 1 a regularization factor for H to induce sparse solutions, so that each document is modeled as a mixture of just a few topics. At the same time, we also add a regularization term on W to prevent its entries from growing too much, and add a prior to exploit available knowledge of the user. In particular, for each of the topics, s/he can identify some terms as relevant, or decide to exclude others. As a result, we obtain the following expression:

$$\min_{W,H\geq 0} ||A - WH||_F^2 + \phi(\alpha_p, W, P) + \psi(H), \tag{4}$$

where $\psi(H) = \beta \sum_{i=1}^{n} ||h_{i\cdot}||_1^2$ and $\phi(\alpha_p, W, P) = \sum_{i=1}^{m} \sum_{j=1}^{k} \alpha_{p_{ij}} (w_{ij} - p_{ij})^2 + \alpha ||W||_F^2$. The notation $h_{i\cdot}$ is used to represent the i-th row of H, the l_1 term promotes sparsity on the rows of H, while the Frobenius norm, (equivalent to l_2 regularization on the columns of W) prevents values in W from growing too large. The prior term P is an $m \times k$ matrix in which entries are either 1 for terms that, according to the available domain knowledge, should be assigned to a certain aspect and 0 otherwise. For example, if the prior terms list of the aspect Food contains the terms 'curry' and 'chicken', the values in the columns corresponding to that aspect in the rows corresponding to these terms are set to 1. Since these seed lists are expected to be very short, matrix P will be a sparse matrix. The values in α_p serve as normalizing factors for the element-wise difference between W and P and to activate/deactivate the prior on a specific term for any of the k aspects. In other words, for what concerns matrix P, if $p_{ij} = 1$ we are suggesting to assign i-th term to aspect j. On the contrary, if $p_{ij} = 0$ we want the i-th term not to be assigned that aspect. This allows us, by manipulating the values in α_p and P, to choose to what extent we want to influence the link between certain topics and the seed terms. To the best of our knowledge, the regularization term based on P is novel and distinguishes our approach from other similar techniques such as [3]. In the same vein as [6], the new update formula for matrices W and H can be obtained in closed form as

$$\begin{aligned} &w_{\cdot k} \leftarrow \frac{[\nu]_+}{\alpha_{p_{\cdot k}} + (HH^T)_{kk}}, \quad h_{i\cdot}^T \leftarrow \left[h_{i\cdot}^T + \xi\right]_+, \\ &\nu = (AH^T)_{\cdot k} - (WHH^T)_{\cdot k} + W_{\cdot k}(HH^T)_{kk}\alpha_{p_{\cdot k}} \odot P_{\cdot k} - 0.5\alpha 1_m, \\ &\xi = \frac{(A^T W)_{\cdot i} - H^T((W^T W)_{\cdot i} + \beta 1_k)}{(W^T W)_{ii} + \beta}. \end{aligned} \tag{5}$$

Here, $\odot$ indicates an element-wise product, $[x]_+ = \max(0, x)$, 1_t indicates a vector of ones of length t and the division in the fraction is element-wise. After we performed the factorization of matrix A into the factors W and H, we normalize each column of matrix H, then, in order to identify the set of relevant topics in a document, we set an Aspect Detection Threshold (ADT) to an appropriate value – according to the total number of considered topics in the data set. Finally, we associate a document to a topic if the corresponding weight in matrix H is greater or equal to the chosen threshold.

Indexing of the Collection. The initialization of the term-document matrix A is a crucial aspect that is often overlooked in the related literature on NMF.

Whenever prior knowledge on the topics is available, we should make sure that all relevant elements with respect to these topics appear in A. Hence, we propose a novel method based on short seed lists of terms, the same used by the user to influence the aspect classification task. In particular, we select the set of terms $\mathcal{D}$ used to index the collection by means of three steps:

1. We add all seed words that appear at least once in our collection to $\mathcal{D}$;
2. Since we do not assume to have complete domain knowledge, the seed lists are extended automatically by means of Word2Vec and included in $\mathcal{D}$;
3. We compute the TF-IDF weight for each term in the collection and add the top few hundreds to $\mathcal{D}$.

As for the Word2Vec model used in the second step, no additional source of information is used because the model is trained on the same data set that has to be analyzed. Even if the latter is not very rich, it turns out that Word2Vec is still capable of mapping seed words close to other words coming from the same topic. More details are provided in Sect. 4.1.

W and H Matrices Initialization. There are different approaches to initialize the matrices W and/or H in the BCD framework and their initialization deeply impacts on the achieved solution. In the context of topic modeling, since we assume that we have some a priori knowledge on which terms should be associated to a specific topic/aspect in the form of words lists, we initialize the matrices according to this knowledge. We begin by extending these lists using a Word2Vec model trained on the same collection of documents that we have to analyze. Specifically, for each term in each list, we add the two closest terms in the Word2Vec model. Then, we perform a search of the few terms from the extended lists corresponding to each topic in each sentence of the collection, and set the corresponding elements of matrix H to 1 or 0 if the sentence contains one of them or not, respectively. We apply the same procedure for the term-topic matrix W. The only pre-processing step involved prior to the training of the Word2Vec model is stopwords removal.

3.2 Sentiment Polarity Classification

After the identification of the relevant aspects in each sentence, we compute the polarity for each of them, classifying the corresponding opinions as positive or negative. Again, we propose a weakly-supervised approach articulated as follows:

- we manually compile two lists of seed terms, one for the positive and one for the negative sentiment terms for each aspect;
- we extend the previously created lists using a Word2Vec model, as we did for document indexing (see Sect. 3.1);
- for each document in our data set, we run a pre-processing step that involves stemming and stopwords removal[1] and we do the same on each extended sentiment terms list;

[1] In this work we employed the stopwords and stemmers provided in Python nltk 3.2.5, https://www.nltk.org.

– finally, we look for these sentiment terms in each document, considering only the sentiment terms relative to the most relevant topics identified in it.

In order to compute the polarity for each aspect, we average over the number of positive and negative terms found in relation to that topic, weighting 1 the positive terms and -1 the negative ones. The final label assigned to the opinion will depend on the sign of this score. We found this approach to be the best performing to extend the seed lists available. We also implemented a simple negation detection system that employs a short manually compiled list of negation terms that can flip the polarity of a word[2]. If we detect one of these terms in a range of 3 tokens before a sentiment term, we flip its polarity. For other languages which follow similar strategies to indicate a negation, this rule can be easily modified to comply with the new structure. In [25], it is shown that the use of negation in these terms can be easily transferred to other languages.

4 Experimental Results

We evaluate WS4ABSA on public data sets[3]:

– 2016 Track 5 Subtask 1 [15], training data set of Restaurant reviews, in English [Rest-EN] (1152 documents, 18779 tokens, 1.18 labels on average for each document);
– 2016 Track 5 Subtask 1 [15], training data set of Restaurant reviews in Spanish [Rest-ES] (1047 documents, 21552 tokens, 1.35 labels on average for each document);
– 2015 Track 12 [16], test data set of Hotel reviews in English [Hotels] (86 documents, 1316 tokens, 1.10 labels on average for each document).

For the aspect classification task, we focused firstly on Restaurant reviews [Rest-EN, Rest-ES] and considered the following aspects:

– Ambiance: the atmosphere or the environment of the restaurant's interior or exterior space;
– Food: the food in general or specific dishes;
– Service: the customer/kitchen/counter service or the promptness and quality of the restaurant's service in general.

The corresponding entities in SemEval data sets are shown in Table 1.

[2] In this work we will consider corpora in English and Spanish (see Sect. 4); the lists of negation terms for English (16 terms) and Spanish (12 terms) are included in our code repository https://gitlab.dei.unipd.it/dl_dei/ws4absa.

[3] The code for deploying and evaluating WS4ABSA is available on https://gitlab.dei.unipd.it/dl_dei/ws4absa.

Table 1. Aspects definitions for the aspect classification task.

Data	Aspect	Labels in SemEval data sets
Rest.	Ambiance	AMBIENCE#GENERAL
	Food	FOOD#PRICES, FOOD#QUALITY, FOOD#STYLE
	Service	SERVICE#GENERAL
Hotel	Ambiance	FACILITIES#DESIGN_FEATURES, ROOMS#DESIGN_FEATURES, HOTEL#DESIGN_FEATURES
	Food	FOOD_DRINKS#PRICES, FOOD_DRINKS#QUALITY, FOOD_DRINKS#STYLE_OPTIONS
	Service	SERVICE#GENERAL

4.1 Training of the Word2Vec Model

Before diving into the experimental results, we report here how we use the available prior knowledge. The word lists provided in input by a user are extended employing a Word2Vec model trained on the data set we are currently analyzing. Even if this model is not an accurate representation of the relations between terms in the considered language in general, we found it good enough for our goal of adding terms related or used in the same context of the available seed terms. We employ these extended lists of terms for document indexing – together with the features selected with TF-IDF – and document classification, with the assumption that words used in the same context are relevant for the same aspect. An example of the resulting word lists obtained with this technique is reported in Table 2. The model has been trained[4] on each collection using the Continuous Bag-Of-Words (CBOW) training algorithm for 10 epochs, generating word embeddings of size 300.

4.2 Evaluation of Aspect Classification

Initially we tackle aspect classification for English reviews. The hyperparameters used in the NMF optimization are listed in Table 3, while the seed lists used to perform aspect classification on the [Rest-EN] data set are reported in Table 4.

Since LDA-based methods are the main alternative to NMF-based ones for the unsupervised document classification task, we choose to compare our approach to two other weakly-supervised methods, LocLDA [2] and ME-LDA [26]. These were developed under the assumption that each sentence is assigned to a single aspect. Thus, we compare WS4ABSA with them on the subset of [Rest-EN] sentences with a single aspect label (972 sentences). The comparison is reported in Table 5. While LocLDA and ME-LDA outperform WS4ABSA in most of the cases, we remark that, differently from WS4ABSA, these approaches

[4] Word2Vec implementation from https://radimrehurek.com/gensim/models/word2vec.html.

Table 2. [Rest-EN]: A few of the terms obtained by extending the English seed lists for aspect classification (from Table 4) with Word2Vec.

Aspect	Seeds included with Word2Vec
Ambiance	Cheap, classic, clean, describe, interesting, Italy, looks
Food	Drip, dumplings, oil, pay, perfect, price, sausages, starter, vegetarian
Service	Happy, help, hookah, hours, personality, professional, recommend, service

Table 3. Hyperparameters used in NMF for the aspect classification task, the ADT, and the number of terms selected with TF-IDF weights for the document indexing (TDI), for each of the considered test data sets. These parameters have been obtained with a grid search over a portion of the dataset, used for validation.

Data set	α	β	α_p	ADT	TDI
[Rest-EN]	1.00	0.10	1.00	0.13	200
[Rest-ES]	0.01	1^{-16}	0.10	0.16	300
[Hotels]	1^{-3}	1.00	1.00	0.19	200

Table 4. Seed lists employed for the aspect classification task in the [Rest-EN] and [Hotels] data sets.

Aspect	Seeds
Ambiance	Bad, beautiful, big, ceilings, chic, concept, cool, cozy, cramped, dark, decor, elegant, expensive, interior, lightning, loud, modern, nice, noisy, setting, trendy, uninspired, vibe, wall
Food	Beef, chewy, chicken, crispy, curry, drenched, dry, egg, groat, moist, onions, over-cooked, pizza, pork, red, roasted, seared, shrimp, smoked, soggy, sushi, tender, tuna, undercooked
Service	Attentive, chefs, efficient, employees, helpful, hostess, inattentive, knowledgeable, making, manager, owner, packed, polite, prompt, rude, staff, unfriendly, wearing, workers

heavily rely on additional resources and can be used only in a single-label context. In particular, in [2] and [26] the authors compute a topic model with 14 topics first, then they examine each of them manually and assign a label to them according to the aspects provided in input. Thus, whenever a new dataset is considered, human inspection of the topic modeling results is required to choose the correct number of topics to use. In addition, in LocLDA and ME-LDA, the discovered topics have to be manually linked to the aspects under examination, while this is not necessary in WS4ABSA, where seed words define the aspects. Moreover, the methods in [2] and [26] both employ some language-dependent resources such as Part-Of-Speech (POS) taggers to identify adjectives in sentences and improve the identification of aspects. Furthermore, in ME-LDA, the authors also employ an annotated dataset to train a Maximum Entropy (ME) classifier. On the contrary, WS4ABSA requires no additional resources beyond the dataset but a list of seed words based on available domain knowledge. Moreover, it is also suitable to deal with the more general multi-label assumption for aspect extraction (indeed, as mentioned above, the average number of labels per sentence is always greater that one in our dataset).

4.3 Evaluation of Sentiment Polarity Classification

For the sentiment polarity classification task, we first perform our experiments on [Rest-EN] data set. We formalize sentiment polarity classification as a single-label multi-class classification problem and used the seed words listed in Table 7. Table 8 describes the results of the sentiment polarity classification task, obtained on the Restaurants data sets. We computed these results considering only the opinions which were classified correctly in the previous aspect classification stage. Our performance results in this task are aligned with other state of the art approaches [16] but stand out for the independence from external resources and for the high language flexibility. As expected, negative polarity is the most challenging to detect. However, we highlight that the we rely on extremely simple rules, described in Sect. 3.2, that may be further enriched to achieve better performance.

Table 5. Aspect classification performance on [Rest-EN] data set, considering the documents with a single relevant aspect in the performance evaluation. We remark that the amount of resources used in our approach is lower than the other methods included in the comparison. Indeed, we only required a short list of seed words defining aspects, while the other two methods are based on language-specific POS tagging, additional annotated data sets and manual topic inspection to retrieve aspects.

	WS4ABSA	LocLDA	ME-LDA
Ambiance: Precision	0.21	0.60	**0.77**
Recall	**0.79**	0.68	0.56
F1 score	0.33	0.64	**0.65**
Food: Precision	0.79	**0.90**	0.87
Recall	0.53	0.65	**0.79**
F1 score	0.64	0.75	**0.83**
Service: Precision	**0.88**	0.80	0.78
Recall	0.39	**0.59**	0.54
F1 score	0.54	**0.68**	0.64
Overall: Precision	0.74	0.77	**0.81**
Recall	0.52	**0.64**	0.63
F1 score	0.56	0.69	**0.70**

Table 6. Terms not present in the seed lists assigned by NMF to the chosen aspects from the Hotels data set.

Term	Ambiance	Food	Service
Curtain	1	≈0	≈0
Pool	1	≈0	≈0
Breakfast	≈0	1	≈0
Buffet	0.03	0.93	0.03
Response	≈0	≈0	1
Management	≈0	≈0	1

4.4 Domain Flexibility Evaluation

To assess the flexibility of WS4ABSA, we use the seed lists defined on Restaurants to perform aspect classification on another data set with similar topics coming from a different domain, i.e. Hotels. The results, shown in Table 9, suggest that WS4ABSA is able to generalize the information provided by seed words to similar aspects from different domains. In this case we consider aspect classification as a multi-label classification problem [19]. This is a more general and challenging scenario. Thus, we measure accuracy in this task by means of the Jaccard index $\text{J} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i \wedge y_i|}{|\hat{y}_i \vee y_i|}$, where N is the total number of samples that have been evaluated (in order to compute the average), $\hat{y}_i$ is a binary vector that is 1 only in the positions corresponding to the aspects predicted for the i-th sample and y_i is another binary vector that is 1 only in the positions corresponding to the true aspects to associate to the i-th sample. These results may be explained by the fact that the method is able to leverage partial prior information, i.e. the seeds play a key role in defining final topics, but they can also be extended automatically to other terms in the collection, if this improves the quality of the factorization. Indeed, recall that we have an active penalization term on W_{ij}, related to the prior, only if domain knowledge suggests that term i is relevant for topic j. Then, we induce a penalization policy that only acts on a subset of the entries of W, denoted by $\mathcal{S} := \left\{(i,j)\,|\, i \in \bar{I},\, j \in \bar{J}\right\}$ for $\bar{I}$ and $\bar{J}$ defined based on prior knowledge on topics. No penalization is imposed for entries $\{W_{i,j}\,|\,(i,j) \notin \mathcal{S}\}$. This approach differs from the penalization strategy adopted by methods, such as Utopian [3], that allows the user to include domain knowledge, but embed it in the form of a distribution over all the available terms. If we assume to set equal to zero all the elements of P corresponding to positions $(i,j) \notin \mathcal{S}$, and we impose a topic-wise penalization, such as in Utopian, i.e.

$$H, W = \underset{H \geq 0, W \geq 0}{\operatorname{argmin}} \; ||A - WH||_F^2 + ||(W-P)D||_F^2,$$

with D diagonal matrix of weights, we force the algorithm towards solutions which do not assign new words to a topic for which seed words were already provided. In fact, the results of this test led to achieve an accuracy of just 0.30 on the [Rest-EN] data set. Therefore, we infer that WS4ABSA can work well

Table 7. Seed lists employed for the sentiment polarity classification task in the [Rest-EN] and [Hotels] data sets.

Aspect	Polarity	Seeds
Ambiance	Positive	beautiful, chic, cool, cozy, elegant, modern, nice, trendy, winner
	Negative	bad, beaten, big, cramped, dark, expensive, loud, noisy, uninspired
Food	Positive	crispy, groat, moist, red, roasted, seared, smoked, tender, winner
	Negative	beaten, chewy, drenched, dry, over-cooked, soggy, undercooked
Service	Positive	attentive, efficient, helpful, knowledgeable, polite, prompt, winner
	Negative	beaten, inattentive, making, packed, rude, unfriendly, wearing

Table 8. Performance results in the sentiment classification task on the Restaurants data set for the Positive and Negative polarities.

	[Rest-EN]	[Rest-ES]
Accuracy	0.85	0.57
Precision (Pos.)	0.86	0.96
Recall (Pos.)	0.89	0.48
Precision (Neg.)	0.84	0.31
Recall (Neg.)	0.80	0.92

Table 9. Aspect classification performance in multi-label classification.

	[Rest-EN]	[Rest-ES]	[Hotels]
Accuracy	0.58	0.52	0.60
Average precision	0.52	0.48	0.58
Average recall	0.74	0.55	0.67
F1 score	0.61	0.51	0.62

with much weaker supervision than Utopian. In this regards, in Table 6, which shows a few rows of the term-topic matrix W, we can see that the algorithm includes some terms that were absent from the initial seed lists. Thus, the proposed classification generalizes well the initial information that was provided in the seed lists.

4.5 Language Flexibility

To assess the flexibility of WS4ABSA with regard to different languages, we considered [Rest-ES] data set and used as seed lists the same words used for the English data set, – see Tables 4 and 7 – translated when necessary with Google Translate. The resulting seed words are shown in in Tables 10 and 11. The results, described in Tables 8 and 9, suggest that our approach can be straightforwardly adapted to different domains and languages by translating the terms in the seed lists with a machine translation system. This is a simple method to leverage the same prior knowledge for cross-domain and cross-languages applications. As for aspect classification task, the performance on [Rest-ES] is very close to what we obtained on [Rest-EN]. As it might be expected, we notice a decrease in the average recall in this case, since some terms that might be very frequently used in a language in a specific context might not be used as frequently in other languages. We see the same effect on the recall of the positive class in Table 8. The impact of the machine translation of seed terms is lower on the recall of the negative class in the same table because we employ a set of terms to recognize negations which was compiled manually based on basic Spanish grammar rules. In fact, these terms would not have been easy to obtain by automatically translating the ones in the list we employed for the datasets in English. We expect that fine tuning of seed words and negation rules could further improve the performance of WS4ABSA. Yet, experiments suggest that an almost automatic adaptation to a different language achieves acceptable performance.

Table 10. Seed lists employed for the aspect classification task in the [Rest-ES] data set. These have been obtained by translating the ones in Table 4.

Aspect	Seeds
Ambiance	Acogedor, ambiente, apretado, bonito, caro, chic, concepto, decoración, elegante, escenario, fuerte, genial, grande, hermoso, interior, malo, moderno, no inspirado, oscuro, pared, relámpago, ruidoso, techos
Food	Ahumado, atún, camarón, carne de res, cauterizado, cebolla, cocido, crujiente, curry, empapado, groat, huevo, húmedo, masticable, mojado, pizza, pollo, puerco, rojo, seco, sobre cocinado, sushi, tierno, tostado
Service	Anfitriona, antipático, atento, cocineros, conocedor, cortés, eficiente, embalado, empleados, falta de atención, gerente, grosero, personal, propietario, rápido, servicial, trabajadores, usar

4.6 Impact of Initialization Policy

We also analyzed how the initialization of matrices W and H affects the results of aspect classification. In particular, we compared the policy described in Sect. 3.1 with 50 random initializations of the matrices W and H in order to evaluate the improvement of our initialization strategy in the classification task on the [Rest-EN] data set. With a random initialization, we obtained an average accuracy in the multi-label classification problem of 0.36, while the proposed initialization approach leads to an accuracy of 0.58, with an improvement of 38%[5] compared to a random initialization of the matrices on average. Furthermore, we also tested our method for feature extraction, i.e. for the document indexing process and

[5] The difference was computed as: difference $\div$ other value $*$ 100.

Table 11. Seed lists employed for the sentiment polarity classification task in the [Rest-ES] data set. These have been obtained by translating the ones in Table 7.

Aspect	Polarity	Seeds
Ambiance	Positive	Acogedor, agradable, chic, de moda, elegante, ganador, genial, hermoso, moderno
	Negative	apretado, caro, fuerte, golpeado, grande, malo, oscuro, ruidoso, sin inspiración
Food	Positive	Ahumado, cauterizado, crujiente, ganador, groat, húmedo, rojo, tierno, tostado
	Negative	batido, demasiado cocido, masticable, mojado, poco cocido, seco
Service	Positive	Atento, conocedor, educado, eficiente, ganador, rápido, útil
	Negative	Antipático, desgastado, embalado, fabricación, falta de atención, golpeado, grosero

the creation of the A matrix. In particular, we compared the results obtained by following the initialization procedure described in Sect. 3.1 with simple TF-IDF initialization on the same [Rest-EN] data set. As a result, we obtained an accuracy of 0.38. Hence noticing a performance improvement with our new feature selection approach of 34%.

5 Conclusions and Future Directions

We propose Weakly-Supervised Approach for ABSA (WS4ABSA), a weakly-supervised approach for ABSA based on NMF that allows users to include domain knowledge in a straightforward fashion by means of short seed lists. Thus, we address one of the drawbacks of most of the available topic modeling strategies, i.e. the fact that the beneficiary of the results is not able to improve them. WS4ABSA can be easily adapted to other domains or languages, as suggested by tests performed on publicly available data sets, and achieves performance comparable with other weakly and semi-supervised approaches in the literature, even though it relies on less external resources. Future research directions include deeper investigations on the effect of the prior on W and possibly H, and also the release of simple rules to deal with negations more effectively in different languages. It might also be useful to implement an on-line version of the NMF classification algorithm, so that it can receive a feedback from the user, and recompute the output on-the-fly more efficiently, i.e. without running again from scratch.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal Mach. Learn. Res. **3**, 993–1022 (2003)
2. Brody, S., Elhadad, N.: An unsupervised aspect-sentiment model for online reviews. In: Human Language Technologies: The 2010 Annual Conference of the North

American Chapter of the Association for Computational Linguistics, pp. 804–812. Association for Computational Linguistics (2010)
3. Choo, J., Lee, C., Reddy, C.K., Park, H.: Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. IEEE Trans. Vis. Comput. Graph. **19**(12), 1992–2001 (2013)
4. Cichocki, A., Phan, A.H.: Fast local algorithms for large scale nonnegative matrix and tensor factorizations. IEICE Trans. Fundam. Electron., Commun. Comput. Sci. **92**(3), 708–721 (2009)
5. García-Pablos, A., Cuadros, M., Rigau, G.: W2vlda: almost unsupervised system for aspect based sentiment analysis. Expert. Syst. Appl. **91**, 127–137 (2018)
6. Kim, H., Park, H.: Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. SIAM J. Matrix Anal. Appl. **30**(2), 713–730 (2008)
7. Kim, J., He, Y., Park, H.: Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. J. Glob. Optim. **58**(2), 285–319 (2014)
8. Kuang, D., Choo, J., Park, H.: Nonnegative matrix factorization for interactive topic modeling and document clustering. In: Celebi, M.E. (ed.) Partitional Clustering Algorithms, pp. 215–243. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-09259-1_7
9. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems, vol. 15. SIAM, Philadelphia (1995)
10. Li, T., Sindhwani, V., Ding, C., Zhang, Y.: Bridging domains with words: Opinion analysis with matrix tri-factorizations. In: Proceedings of the 2010 SIAM International Conference on Data Mining, pp. 293–302. SIAM (2010)
11. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
12. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-vol. 1, pp. 142–150. Association for Computational Linguistics (2011)
13. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
14. Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics **5**(2), 111–126 (1994)
15. Pontiki, M., et al.: Semeval-2016 task 5: aspect based sentiment analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 19–30 (2016)
16. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: Semeval-2015 task 12: aspect based sentiment analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 486–495 (2015)
17. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. **24**(5), 513–523 (1988)
18. Toh, Z., Su, J.: 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), pp. 282–288 (2016)
19. Tsoumakas, G., Katakis, I.: Multi-label classification: an overview. Int. J. Data Warehous. Min. **3**(3), 1–13 (2006)

20. Varghese, R., Jayasree, M.: Aspect based sentiment analysis using support vector machine classifier. In: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1581–1586. IEEE (2013)
21. Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM J. Optim. **20**(3), 1364–1377 (2009)
22. Wang, F., Li, T., Zhang, C.: Semi-supervised clustering via matrix factorization. In: Proceedings of the 2008 SIAM International Conference on Data Mining, pp. 1–12. SIAM (2008)
23. Xiang, B., Zhou, L.: Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 434–439 (2014)
24. Yan, X., Guo, J., Liu, S., Cheng, X., Wang, Y.: Learning topics in short texts by non-negative matrix factorization on term correlation matrix. In: Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 749–757. SIAM (2013)
25. Zagibalov, T., Carroll, J.: Automatic seed word selection for unsupervised sentiment classification of chinese text. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1, pp. 1073–1080. Association for Computational Linguistics (2008)
26. Zhao, W.X., Jiang, J., Yan, H., Li, X.: Jointly modeling aspects and opinions with a maxent-lda hybrid. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 56–65. Association for Computational Linguistics (2010)