Aspect-based Sentiment Analysis using Semi-supervised Learning in Bipartite Heterogeneous Networks

Ivone P. Matsuno[1,2], Rafael G. Rossi[2], Ricardo M. Marcacini[2], Solange O. Rezende[1]

[1] Institute of Mathematics and Computer Science - University of São Paulo, Brazil
ivone.matsuno@usp.br,rezende@icmc.usp.br
[2] Federal University of Mato Grosso do Sul, Brazil
{rafael.g.rossi,ricardo.marcacini}@ufms.br

Abstract. Aspect-Based Sentiment Analysis (ABSA) allows to analyze the sentiment from each product aspect, e.g., the camera quality, operating system and the storage capacity of a smartphone. Two main tasks to perform ABSA are: (i) the terms/words related to the aspects and (ii) performing sentiment analysis for each identified aspect. Several approaches to treat these tasks are found in the literature, such as those based on the sentiment lexicon, syntactical relations and topic models. The main disadvantages of these methods are the time required and the need of specialists to build the lexicon or to define the rules for different languages and domains. Alternatively, supervised machine learning techniques are employed to perform both aspect identification and sentiment analysis of the extracted aspects. Although these techniques are language and domain independent and avoid the use of a predefined lexicon or the manual building of rules, the use of supervised learning for ABSA requires labeling a significant number of aspects and their polarities, which limits the use of such approach in real applications. In this article we propose an approach to perform ABSA through semi-supervised learning techniques, i.e., significantly less amount of labeled data are required to perform learning. This makes the use of our proposal to perform ABSA easier in real applications. In our proposal, we model the data into networks to perform the semi-supervised learning. Specifically, we propose the use of bipartite networks to represent the data, since network-based approaches have been successfully used to perform semi-supervised learning and the bipartite networks are parameter-free and fast to be generated. So, this type of network is advisable to be used in practical situations. The results obtained in a rigorous experimental evaluation demonstrate that the proposed approach for ABSA obtains better results than existing approaches based on machine learning for ABSA.

Categories and Subject Descriptors: G.2.2 [**Graph Theory**]: Graph Labeling; H.2.8 [**Database Applications**]: Data Mining; H.2.4 [**Systems**]: Textual Databases

Keywords: aspect-based sentiment analysis, heterogeneous network, machine learning, opinion mining, sentiment analysis

1. INTRODUCTION

Traditional approaches for Sentiment Analysis (SA) aims to classify the sentiment polarity of textual documents as positive, negative or neutral. Usually, this classification is performed considering the documents as a single unit (document-level sentiment analysis) or considering each sentence of a document (sentence-level sentiment analysis) [Liu 2012]. In both cases, information about the aspects, i.e., features or properties of products or services, is not explored. However, the sentiment polarity might be different for different aspects of the same product or service [Chen et al. 2014]. For instance, in the following sentence "*I liked the image resolution of the TV, but its remote control is terrible*", there is a positive opinion about the *image resolution* and a negative opinion about the *remote control*. To deal with this type of scenario, we can use Aspect-Based Sentiment Analysis to enhance decision making through specific information on the sentiment polarity of each aspect of a product or service.

The authors thank FAPESP (Process 2014/08996-0 and 2011/12823-6), PROPP/UFMS, CNPq, CAPES, FUNDECT-MS(Process 147/2016) and FINEP for the financial support.

ABSA is more challenging and complex than document-level or sentence-level sentiment analysis [Liu 2012; Jiménez-Zafra et al. 2015]. Despite the many steps that can be carried out to perform ABSA, we can divide it into two main steps: (i) the terms/words related to the aspects, which carry the information about the sentiment must also be identified and (ii) performing sentiment analysis for each identified aspect. Promising researches about ABSA combine Natural Language Processing (NLP) and Machine Learning (ML) techniques for both aspect extraction and sentiment analysis [Liu 2015; Ganeshbhai and Shah 2015]. Unsupervised learning is applied for sentiment analysis considering Topic Modeling techniques, such as Probabilistic Latent Semantic Analysis (PLSA) [Hofmann 1999] and Latent Dirichlet Allocation (LDA) [Blei et al. 2003], to generate a *Topic-Sentiment Model* [Jiménez-Zafra et al. 2015; Moghaddam and Ester 2012; Zhao et al. 2010; Titov and McDonald 2008a; 2008b]. In such approach, each topic is composed by of set of terms. Then, predefined list of aspects and other predefined list of positive and negative terms (words and expressions) are used to define the sentiment polarity. The disadvantage of this approach is to define the correct number of topics and the need of a predefined lexicon resource. Moreover, the previous definition of lexicon resource is often problematic in real scenarios. For instance, the word "*fastly*" has a positive sentiment for "*boot time of the operating system*" and a negative sentiment for "*cell battery discharge*". Thus, the polarity of the word may change for different contexts of domains [Zha ; Bollegala et al. 2011].

Machine learning-based methods for ABSA are potentially useful alternatives to overcome these limitations. In this case, the goal is to learn a model that automatically extracts the aspects from the texts, by identifying the importance of the lexicons for the context domain, linguistic features, and syntactic relations of the features [Liu 2012; Chen et al. 2014]. Although supervised machine learning techniques provide better results [Liu 2015; Ganeshbhai and Shah 2015], such approaches require a huge human effort to label a large number of aspects and their sentiments to learn a model [Chen et al. 2014; Pontiki et al. 2014]. In this article we propose a semi-supervised based approach to perform ABSA, that allows learning a model using few labeled data, which makes it more applicable to real-world scenarios. In our approach, named ASPHN (*Aspect-Based Sentiment Propagation on Heterogeneous Networks*), we proposed a semi-supervised learning technique based on graph models, in which linguistic features, candidate aspects and sentiment labels are modeled by means of a heterogeneous network. Specifically, we considered bipartite heterogeneous networks to model the data. This type of network is parameter-free, quickly generated, and has presented competitive or better results than the other types of networks [Rossi et al. 2014; Rossi et al. 2014]. In the proposed ASPHN we generated two bipartite networks. To identify the aspects we generated a bipartite network composed of two types of objects: linguistic features and aspect candidates. Finally, to analyze the sentiment of the aspects, we generated a bipartite network composed of the aspects identified in the previous step and terms that will be processed to identify the sentiment (positive, negative, or neutral) of the aspect. For both networks, we label few objects according to the task and perform label propagation through the network to classify the remaining objects.

Some concepts of this work have been briefly introduced in a previous work [Matsuno et al. 2015] and we present a substantially improved version in this article. The major improvements are: (i) we included more semi-supervised learning algorithms, based on bipartite heterogeneous networks, in the proposed approach and compare them in the experimental evaluation, (ii) we included more supervised learning algorithms in comparison with the proposed approach to demonstrate how semi-supervised learning can be useful to perform ABSA, and (iii) we presented an extension about the literature review and the concepts about the proposed approach.

We conduct a robust comparative evaluation of the proposed ASPHN approach with traditional and state-of-the-art methods. The evaluation carried out in this article highlights the benefits about the use of bipartite networks and semi-supervised learning to perform both aspect identification and sentiment analysis. We also present the performance of aspect identification and sentiment analysis for a different range of labeled documents. The obtained results show that our proposal obtained better results than traditional semi-supervised learning or supervised learning algorithms, even using very few labeled examples.

The remainder of this article is organized as follows. Section 2 presents background and related works about the main tasks to perform Aspect-based Sentiment Analysis (ABSA). Section 3 presents the details of the proposed ASPHN to perform aspect identification and sentiment analysis through semi-supervised learning based on bipartite heterogeneous networks. Section 4 presents details of the experimental evaluation and the results. Finally, Section 5 presents the conclusions and points to future work.

2. BACKGROUND AND RELATED WORK

Aspect-Based Sentiment Analysis (ABSA) can be defined as a process to extract a set of opinions expressed in a set of documents (reviews, tweets, news) where each opinion O is represented by the triple $O = (e_i, a_{ij}, s_{ij})$, in which e_i is an entity i representing a product, service, personality, institute or organization, $a_{i,j}$ is an aspect j (feature or property) of the entity e_i, and $s_{i,j}$ is the polarity (e.g., positive, negative, or neutral) of the sentiment of the aspect $a_{i,j}$[1] [Liu 2012; Jiménez-Zafra et al. 2015].

ABSA requires several previous "tasks" such as feature extraction, entity identification and categorization, aspect identification and categorization, and aspect sentiment classification. However, these tasks can be organized into two major steps: (i) identifying which terms/words are aspects and (ii) performing sentiment analysis for each aspect individually. In this section, we present the work related to these two tasks.

2.1 Aspect Identification

The goal of the aspect identification task is to extract or identify the terms that indicate entity features related to the aspects. Some approaches are based on the frequency in which words or terms occur in the documents. These approaches consider that the most common nouns are used to generate a set of candidate aspects and propose a selection criterion to restrict this set [Hu and Liu 2004; Popescu and Etzioni 2005; Long et al. 2010]. These restrictions are different techniques, such as part-of-speech, distance between candidates aspects, co-occurrence of the candidates aspects and sentiment words [Hu and Liu 2004; Zhuang et al. 2006], the measure of PMI (Point-wise Mutual Information) which can be used to calculate the frequency in a word candidate aspect to any term that discriminates aspect [Popescu and Etzioni 2005], and the information distance based on Google [Cilibrasi and Vitanyi 2007].

In some studies, syntactic relations of words are explored when the words and sentiment expressions are known *a priori*. For instance, if the documents do not contain a frequent aspect, but contain some sentiment words, the nearest noun (or noun phrase) of the sentiment word may be extracted as an aspect [Hu and Liu 2004; Zhuang et al. 2006; Kobayashi et al. 2007; Somasundaran et al. 2009; Yang and Cardie 2013]. The requirement for a list of sentiment words of the application domain and the need of an extractor (parser tool) of syntactic relations for a particular language are major limitations of these approaches.

Some approaches use probabilistic topic model for aspect identification [Mei et al. 2007; Guo et al. 2009; Titov and McDonald 2008a; 2008b; Lu et al. 2009; Moghaddam and Ester 2010; Wu and Ester 2015]. The two main widely-used probabilistic topic model algorithms are *Latent Dirichlet Allocation* (LDA) [Blei et al. 2003] and *Probabilistic Latent Semantic Indexing* (PLSI) [Hofmann 1999]. Zhao et al. [2010] proposed a hybrid method that combines Maximum Entropy and LDA to identify words that are aspects and identify sentiments words to specific aspects. However, the probabilistic topic

[1]In this article we consider that the entity is previously defined, in a way that the goal is to identify the aspects and their polarities for an specific entity [Pang and Lee 2008; Liu 2012].

model has some limitations that hinder its use in Sentiment Analysis in real-life applications, because a lot of data is necessary to build the model [Blei et al. 2003; Wu and Ester 2015].

Machine learning approaches to identify aspects are usually based on classification algorithms. For instance, Ghani et al. [2006] formulate the aspect extraction as a classification problem and use Naïve Bayes to calculate the probability of each feature to identify an aspect. Analogously, Kobayashi et al. [2007] use inductive algorithms to classify the syntactic relation "aspect-of" by using a specific labeled dataset. Yu et al. [2011] uses noun phrases as aspect candidates and the SVM algorithm to build a model based on labeled data. Kovelamudi et al. [2011] designed a framework for aspect identification of a product by using the Wikipedia dataset to extract aspect candidates and classification algorithms are used for the identification of the aspects. A common limitation of these studies is that by using supervised learning methods, there is the need of a large labeled data that usually is not available for tasks involving ABSA.

2.2 Aspect Sentiment Classification

The task of aspect sentiment classification has two main approaches [Liu 2012]: (i) lexicon-based and (ii) machine learning-based. Lexicon-based approaches are typically unsupervised. To determine the sentiment orientation on each aspect in a text, lexicon-based approaches make use of the lexicon resource (words, phrases and expressions which represent sentiments), composition rules, sentiment aggregation function or a set of sentiment and relationships derived from the parse tree of the texts. Words that can modify or intensify sentiment, for example, the words *"no"*, *"but"*, *"very"*, *"little"* and other grammatical constructions that can affect sentiments can also be considered by lexicon-based approaches. In [Taboada et al. 2011; Liu 2010; Zhu et al. 2009; Ding et al. 2008; Wan 2008; Hu and Liu 2004; Kim and Hovy 2004] some lexicon-based approaches to aspect sentiment classification are presented. There are different lexicon techniques for different languages and the performance of these techniques is usually worse than the use of machine learning-based approaches [Ding et al. 2008; Cruz et al. 2013]. Besides, lexicon resources must be adapted to some specific domains. For example, the word *"hot"* in a domain about *food* indicates a positive sentiment, but the same word in domain about *beer* indicates a negative sentiment [Zha ; Bollegala et al. 2011].

Traditional machine learning techniques, such as Multinomial Naïve Bayes, Support Vector Machines, KNN, applied on vector space representations can be used to perform aspect sentiment classification, where terms can be used as features. In this case, each identified aspect is labeled as positive, negative or neutral and a machine learning technique is applied to learn a classification model and classify the sentiment of the unseen aspects. The main advantage about using machine learning algorithms is the fact that rules, the polarity of the words, and other useful characteristics to perform sentiment analysis, are obtained automatically given a set of labeled data. This fact allows the machine learning approach to be applied in ABSA for texts from different domains, goals and languages. However, most machine learning-based approaches consider supervised learning to perform sentiment analysis. Thus, the same limitations previously mentioned are also present in the aspect sentiment classification, i.e., requires a huge amount of labeled data to obtain an accurate classifier.

We highlight that labeling data for aspect-based sentiment analysis is time consuming and makes the use of supervised learning approaches unfeasible in practical situations. In this context, first we need to label which textual expressions are aspects for a particular application domain. Secondly, we need to label the phrases in which the aspects occurred to define the sentiment (positive, negative or neutral). Aiming to reduce this problem, semi-supervised machine learning has been used, since it requires few labeled data. Although it has presented good results in many applications, the use of semi-supervised learning has been little explored to address the problem of ABSA. Existing studies in this field investigate document-level sentiment analysis. According to the literature, there are few studies about aspect-based sentiment analysis with semi-supervised learning [Wan 2009; Dasgupta and Ng 2009; Zhou et al. 2010; He and Zhou 2011; 2011; Mukherjee and Liu 2012; Tang et al. 2015].

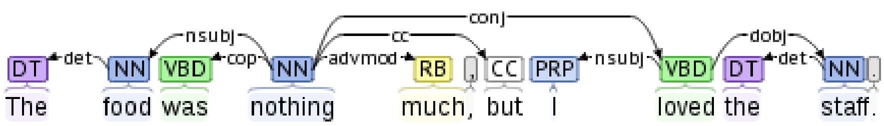


Fig. 1. Example of a syntactic tree extracted from the sentence *"The food was nothing much, but I loved the staff."*
.

3. PROPOSED APROACH: *ASPECT-BASED SENTIMENT PROPAGATION FOR HETEROGENEOUS NETWORKS* (ASPHN)

Our proposal, named Aspect-Based Sentiment Propagation for Heterogeneous Networks (ASPHN), makes use of heterogeneous networks to perform ABSA. We chose to model our problem as a heterogeneous network and perform label propagation since this combination of techniques has been successfully used in semi-supervised learning [Rossi et al. 2016; Ji et al. 2010]. Besides, as we will present in the next sections, such design of the proposed approach is advisable for ABSA through semi-supervised learning.

In fact, we make use of two bipartite heterogeneous networks, one for each step of ABSA presented in Section 2. In order to do so, we have to: (i) generate the bipartite heterogeneous networks in a way that allows the aspect identification and the aspect sentiment classification and (ii) perform label propagation in such network to classify the objects as `aspects` and `non aspects` (first step) and as `positive`, `negative` and `neutral` (second step). In the next section we detail how to generate the bipartite heterogeneous networks and how to perform semi-supervised learning through label propagation in such networks.

3.1 Bipartite Heterogeneous Network Generation

Formally, a bipartite heterogeneous network can be defined as $N = \langle \mathcal{O}, \mathcal{E}, \mathcal{W} \rangle$, in which $\mathcal{O}$ represents two sets of network objects (also called vertices or nodes), $\mathcal{E}$ represents the set of connections (also called relations or links) which occurs just from objects of one set to another set, $\mathcal{E}$ represents the set of connections (also called relations or links) among the objects, and $\mathcal{W}$ represents the weights of the connections.

The first bipartite heterogeneous network is composed by (i) aspect candidates $\mathcal{C} = \{c_1, ..., c_r\}$ and (ii) linguistic features $\mathcal{L} = \{l_1, ..., l_q\}$. An aspect candidate is connected to the linguistic feature if there is a relation between them. In our proposal, the aspect candidates can be nouns, verbs[2], adjectives and adverbs [Liu 2015]. Linguistic features are extracted from syntactic dependency trees of the sentences. For example, in the following sentence "*The food was nothing much, but I loved the staff.*", we can obtain the syntactic tree presented in Figure 1[3]. From this syntactic tree, we can extract the linguistic features considering each syntactic function or each syntactic relation as a feature. We removed the stop words (prepositions, articles, conjunctions and pronouns) from the sentences. The weight of the feature is equal to 1 if the aspect candidate enters into a syntactic relation, and 0 otherwise. Table I presents a vector space representation extracted from the syntactic tree presented in Figure 1. We differentiate the syntactic function features from the syntactic relation features using upper case and lower case respectively.

From the vector space representation, we can extract the bipartite heterogeneous network, which consists of a direct mapping from the vector space to the graph. In such case, we have two types of objects in the network: aspect candidates and linguistic features. An aspect candidate is connected to a linguistic feature if the weight in the vector space representation is equal to 1. In Figure 2 we

[2]Some state verbs are considered stop words, such as forms of the verb "to be".
[3]The syntactic tree presented in Figure 1 was generated using the *Stanford CoreNLP* toolkit [Manning et al. 2014].

Table I. Vector space representation extracted from the syntactic tree in Figure 1.

	NN	**RB**	**JJ**	**PR**	**DT**	**VB**	**CC**	**det**	**nsubj**	**cop**	**advmod**	**cc**	**conj**	**dobj**
food	1	0	0	0	0	0	0	1	1	0	0	0	0	0
nothing	1	0	0	0	0	0	0	0	0	1	1	1	1	0
much	0	1	0	0	0	0	0	0	0	0	1	0	0	0
staff	1	0	0	0	0	0	0	1	0	0	0	0	0	1

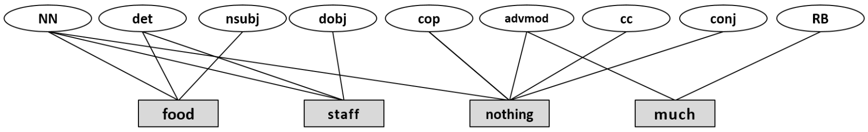


Fig. 2. Example of a bipartite heterogeneous network extracted from the vector space representation in Table I.

present an example of a bipartite heterogeneous network extracted from the vector space representation presented in Table I.

The second bipartite heterogeneous network is composed by (i) aspects $\mathcal{A} = \{a_1, ..., a_s\}$ (identified in the previous step), and (ii) terms $\mathcal{V} = \{v_1, ..., v_n\}$. The aspects are connected to the terms which are present in the same sentence. We highlight that we removed the stop words from the sentences. For instance, consider these three sentences: (i) *"The french fries were excellent. The drinks were terrible"*, (ii) *"In general, the snacks and french fries were good."*, and (iii) *"In general, drinks were terrible, mainly the wine."*. Considering our proposal, we generate a vector space representation as presented in Table II and from this representation, we extract a bipartite network as presented in Figure 3.

3.2 Semi-Supervised Learning based on Bipartite Heterogeneous Networks

Semi-supervised learning based on bipartite heterogeneous networks has presented satisfactory results compared to other types of networks [Ji et al. 2010; Yin et al. 2009], mainly when considering textual data [Rossi et al. 2016; 2014]. Besides the fact that the bipartite network is a direct map from a vector space representation, the use of such network allows to make use of graph regularization to perform learning. Graph regularization has been demonstrated to be the state-of-the art to perform semi-supervised learning [Rossi et al. 2016; Zhu and Goldberg 2009; Chapelle et al. 2010]. The idea behind semi-supervised learning based on bipartite networks is to classify one type of objects, called target objects, using the other type of object, bridge objects, to propagate class information among the target objects.

Table II. Vector space representation to aspects identified in the previous step in three sentences.

	excellent	**good**	**general**	**mainly**	**terrible**
french-fries	1	1	0	0	0
snacks	0	1	1	0	0
wine	0	0	1	1	1
drinks	0	0	0	1	1

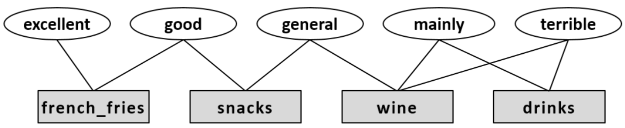


Fig. 3. Example of a bipartite heterogeneous network extracted from the vector space representation in Table II.

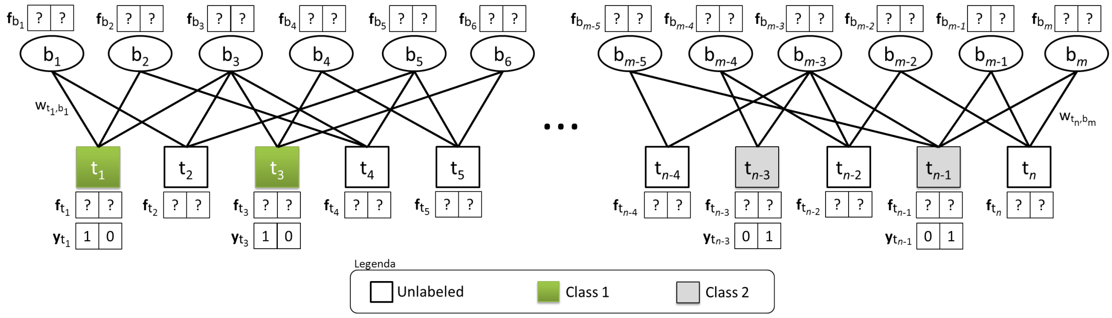
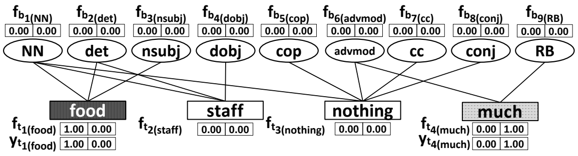
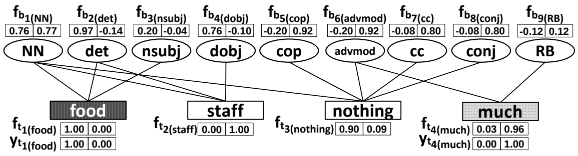


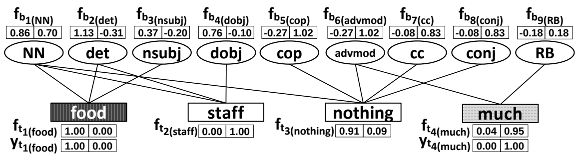
Fig. 4. Example of a bipartite network-based representation for semi-supervised learning.



(a) Initial state of the bipartite network structure for aspect identification.



(b) Intermediate state of the bipartite network.



(c) Final state of the bipartite network with weights assigned to all network objects.

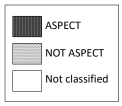
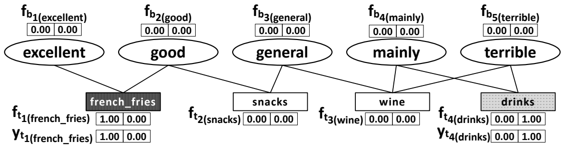


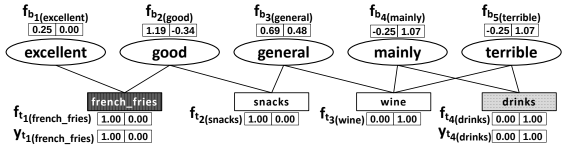
Fig. 5. Example of semi-supervised learning based on bipartite network considering the TCBHN algorithm with $\eta = 0.5$ to perform aspect identification.

In order to do so, let $\mathcal{C} = \{c_1, c_2, \ldots, c_l\}$ represent the set of class labels, let $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$ be the set of bridge objects, and let $\mathcal{T} = \{t_1, t_2, \ldots, t_n\}$ be the set of target objects. In a semi-supervised learning scenario, $\mathcal{T} = \mathcal{T}^L \cup \mathcal{T}^U$, in which $\mathcal{T}^L$ represents the set of labeled target objects and $\mathcal{T}^U$ represent the set of unlabeled target objects. Also, let $\mathbf{f}_{o_i} = \{f_{c_1}, f_{c_2}, \ldots, f_{c_{|\mathcal{C}|}}\}$ be the weight vector of an network object o_i which stores the weights of an object o_i for all classes in $\mathcal{C}$ assigned during semi-supervised learning. Hence it is also referred to as class information vector. Let $\mathbf{F}(\mathcal{O}) = \{\mathbf{f}_{o_1}, \mathbf{f}_{o_2}, \ldots, \mathbf{f}_{o_{|\mathcal{O}|}}\}^T$ be a matrix which stores all the weight vectors of the objects. The predefined labels for a target $t_i \in \mathcal{T}^L$ are stored in a weight vector $\mathbf{y}_{t_i} = \{t_1, t_2, \ldots, t_{|\mathcal{C}|}\}$,which has the value 1 in the position corresponding to the class of the target object t_i and 0 to the others. The weights of the links among network objects are stored in a matrix $\mathbf{W}$. Link weights among a network object o_i and other network objects are represented by a vector $\mathbf{w}_{o_i} = \{w_{o_1}, w_{o_2}, \ldots, w_{o_{|\mathcal{D}|}}\}$. This vector is also used to represent a document d_i in a vector space model and store the frequency, or other frequency-based measure, of terms in the document d_i. Figure 4 illustrates a bipartite network that uses the structures described above.

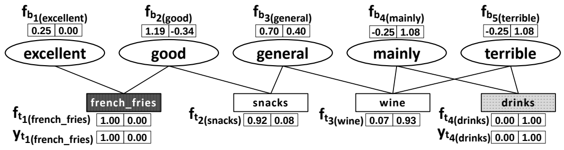
Figure 5 and Figure 6 present the example of semi-supervised learning based on bipartite network considering the TCBHN algorithm with $\eta = 0.5$. In the first example, we consider the bipartite heterogeneous network presented in Figure 2 to perform aspect identification, where $\mathcal{C} = \{ASPECT, NOT\text{-}ASPECT\}$, $\mathcal{B} = \{NN,\ det,\ nsubj,\ dobj,\ cop,\ advmod,\ cc,\ conj,\ RB\}$, $\mathcal{T}^L =$



(a) Initial state of the bipartite network structure for aspect sentiment analysis.



(b) Intermediate state of the bipartite network.



(c) Final state of the bipartite network with weights assigned to all network objects.

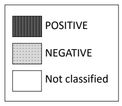


Fig. 6. Example of semi-supervised learning based on bipartite network considering the TCBHN algorithm with $\eta = 0.5$ to perform aspect sentiment analysis.

{"*food*", "*much*"} and $\mathcal{T}^U = \{$"*staff*", "*nothing*"$\}$. In the second example, we consider the bipartite heterogeneous network presented in Figure 3 to perform aspect sentiment analysis, where $\mathcal{C} = \{POSITIVE, NEGATVIE\}$, $\mathcal{B} = \{excellent,\ good,\ general,\ mainly,\ terrible\}$, $\mathcal{T}^L = \{$"*french–fries*", "*drinks*"$\}$ and $\mathcal{T}^U = \{$"*snacks*", "*wine*"$\}$ To facilitate illustration of the operation, the frequencies of terms are equal to 1 and $\eta = 0.5$.

The graph regularization has to satisfy two assumptions: (i) two objects connected in the graph tend to be classified with the same label and (ii) the object labels should be close to the real class information (training set). We present below the algorithms to perform graph regularization on bipartite networks used in this article, with its respective graph regularization functions. In each regularization function the first term is related to the first assumption and, analogously, the second term describes the second assumption.

—**Label Propagation based on Bipartite Heterogeneous Network (LPBHN)** [Rossi et al. 2014]: this is a parameter-free algorithm to perform semi-supervised learning on bipartite networks. This algorithm is an extension of the Gaussian Fields and Harmonic Function (GFHF) algorithm [Zhu et al. 2003] to bipartite heterogeneous networks. The regularization function to be minimized by LPBHN is:

$$Q(\mathbf{F}) = \frac{1}{2}\sum_{t_i \in \mathcal{T}}\sum_{b_j \in \mathcal{B}} w_{t_i,b_j}(\mathbf{f}_{t_i} - \mathbf{f}_{b_j})^2 + \lim_{\mu \to \infty} \mu \sum_{t_i \in \mathcal{T}^L}(\mathbf{f}_{t_i} - \mathbf{y}_{t_i})^2 \tag{1}$$

There is a restriction that $\mathbf{f}_{t_i} = \mathbf{y}_{\mathbf{t_i}}$, so the second term of Equation 1 has a value tending to infinity.

—**GNetMine** [Ji et al. 2010]: this is an extension of the Learning with Local and Global Consistensy (LLGC) algorithm [Zhou et al. 2004]. The regularization function to be minimized by GNetMine is:

$$Q(\mathbf{F}) = \sum_{t_i \in \mathcal{T}}\sum_{b_j \in \mathcal{B}} w_{t_i,b_j}\left\|\frac{\mathbf{f}_{t_i}}{\sqrt{\sum_{b_k \in \mathcal{B}} w_{t_i,b_k}}} - \frac{\mathbf{f}_{b_j}}{\sqrt{\sum_{t_k \in \mathcal{T}} w_{t_k,b_j}}}\right\|^2 + \sum_{t_i \in \mathcal{T}^L}\alpha_{t_i}(\mathbf{f}_{t_i} - \mathbf{y}_{t_i}), \tag{2}$$

in which $0 < \alpha < 1$ gives the importance of each term of the Equation 2.

—**Tag-based Model (TM)** [Yin et al. 2009]: this algorithm was initially proposed to classify web objects connected to social tags. In our context, the regularization function to be minimized by TM is:

$$Q(\mathbf{F}) = \Big(\beta \sum_{b_i \in \mathcal{B}^L} ||\mathbf{f}_{b_i} - \mathbf{y}_{b_i}||^2 + \gamma \sum_{B_i \in \mathcal{B}^U} ||\mathbf{f}_{b_i} - \mathbf{y}_{b_i}||^2\Big) + \Big(\sum_{b_i \in \mathcal{B}}\sum_{t_j \in \mathcal{T}} w_{b_i,t_j}||\mathbf{f}_{b_i} - \mathbf{f}_{t_j}||^2\Big), \tag{3}$$

in which the parameters β and γ control the importance given to term of the Equation 3.

—**Transductive Classification based on Bipartite Heterogeneous Network (TCBHN)** [Rossi et al. 2016]: this is an algorithm which performs optimization and label propagation to minimize the following regularization function:

$$Q(\mathbf{F}) = \frac{1}{2}\left(\sum_{c_k \in \mathcal{C}} \Big(\sum_{t_i \in \mathcal{T}^U} f_{t_i,c_k} - \sum_{b_j \in \mathcal{T}} w_{t_i,b_j} \cdot f_{b_j,c_k} \Big)\right)^2 + \frac{1}{2}\left(\sum_{c_k \in \mathcal{C}} \Big(\sum_{t_i \in \mathcal{T}^L} y_{t_i,c_k} - \sum_{b_j \in \mathcal{B}} w_{t_i,b_j} \cdot f_{b_j,c_k} \Big)\right)^2 \quad (4)$$

In our ASPHN approach, we use the first bipartite network for the classification of aspects in which the vertices are the linguistic features and few aspects are labeled as “yes” or “no”. After the label propagation of this stage, the aspects classified as “yes” are used in the second bipartite network, in which the labels of the aspects are “positive”, “negative” or “neutral”; and the vertices are the terms. The idea of this second bipartite network is that if a term is connected with a labeled aspect, then this label information is propagated to unlabeled aspects connected to the same term. After this stage, the heterogeneous network contains the classified aspects as well as the polarity of the aspect.

4. EXPERIMENTAL EVALUATION

In this experimental evaluation, we analyze the feasibility of our proposal, ASPHN, to perform aspect identification and sentiment analysis. More details about the datasets, experimental setup and evaluation criteria, and the results are presented in the next sections.

4.1 Datasets and Structured Representation

We considered two datasets which contain aspect annotation and the polarity of the texts. The first dataset is composed of 3.044 English reviews about restaurants and the second is composed by 3.048 about laptops. Both are available in [Pontiki et al. 2014]. Each review may contain one or more than one aspect. The aspect polarity can be positive, negative or neutral. The Restaurant dataset about restaurants contains 3699 aspects with the following polarities sentiments: 2164 are positive, 637 are neutral and 807 are negative. The Laptop dataset contains 2373 aspects with the following polarity sentiments: 994 are positive, 464 are neutral and 870 are negative. We calculate the S-Index [Rossi et al. 2013] for both restaurants and laptops datasets. S-Index intend to measure the overlap among the classes. When S-Index value is close to 1 means that the classes are well separated, i.e, there is no overlap among the classes. Otherwise, the value is close to 0. Therefore, S-Index can be used to understand the classification complexity for each dataset in controlled scenarios. The S-Index values for the aspect identification and sentiment classification in the laptop dataset are 0.82 and 0.75, respectively. Similarly, the S-Index values are 0.89 and 0.77 for the laptops dataset. Thus, we can observe (*apriori*) that the restaurant dataset is a more complex classification problem than the laptops dataset. Due to reasons concerning reproducibility, all source codes and datasets used in our experimental evaluation are freely available at `http://gepic.ufms.br/asphn2016/`.

To represent the dataset for machine learning algorithms to perform aspect identification, we extract linguistic features as presented in Section 3.1. We generated a vector space-model representation, as presented in Table I, and then bipartite networks were generated, as presented in Figure 2. We generated 75 linguistic features composed of grammatical structures and syntactic dependencies. The vector-space model representations and the bipartite network representations were generated according to the steps presented in Section 3.1, as presented in Table II and Figure 3 respectively.

4.2 Experiment configuration and evaluation criteria

We compared our proposal with other semi-supervised learning algorithms based on the vector space model. Moreover, we also compared our proposal with supervised learning algorithms, which is traditionally used to perform sentiment analysis. We carried out these comparisons for two reasons: (i) to verify if the use of semi-supervised learning based on bipartite heterogeneous networks is superior to the performance of semi-supervised learning based on vector space model and (ii) to analyze if our proposal is superior to the performance obtained by traditional machined learning, i.e., supervised learning based on vector space model and large labeled data. Moreover, we evaluated how much unlabeled documents are useful to improve aspect identification and sentiment analysis.

The semi-supervised learning algorithms based on bipartite networks presented in Section 3.2 were used in our experimental evaluation. The algorithms and the respective parameters are [Rossi et al. 2016]:

—**Tag-based Model) (TM)**: we used $\beta = \{0.1, 1, 10, 100, 1000\}$, and $\gamma = \{0.1, 1, 10, 100, 1000\}$.
—**GNetMine**: we used $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$.
—**Transductive Categorization based on Bipartite Heterogeneous Networks (TCBHN)**: we considered the iterative solution of TCBHN presented in [Rossi et al. 2016]. This iterative solution has two parameters η (error correction rate) and ϵ (minimum squared error). We used $\eta = \{0.01,\ 0.05, 0.1, 0.5\}$, $\epsilon = 0.01$, 10 as the maximum number of global iterations and 100 as the maximum number of local iterations, which gives a total of 1000 iterations.
—**Label Propagation using Bipartite Heterogeneous Networks (LPBHN)**: this is a parameter-free semi-supervised learning algorithm.

In this experimental evaluation, we considered the traditional semi-supervised learning algorithms based on vector space model [Zhu and Goldberg 2009]. The algorithms and their parameters are [Rossi et al. 2016]:

—**Self-Training**: at each step of the Self-Training approach, the X most confident classified examples considering a previous classification model (induced through supervised learning) are added to the set of labeled examples. This process is repeated until all unlabeled documents were added to the set of labeled documents. We considered $X = \{5, 10, 15, 20\}$. We considered Multinomial Naïve Bayes (MNB) as the inductive learning algorithm for Self-Training since it presents the best trade-off between classification performance and time for textual data [Rossi et al. 2014; Nigam et al. 2000].
—**Expectation Maximization (EM)**: we considered the EM for textual data classification presented in [Nigam et al. 2000]. In this algorithm we have to define the parameter λ (weight of unlabeled examples during semi-supervised learning) and the number of components for each class. We used $\lambda = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and 1, 2, 5, 10 components for each class.

We also run state-of-the art inductive supervised learning algorithms for analysis and to compare with our semi-supervised learning proposal. This also allows us to analyze if the use of unlabeled documents actually improves classification performance. The algorithms, parameters, and considerations of the inductive supervised learning algorithms are [Rossi et al. 2016]:

—**Multinomial Naïve Bayes (MNB)**: we considered MNB since it is the learning algorithm used in Self-Training, Co-Training and Expectation Maximization. This allowed us to measure the difference in classification performance and time to move from inductive supervised learning to transductive learning for algorithms based on the vector space model. There are no parameters for MNB.
—**Support Vector Machine (SVM)**: we considered three types of kernel: Linear, Polynomial (exponent = 2) and RBF (Radial Basis Function). Since the parameter C is real and positive, some authors set this value as 10^Y. For each type of kernel we considered $Y = \{-5; -4; -3; -2; -1; 0; 1; 2; 3; 4; 5\}$. We use the SVM implementation available in Weka tool[4].

[4]Weka: `http://www.cs.waikato.ac.nz/ml/weka/`

—***k-NN***: we considered $k = \{7; 17; 37; 57\}$ [Rossi et al. 2014]. We also considered *k-NN* algorithm without and with a weighted vote, which gives for each of the nearest neighbors a weighted vote equal to $(1 - s)$, where s is a similarity measure among neighbors. We adopted cosine as the similarity measure.

We used all iterative solutions for all algorithms which have iterative solutions (EM, LPBHN, GNetMine, TCBHN, and TM). The maximum number of iterations was set to 1000 [Rossi et al. 2016].

We used the F^1 measure to compare the classification results. F^1 is the harmonic mean of precision and recall measures, in which both measures have the same weight, i.e.

$$F^1 = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{5}$$

Precision and recall were computed for each class in multiclass evaluation. The precision and recall of a class c_i are:

$$Precision_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FP_{c_i}}, \quad (6a) \qquad Recall_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FN_{c_i}}, \quad (6b)$$

where TP (True Positive) means the number of test documents correctly assigned to class c_i, FP (False Positive) means the number of test documents from class c_j ($c_j \neq c_i$) but assigned to class c_i, and FN (False Negative) is the number of test documents from class c_i but assigned to class c_j ($c_j \neq c_i$). The Precision measure returns the percentage of documents correctly classified as c_i considering all documents classified as c_i. The Recall measure returns the percentage of documents correctly classified as c_i considering all documents which actually belong to class c_i.

Two strategies to summarize the results of precision and recall computed for each class of a text collection are: (i) **micro-averaging** and **macro-averaging** [Sokolova and Lapalme 2009]. The micro-averaging strategy performs a sum of the terms of the evaluation measures. Therefore, the precision and recall using the micro-averaging strategy are:

$$Precision^{Micro} = \frac{\sum_{c_i \in \mathcal{C}} TP_{c_i}}{\sum_{c_i \in \mathcal{C}}(TP_{c_i} + FP_{c_i})}, \quad (7) \qquad Recall^{Micro} = \frac{\sum_{c_i \in \mathcal{C}} TP_{c_i}}{\sum_{c_i \in \mathcal{C}}(TP_{c_i} + FN_{c_i})}. \quad (8)$$

The macro-averaging strategy performs an average over the evaluation measures for each class. Therefore, the precision and recall using macro-averaging strategy are:

$$Precision^{Macro} = \frac{\sum_{c_i \in \mathcal{C}} Precision_{c_i}}{|\mathcal{C}|}, \quad (9) \qquad Recall^{Macro} = \frac{\sum_{c_i \in \mathcal{C}} Recall_{c_i}}{|\mathcal{C}|}. \quad (10)$$

Micro-averaging scores are dominated by the number of TP. Therefore, large classes dominate small classes in micro-averaging scores. On the other hand, macro-averaging gives equal weight to each class. In this case, the number of TP in small classes are emphasized in macro-averaging scores. These two strategies give different scores and are complementary to each other. We denote F^1 computed through micro-averaging of precision and recall by *Micro-*F^1, and through macro-averaging by *Macro-*F^1.

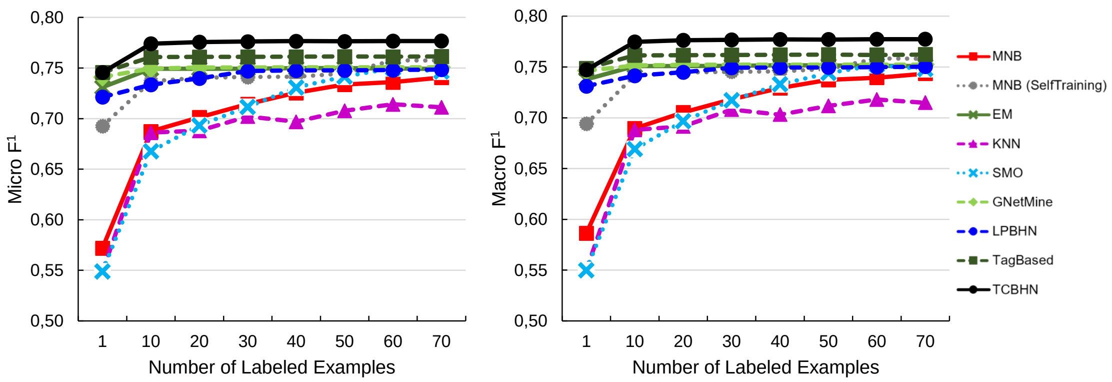
To obtain *Micro-*F^1 and *Macro-*F^1, we first carried out a 10-fold cross-validation process. For each training set (9 folds) we carried out 10 runs to induce a classification model considering N randomly selected labeled documents in each run. We considered $N = \{1, 10, 20, 30, 40, 50, 60, 70\}$. This variation in the number of labeled documents allowed us to better demonstrate the behavior of the algorithms for different number of labeled documents, a trade-off between the number of labeled documents and classification performance, and the differences among the inductive supervised learning algorithms and semi-supervised learning algorithms as we increase the number of labeled documents. The remaining training examples were considered as unlabeled examples for semi-supervised learning algorithms. Thus, 100 executions were carried out and, in each execution, we obtained an accuracy value. The final *Micro-*F^1 and *Macro-*F^1 values presented in the next section were an average of the 100 values obtained in the 10-fold cross-validation. Besides the direct analysis comparing the classification performances, we also compare the algorithms considering the Friedman statistical significant test and Nemenyi's post hoc test with 95% of confidence level to assess statistically significant differences among the classification results [Demsar 2006].

4.3 Results

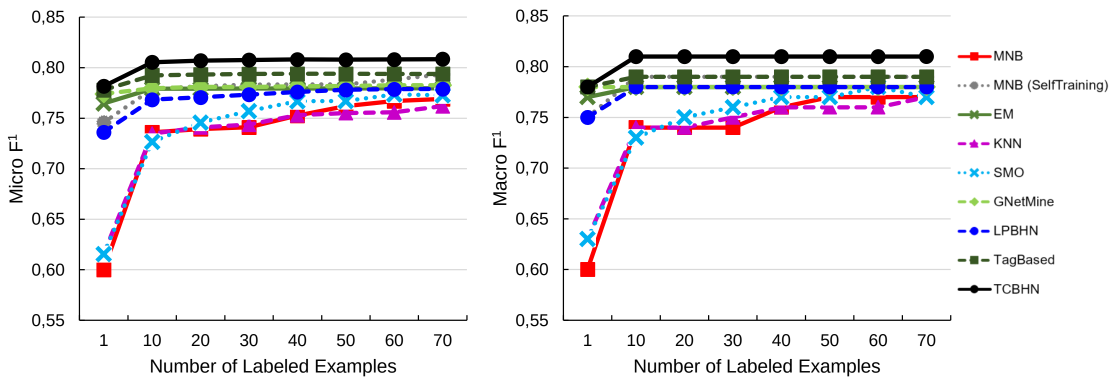
We analyze the results considering each one of the steps to perform ABSA individually, i.e., an analysis about aspect identification and another about sentiment analysis.

4.3.1 *Aspect Identification.* Figure 7(a) presents the classification performance of aspects for the Laptop dataset and Figure 7(b) presents the classification performance for the Restaurants dataset. Considering both datasets and both Micro-F^1 and Macro-F^1 measure, TCBHN presented the highest values for all numbers of labeled examples used in the experimental evaluation. TM algorithms presented the second best Micro-$F1$ values for all numbers of labeled examples.

In Figure 8 we present the critical difference diagrams among the different learning algorithms applied for aspect identification considering the results of Micro-F^1 and Macro-F^1 measures presented in Figure 4.3.1. In this diagram, each algorithm is sorted according to the rank of the Friedman's Statistical Significant Test [Demsar 2006]. The methods connected by a line do not present statistical significant difference among them.

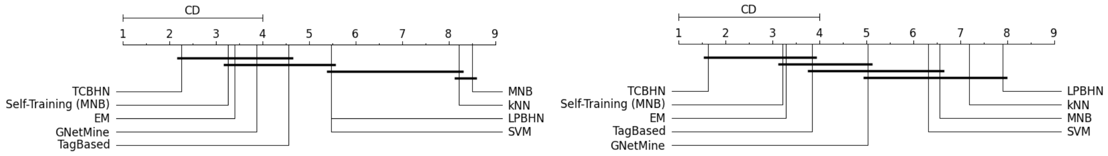


(a) Laptop dataset



(b) Restaurant dataset

Fig. 7. Aspect identification results.

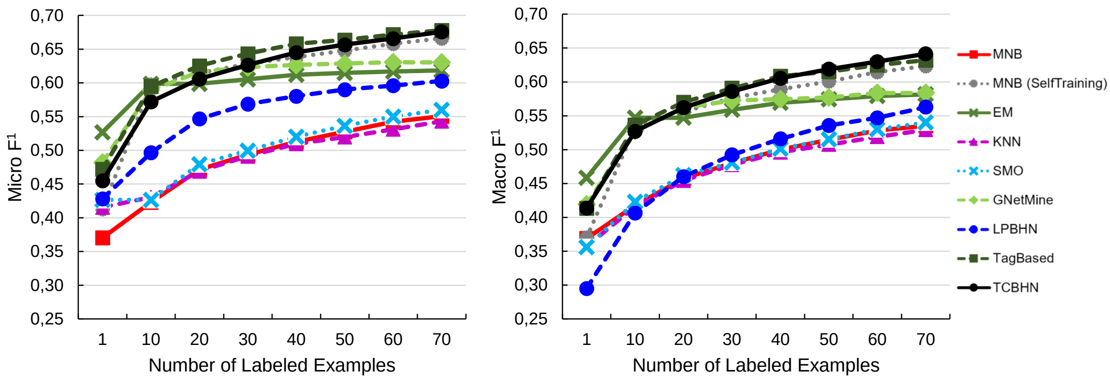


(a) Micro-F^1. (b) Macro-F^1.

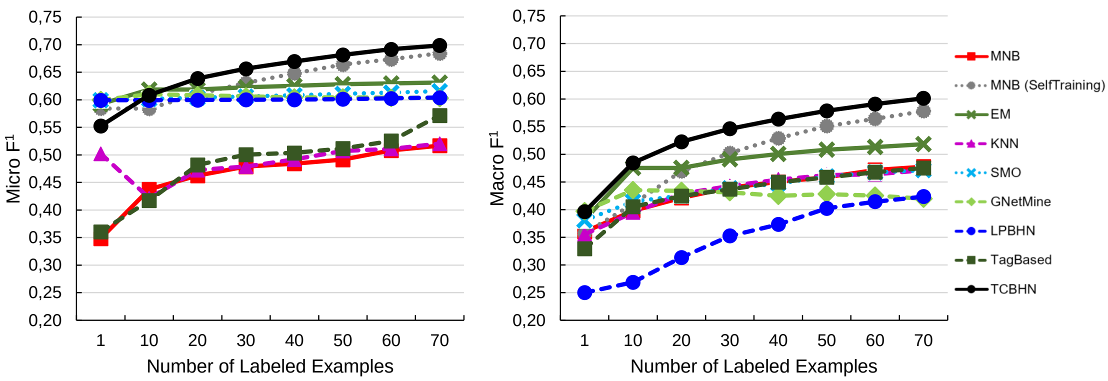
Fig. 8. Critical difference diagrams for aspect identification.

According to the critical difference diagrams, TCBHN obtained the best average ranking for Micro-F^1 and Macro-F^1. Moreover, TCBHN presented better results with statistically significant differences than all supervised learning algorithms. In general, there were not statistically significant differences among the semi-supervised learning algorithms. However, we highlight that algorithms based on bipartite networks present a lower computation cost than algorithms based on the vector space model [Rossi et al. 2016]. Thus, we can conclude that the use of unlabeled examples is also helpful to perform aspect identification through machine learning.

4.3.2 *Aspect Sentiment Classification.* Figure 9(a) classification performance of sentiment polarities for the Laptop dataset. Considering the Micro-F^1 measure, EM algorithm presented the highest values when considering 1 and 10 labeled examples. TM and presented the highest Micro-F^1 values when considering 20 or more labeled examples per class. TCBHN and Self-Training presented results close to TM when using 20 or more labeled examples per class.

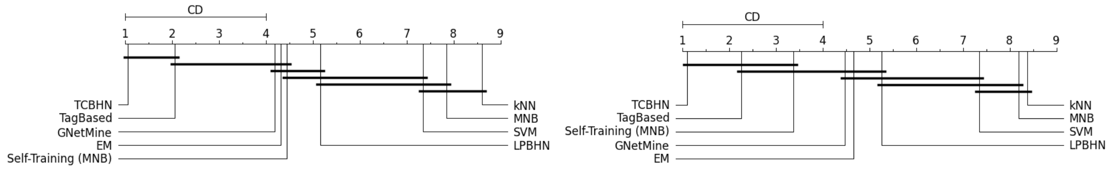


(a) Laptop dataset



(b) Restaurant dataset

Fig. 9. Aspect sentiment classification results.



(a) Micro-F^1. (b) Macro-F^1.

Fig. 10. Critical difference diagrams for aspect sentiment classification.

Considering the Macro-F^1 measure, EM algorithm presented the highest values when considering 1 and 10 labeled examples again. TM and TCBHN algorithms presented the highest values when considering 20 or more labeled examples. Figure 9(b) presents the performance for the Restaurant dataset. Considering the Micro-F^1 measure, GnetMine presented the highest values for 1 labeled example per class, EM presented the highest values for 10 labeled examples and TCBHN the highest values when considering 20 or more labeled examples for each class. Considering the Macro-F^1 measure, TCBHN presented the highest values for all numbers of labeled examples. For both Laptops and Restaurants datasets, and for both Micro-F^1 and Macro-F^1, semi-supervised learning algorithms presented better performance than supervised learning algorithms. The exception is the algorithm LPBHN which presented inferior results than the supervised learning algorithms for the Restaurant dataset on the Macro-F^1 measure.

In Figure 10 we present the critical difference diagrams among the different learning algorithms applied for aspect sentiment classification considering the results of Micro-F^1 and Macro-F^1 measures presented in Figure 4.3.2. We can observe that TCBHN obtained the best average ranking and TM obtained the second best ranking considering Micro-F^1 and Macro-F^1 measures. We highlight that both algorithms are based on bipartite heterogeneous networks.

TCBHN and TM algorithms presented better results with statistically significant differences than the supervised learning algorithms. Thus, we can conclude that the bipartite modeling and the use of unlabeled examples were again useful to improve aspect sentiment classification.

5. CONCLUSIONS AND FUTURE WORK

In this article we presented the Aspect-Based Sentiment Propagation for Heterogeneous Networks (ASPHN) approach. Our proposal performs aspect identification and the sentiment analysis of the aspect through semi-supervised learning based on heterogeneous networks. To the best of our knowledge, there is no research which performs both aspect identification and sentiment analysis through semi-supervised learning based on heterogeneous networks.

In our proposal, we demonstrate how to represent an aspect-base sentiment analysis (ABSA) by using bipartite heterogeneous networks and how to use semi-supervised learning algorithms based on heterogeneous networks to perform both aspect identification and sentiment analysis. The obtained results demonstrated that our proposal obtained better results than the traditional semi-supervised learning or supervised learning algorithms based on the vector space model. Therefore, we allow the user to label fewer examples and obtain equal or better performance than traditional machine learning algorithms that use large labeled datasets.

As future work we intend to verify the feasibility of our proposal to perform ABSA in texts from other domains and texts written in Portuguese. We also intend to consider transfer learning, i.e., analyze the feasibility to learn a classification model in a certain domain and apply the learned model to perform ABSA in texts from other domains.

REFERENCES

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of Machine Learning Research* vol. 3, pp. 993–1022, 2003.

Bollegala, D., Weir, D., and Carroll, J. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In *Proc. of the 49th ACL Annual Meeting: Human Language Technologies*. pp. 132–141, 2011.

Chapelle, O., Schlkopf, B., and Zien, A. *Semi-Supervised Learning.* The MIT Press, 2010.

Chen, Z., Mukherjee, A., and Liu, B. Aspect extraction with automated prior knowledge learning. In *Proc. of the 52nd ACL Annual Meeting*. pp. 347–358, 2014.

Cilibrasi, R. L. and Vitanyi, P. M. B. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering* 19 (3): 370–383, 2007.

Cruz, F. L., Troyano, J. A., Enríquez, F., Ortega, F. J., and Vallejo, C. G. 'long autonomy or long delay?'the importance of domain in opinion mining. *Expert Systems with Applications* 40 (8): 3174–3184, 2013.

Dasgupta, S. and Ng, V. Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. In *Proc. of the Joint Conference of the 47th ACL Annual Meeting and the 4th Int. Joint Conference on Natural Language Processing of the AFNLP*. Vol. 2 and 2. pp. 701–709, 2009.

Demsar, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* vol. 7, pp. 1–30, 2006.

Ding, X., Liu, B., and Yu, P. S. A holistic lexicon-based approach to opinion mining. In *Proc. of the 2008 Int. Conference on Web Search and Data Mining*. pp. 231–240, 2008.

Ganeshbhai, S. Y. and Shah, B. K. Feature based opinion mining: A survey. In *IEEE Int. Advance Computing Conference*. pp. 919–923, 2015.

Ghani, R., Probst, K., Liu, Y., Krema, M., and Fano, A. Text mining for product attribute extraction. *ACM SIGKDD Explorations Newsletter* 8 (1): 41–48, 2006.

Guo, H., Zhu, H., Guo, Z., Zhang, X., and Su, Z. Product feature categorization with multilevel latent semantic association. In *Proc. of the 18th Conference on Information and Knowledge Management*. pp. 1087–1096, 2009.

He, Y. and Zhou, D. Self-training from labeled features for sentiment analysis. *Information Processing & Management* 47 (4): 606–616, 2011.

Hofmann, T. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 50–57, 1999.

Hu, M. and Liu, B. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, 2004.

Ji, M., Sun, Y., Danilevsky, M., Han, J., and Gao, J. Graph regularized transductive classification on heterogeneous information networks. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part I*. pp. 570–586, 2010.

Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Martínez-Cámara, E., and Ureña-López, L. A. Combining resources to improve unsupervised sentiment analysis at aspect-level. *Journal of Inf. Science* 42 (2): 213–229, 2015.

Kim, S.-M. and Hovy, E. Determining the sentiment of opinions. In *Proc. of the 20th Int. Conference on Computational Linguistics*, 2004.

Kobayashi, N., Inui, K., and Matsumoto, Y. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proc. of the EMNLP-CoNLL*. pp. 1065–1074, 2007.

Kovelamudi, S., Ramalingam, S., Sood, A., and Varma, V. Domain independent model for product attribute extraction from user reviews using wikipedia. In *Proc. of the 5th IJCNLP*. pp. 1408–1412, 2011.

Liu, B. Sentiment analysis and subjectivity. *Handbook of natural language processing* vol. 2, pp. 627–666, 2010.

Liu, B. *Sentiment Analysis and Opinion Mining.* Morgan & Claypool, 2012.

Liu, B. *Sentiment Analysis - Mining Opinions, Sentiments, and Emotions.* Cambridge University Press, 2015.

Long, C., Zhang, J., and Zhut, X. A review selection approach for accurate feature rating estimation. In *Proc. of the 23rd Int. Conference on Computational Linguistics: Posters*. pp. 766–774, 2010.

Lu, Y., Zhai, C., and Sundaresan, N. Rated aspect summarization of short comments. In *Proc. of the 18th Int. Conference on World Wide Web*. pp. 131–140, 2009.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. The Stanford CoreNLP natural language processing toolkit. In *Proc. of the ACL: System Demonstrations*. pp. 55–60, 2014.

Matsuno, I. P., Rossi, R. G., Marcacini, R. M., and Rezende, S. O. Análise de sentimentos baseada em aspectos usando aprendizado semissupervisionado em redes heterogêneas. In *Proc. of the Symposium on Knowledge Discovery, Mining and Learning*. pp. 90–93, 2015.

Mei, Q., Ling, X., Wondra, M., Su, H., and Zhai, C. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *Proc. of the 16th Int. Conference on World Wide Web*. pp. 171–180, 2007.

Moghaddam, S. and Ester, M. Opinion digger: An unsupervised opinion miner from unstructured product reviews. In *Proc. of the 19th Int. Conference on Information and Knowledge Management*. pp. 1825–1828, 2010.

Moghaddam, S. and Ester, M. On the design of lda models for aspect-based opinion mining. In *Proc. of the 21st ACM Int. Conference on Information and Knowledge Management*. pp. 803–812, 2012.

Mukherjee, A. and Liu, B. Aspect extraction through semi-supervised modeling. In *Proc. of the 50th ACL Annual Meeting: Long Papers*. ACL, pp. 339–348, 2012.

Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. Text classification from labeled and unlabeled documents using EM. *Machine Learning* 39 (2-3): 103–134, 2000.

Pang, B. and Lee, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retrieval* 2 (1-2): 1–135, 2008.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. Semeval-2014 task 4: Aspect based sentiment analysis. In *Workshop on Semantic Evaluation*. pp. 27–35, 2014.

Popescu, A.-M. and Etzioni, O. Extracting product features and opinions from reviews. In *Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. pp. 339–346, 2005.

Rossi, R. G., de Andrade Lopes, A., de Paulo Faleiros, T., and Rezende, S. O. Inductive model generation for text classification using a bipartite heterogeneous network. *Journal of Computer Science and Technology* 3 (29): 361–375, 2014.

Rossi, R. G., Lopes, A. A., and Rezende, S. O. A parameter-free label propagation algorithm using bipartite heterogeneous networks for text classification. In *Proc. of the 29th ACM SAC*. pp. 79–84, 2014.

Rossi, R. G., Lopes, A. d. A., and Rezende, S. O. Optimization and label propagation in bipartite heterogeneous networks to improve transductive classification of texts. *Inf. Processing & Management* 52 (2): 217–257, 2016.

Rossi, R. G., Marcacini, R. M., and Rezende, S. O. Benchmarking text collections for classification and clustering tasks. *Institute of Mathematics and Computer Sciences, University of Sao Paulo*, 2013.

Sokolova, M. and Lapalme, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45 (4): 427–437, 2009.

Somasundaran, S., Namata, G., Getoor, L., and Wiebe, J. Opinion graphs for polarity and discourse classification. In *Proc. of the Workshop on Graph-based Methods for Natural Language Processing*. pp. 66–74, 2009.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. Lexicon-based methods for sentiment analysis. 37 (2): 267–307, 2011.

Tang, J., Nobata, C., Dong, A., Chang, Y., and Liu, H. pp. 577–585. In , *Propagation-based Sentiment Analysis for Microblogging Data*. pp. 577–585, 2015.

Titov, I. and McDonald, R. Modeling online reviews with multi-grain topic models. In *Proc. of the 17th Int. Conference on World Wide Web*. pp. 111–120, 2008a.

Titov, I. and McDonald, R. T. A joint model of text and aspect ratings for sentiment summarization. In *Proc. of the ACL: Human Language Technology*. Vol. 8. pp. 308–316, 2008b.

Wan, X. Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In *Proc. of the Conference on Empirical Methods in Natural Language Processing*. pp. 553–561, 2008.

Wan, X. Co-training for cross-lingual sentiment classification. In *Proc. of the 47th Conference of the ACL and the 4th Int. Conference on Natural Language Processing of the AFNLP*. Vol. 1 and 1. pp. 235–243, 2009.

Wu, Y. and Ester, M. Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *Proc. of the Eighth ACM Int. Conference on Web Search and Data Mining*. pp. 199–208, 2015.

Yang, B. and Cardie, C. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st ACL Annual Meeting*. pp. 1640–1649, 2013.

Yin, Z., Li, R., Mei, Q., and Han, J. Exploring social tagging graph for web object classification. In *Proc. of the Int. Conference on Knowledge Discovery and Data Mining*. pp. 957–966, 2009.

Yu, J., Zha, Z.-J., Wang, M., and Chua, T.-S. Aspect ranking: Identifying important product aspects from online consumer reviews. In *Proc. of the 49th ACL Annual Meeting: Human Language Technologies*. pp. 1496–1505, 2011.

Zhao, W. X., Jiang, J., Yan, H., and Li, X. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing*. pp. 56–65, 2010.

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., and Schölkopf, B. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*. Vol. 16. pp. 321–328, 2004.

Zhou, S., Chen, Q., and Wang, X. Active deep networks for semi-supervised sentiment classification. In *Proc. of the 23rd Int. Conference on Computational Linguistics*. pp. 1515–1523, 2010.

Zhu, J., Wang, H., Tsou, B. K., and Zhu, M. Multi-aspect opinion polling from textual reviews. In *Proc. of the 18th ACM Conference on Information and Knowledge Management*. pp. 1799–1802, 2009.

Zhu, X., Ghahramani, Z., and Lafferty, J. Semi-supervised learning using gaussian fields and harmonic functions. In *Proc. of the Int. Conference on Machine Learning*. pp. 912–919, 2003.

Zhu, X. and Goldberg, A. B. *Introduction to semi-supervised learning*. Morgan and Claypool, 2009.

Zhuang, L., Jing, F., and Zhu, X.-Y. Movie review mining and summarization. In *Proc. of the 15th ACM Int. Conference on Information and Knowledge Management*. pp. 43–50, 2006.