Noname manuscript No.
(will be inserted by the editor)

Survey on Aspect Detection for Aspect-Based Sentiment Analysis

Maria Mihaela Truşcă · Flavius Frasincar

Received: date / Accepted: date

Abstract Sentiment analysis is an important tool to automatically understand the user-generated content on the Web. The most fine-grained sentiment analysis is concerned with the extraction and sentiment classification of aspects and has been extensively studied in recent years. In this work, we provide an overview of the first step in aspect-based sentiment analysis that assumes the extraction of opinion targets or aspects. We define a taxonomy for the extraction of aspects and present the most relevant works accordingly, with a focus on the most recent state-of-the-art methods. The three main classes we use to classify the methods designed for the detection of aspects are pattern-based, machine learning, and deep learning methods. Despite their differences, only a small number of works belong to a unique class of methods. All the introduced methods are ranked in terms of effectiveness. In the end, we highlight the main ideas that have led the research on this topic. Regarding future work, we deemed that the most promising research directions are the domain flexibility and the end-to-end approaches.

Keywords Aspect-based sentiment analysis · Aspect detection · Taxonomy of methods · Introductory and survey · Neural nets

1 Introduction

The recent growth of user-generated content available online has encouraged many researchers to find solutions for natural language processing (NLP). One important

Maria Mihaela Truşcă
Department of Informatics and Economic Cybernetics, Bucharest University of Economic Studies, Piata Romana 6, 010374 Bucharest
E-mail: maria.trusca@csie.ase.ro

Flavius Frasincar
Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands
Tel.: (+31)(10)4081340
Fax: (+31)(10)4089162
E-mail: frasincar@ese.eur.nl

tool in the automatic understanding of texts is sentiment analysis, widely applied in a broad range of fields. In the beginning, sentiment analysis was considered a problem of text classification whose major aim was to find the overall polarity of a text (document-level sentiment analysis) or sentence (sentence-level sentiment analysis) (Liu (2015)). However, a document or a sentence can hold different opinions about different subjects, which means that a finer-grained analysis turns out to be necessary (Schouten and Frasincar (2015)). The difference between these two approaches is investigated by Jiang et al. (2011). Using a self-defined Twitter corpus, the authors prove that the lack of opinion analysis focused on subjects is responsible for 40% of the classification error.

According to the terminology of sentiment analysis, subjects are represented by entities and aspects. While the entity is the object of analysis, the aspect represents its features. Given that the aspects have different levels of abstraction, the pair (entity, aspects) can be represented by a hierarchical structure where the entity is the root of the tree. Figure 1 shows two examples of hierarchical entity-aspect relations. For simplicity reasons, all sentiment subjects (entities and aspects with different levels of abstractions) are called aspects.

The main tasks of aspect-based sentiment analysis (ABSA) are aspect detection and sentiment classification. These tasks were settled in the initial works for sentiment summarisation. However, more recent methods for ABSA refer to the SemEval workshop that sets a different classification of ABSA tasks. According to the SemEval 2014 workshop (Pontiki et al. (2014)), ABSA has four tasks used to extract aspect terms and their categories (or entities) and to assign sentiments per aspect and category. SemEval 2015 and SemEval 2016 workshops (Pontiki et al. (2015, 2016)) refined the tasks of the SemEval 2014 workshop and considered that the sentiment polarity of a category and aspect term should be the same, meaning that the four tasks were resumed to only three.

Currently, the sentiment prediction task is largely addressed, and it is a subject of many research works and surveys that presents it either alone (Tang et al. (2009); Medhat et al. (2014); Ravi and Ravi (2015); Zhang et al. (2018)) or as a task of ABSA (Schouten and Frasincar (2015); Do et al. (2019); Liang et al. (2022)). Given that the identification of aspects is crucial to the sentiment classification task, in this survey, we focus on the review of the most relevant works in this area. A complete comparative review about aspect detection is presented by Rana and Cheah (2016), but despite being a comprehensive study for 2016 it does not include the most recent and effective approaches like deep learning methods. On the other hand, more recent surveys about the extraction of aspects (Maitama

Fig. 1 Hierarchical relations between entity and aspects

et al. (2020a), Maitama et al. (2020b)) mainly have a synthetic nature that aims to group the proposed solutions in terms of the employed techniques and methods, or in terms of languages and dataset domains. Given all these previous researches, our work can be considered complementary, providing a deep dive into all the presented methods that can help future researchers to understand the current status and the required improvements specific to aspect detection. All the research works presented in this survey were found in the ACM, Scopus, Web of Science, DBLP, Semantic Scholar, and IEEE Xplore databases using keywords as "aspect detection", "aspect extraction", and "aspect-based sentiment analysis".

Depending on the presence of aspects in a text, Hu and Liu (2004) distinguish between explicit and implicit aspects. While the explicit aspects are represented by a set of tokens in a sentence (e.g., the sentence "the battery life of my cell is short" has the explicit aspect "battery life"), implicit aspects can only be deducted from the context (e.g., the sentence "my new camera is too big to keep it in my pocket" has the implicit aspect "camera size"). As it was mentioned above, all entities and aspect categories are considered aspects, which implies that the two tasks of the SemEval workshop for the extraction of aspect terms and aspect categories are treated as a singular aspect detection task. The detection of explicit aspects resembles the extraction of aspect terms that might have a lower or higher granularity. As the implicit aspects can be extracted only conceptually, we need to rely either on the detection of descriptors or on the assignment of aspect labels at the document or sentence level. Therefore, some of the methods proposed for the aspect categorization are included in the section of implicit aspects.

Besides the type of target aspects, the current methods are classified based on the type of learning and the employed approach. The learning techniques are supervised and unsupervised, depending on the availability of training labels. While supervised learning is usually the most effective, its major shortcoming is the requirement of prior annotations. To face this problem, some methods utilise semi-supervision and weak supervision. Semi-supervised learning assumes that only a small subset of the available data is annotated. The remaining unlabeled data is used for both testing and training in a repetitive process that ends when a convergence criterion is reached. Weak supervision aims to allocate inexact labels to unsupervised data in order to benefit from supervised learning. The labels are cheap, usually determined by a pattern-based method, and do not exactly represent the desired information necessary for prediction.

The main two methods developed for the extraction of aspects leverage either on linguistic observations from different sources and data, or on machine learning concepts. Among these two options, pattern-based systems were introduced first. Their approaches mainly create patterns that embody linguistic concepts like part-of-speech (POS) tags, dependency relations, domain lexicons or co-occurrences between words. Pattern-based systems were followed by machine learning methods classified as discriminative and generative. The difference between the two categories lays in the modeling of the input-output relation. The discriminative methods define the boundary between classes by learning the probability of output conditioned by the given input. On the other hand, the generative methods learn how data was generated by modeling the joint probability between input and output. The conditional probabilities of the output given the input can also be learnt by the generative methods, but an extra Bayesian step is required. While the generative class is represented in the literature only by the Hidden Markov

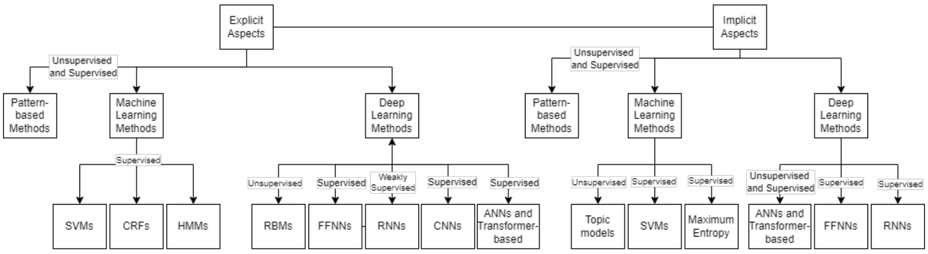

Fig. 2 Overview of the presented methods for aspect detection. Due to the lack of comparative results, we do not present the topic modeling techniques in the following sections.

Models (HMMs) (Rabiner (1989)) and Restricted Boltzmann Machines (RBMs) (Salakhutdinov et al. (2007)), the discriminative class includes methods like Conditional Random Fields (CRFs) (Lafferty et al. (2001)), Maximum Entropy (Berger et al. (1996)), Support Vector Machine (SVM) (Boser et al. (1992)), and the majority of neural networks. As the latter ones have received more attention recently, we decided to present the neural networks in a separate section from the machine learning one. Even if the pattern-based and machine learning methods are different, their concepts are usually combined. Figure 2 presents the current methods proposed till now to detect aspects.

The remaining part of the paper is organised, as follows. Section 2 gives an overview of the most relevant works for the detection of explicit aspects and Sect. 3 introduces the methods proposed for the identification of implicit aspects. Section 4 discusses the presented methods and indicates directions for further improvements. Section 5 presents the conclusions and future research directions.

2 Explicit Aspects

This section is dedicated to aspects explicitly mentioned in the text. The following sections present the solutions inspired by pattern-based systems (Sect. 2.1), machine learning methods (Sect. 2.2), and neural networks (Sect. 2.3).

2.1 Pattern-Based Methods

Usually, pattern-based methods are unsupervised, and even if the aspect labels exist, they are used only for the final check. Table 1 lists the performance results of the presented pattern-based systems together with the details about the employed datasets.

Unsupervised Learning. Yi et al. (2003) were the first that propose an unsupervised method for aspect-based sentiment summarisation considering the generation of aspects as a subtask. According to their approach, the aspects are detected using noun-based patterns. The idea of considering nouns as aspect candidates is common and supported by the claim of Liu (2011) according to which 60-70% of the aspects are nouns. Then, non-aspect nouns are filtered out using two sorting methods. The first method is a mixture model (Zhai and Lafferty (2001)) that computes for each word a score as a linear combination between its corpus-level

Table 1 Pattern-based methods for the detection of explicit aspects

References	Method	Dataset	Language	Domain	Performance
Tubishat et al. (2021)	IWOA+PA	Hu and Liu (2004)	English	Electronic Products	Recall: 93% Precision: 92% F_1: **92%**
Rana and Cheah (2017)[1]	TF-RBM	Hu and Liu (2004)	English	Electronic Products	Recall: 92% Precision: 87% F_1: **89.43%**
Rana and Cheah (2019)[1]	SPR	Hu and Liu (2004)	English	Electronic Products	Recall: 91% Precision: 86% F_1: **89%**
Liu et al. (2016)[1]	RSLS+	Hu and Liu (2004)	English	Electronic Products	Recall: 91.1% Precision: 84.9% F_1: **87.9%**
Qiu et al. (2009)	Prop-Dep	Hu and Liu (2004)	English	Electronic Products	Recall: 83.0% Precision: 88.0% F_1: **86.0%**
Popescu and Etzioni (2007)	OPINE	Hu and Liu (2004)	English	Electronic Products	Recall: 77.0% Precision: 94.0% F_1: **84.65**%
Rana and Cheah (2017)[2]	TF-RBM	Hu and Liu (2004)	English	Electronic Products	Recall: 80% Precision: 79% F_1: **79.45%**
Rana and Cheah (2019)[2]	SPR	Hu and Liu (2004)	English	Elctronic Products	Recall: 76% Precision: 81% F_1: **78%**
Liu et al. (2016)[2]	RSLS+	Hu and Liu (2004)	English	Electronic Products	Recall: 76.7% Precision: 78.2% F_1: **77.3%**
Hu and Liu (2004)[3]	FBS	Self-defined	English	Electronic Products	Recall: 80.0% Precision: 72.0% F_1: **75.8**%
Wu et al. (2009)[4]		Hu and Liu (2004) & Jindal and Liu (2008)	English	Mixed	Recall: 85.5% Precision: 42.8% F_1: **57.0%**
Kang and Zhou (2017)[5]	RubE	Hu and Liu (2004) & Zhang (2013)	English	Mixed	Recall: 87.0% Precision: 88.0% F_1: **87.0%**
Dragoni et al. (2019)		SemEval 2015 (Task 12)	English	Laptop	Recall: 41.57% Precision: 67.02% F_1: **51.31%**
Dragoni et al. (2019)		SemEval 2015 (Task 12)	English	Restaurant	Recall: 53.68% Precision: 68.95% F_1: **60.36%**
Yi et al. (2003)[6]	bBNP-L	Self-defined	English	Mixed	Precision: 98.35%
Liu et al. (2005)[7]	Opinion Observer	Self-defined	English	Electronic Products Pros	Recall: 90.20% Precision: 88.90% F_1: **89.55%**
Liu et al. (2005)[7]	Opinion Observer	Self-defined	English	Electronic Products Cons	Recall: 82.40% Precision: 79.1% F_1: **80.72%**
Blair-Goldensohn et al. (2008)[8]	Combined Mtd. (Dynamic & Static)	Self-defined	English	Restaurant	Recall: 66.10% Precision: 88.20% F_1: **75.50**%
Blair-Goldensohn et al. (2008)[8]	Combined Mtd. (Dynamic & Static)	Self-defined	English	Hotel	Recall: 68.15% Precision: 83.70% F_1: **75.10**%
Somasundaran and Wiebe (2009)[9]	OpPr+Disc	Self-defined	English	Mixed	Accuracy: 65.81% Recall: 65.81% Precision : 68.16% F_1: **66.92%**
Zhang et al. (2010)[10]		Self-defined	English	Mixed	Recall: 60.25% Precision: 67% F_1: **63.15%**

[1] The result is reported based on multiple occurrences of aspects (a given aspect with n occurrences in a corpus is considered properly extracted if it is determined at least once).
[2] The result is reported based on distinct occurrence of aspects (all of the n occurrences of a given aspect in a corpus should be determined separately).
[3] The result is reported as the weighted average for the aspects *digital camera*, *DVD player*, *mp3 player*, and *cellular phone*.
[4] The result is reported for the aspects *cell phone*, *DVD player*, *digital camera*, *mp3 player*, and *diaper*.
[5] The result is reported for the aspects *cell phone*, *DVD player*, *digital camera*, *mp3 player*, and *movie*.
[6] The result is reported as the weighted average for the aspects *digital camera*, and *music*.
[7] The result is reported without differentiating between explicit and implicit aspects.
[8] The result is reported as the average for the aspects *service* and *value* identified.
[9] The result is reported as the weighted average for aspects detected from four dual debates *Windows vs Mac*, *Sony Ps3 vs. Nintendo Wii*, *Firefox vs. Opera*, and Firefox vs. Internet Explorer.
[10] The result is reported as the average for the aspects *cars*, *mattress*, *cellular phone*, and *LCD*. The corpus considers 2000 sentences for each aspect.

References	Method	Dataset	Language	Domain	Performance
Moghaddam and Ester (2010)	Opinion Digger	Self-defined	English	Mixed	Recall: 87% Precision: 80% F_1: **83.35%**
Hai et al. (2013)	FOM-IEDR	Self-defined	Chinese	Cellular Phone	Recall: 61.71% Precision: 65.60% F_1: **63.60%**
Hai et al. (2013)	FOM-IEDR	Self-defined	Chinese	Hotel	Recall: 54.30% Precision: 50.37% F_1: **52.26%**
Tubishat et al. (2021)[12]	IWOA+PA	Liu et al. (2016)	English	Electronic Product	Recall: 90% Precision: 96% F_1: **93%**
Liu et al. (2016)[1, 11]	RSLS+	Self-defined	English	Electronic Products	Recall: 84.6% Precision: 81.9% F_1: **83.2%**
Liu et al. (2016)[2, 11]	RSLS+	Self-defined	English	Electronic Products	Recall: 73.1% Precision: 74.7% F_1: **73.8%**

[11] The result is reported for the aspects *computer*, *speaker*, and *wireless router*.
[12] The result is reported only for the aspects *computer* and *speaker*.

frequency and the degree of significance for a possible topic. The second method applies the Dunning log-likelihood test (Dunning (1993)).

Since opinion words usually hint at the presence of aspects, numerous pattern-based methods use them as a starting point for the extraction of aspects. Hu and Liu (2004) propose a solution that extracts both frequent and infrequent aspects. According to their approach, the frequent aspects are represented by nouns found using association rules (Liu et al. (1998)). As the number of detected aspects might be too large, the authors prune all the multi-word aspects with different word ordering and all single-word aspects that appear only in a small number of sentences and are a subset of another found aspect. To extract infrequent aspects, the authors filter out the sentences with frequent aspects and extract all nouns connected to adjectives (considered by default candidate opinion words).

Another way to use opinion words is to frame them together with the target aspects in predefined patterns. Using dual-topic debate data to find recommendations of users at the post level, Somasundaran and Wiebe (2009) introduce a set of opinion-based patterns to identify both high-level and low-level aspects, explicitly mentioned in the text. The patterns are used to determine the probability of a high-level aspect (topic) with a given polarity to be conditioned by the presence of a low-level aspect with a given polarity. The co-occurrence of the two types of aspects is mainly defined based on their vicinity but also takes into account the conjunctions between sentences that may affect the polarity of aspects. The probabilities together with the aspect-sentiment pairs represent the prior knowledge of a linear programming problem used to find the recommended topic of a post. The method works as an unsupervised classification task at the document-level.

Other similar works are presented by Poria et al. (2014) and Qiu et al. (2009). While the pattern-based method introduced by Poria et al.[1] embodies fixed dependencies relations, Qiu et al. propose a more flexible approach. Precisely, Qiu et al.'s patterns can recognise both direct and indirect relations between aspects and sentiment-bearing words. Indirect relations are limited to only one intermediate object. Using a set of sentiment-bearing seed words, the method relies on a double propagation algorithm to extend lists of opinion words and aspects. From an aspect detection perspective, there are three types of patterns. The first type

[1] Since the method extracts both implicit and explicit aspects simultaneously, the results are not presented in Table 1.

assumes the extraction of aspects using opinion words. The patterns of the second type detect aspects based on the already identified aspects. Different from the first two types of patterns, the patterns of the third type assume that the extracted aspects (and the extracted opinion words) are used as input to extract new opinion words.

The downside of Qiu et al.'s double propagation algorithm is pointed out by Zhang et al. (2010). Precisely, the algorithm is suitable only for medium size corpora, leading to either low precision for the case of large corpora or low recall for the small ones. To address the small recall, Zhang et al. introduce new patterns to detect aspects in part-whole and negation relations. In addition, new pruning methods based on relevance and frequency criteria are considered to better control the precision.

Later on, the above works (Qiu et al. (2009); Zhang et al. (2010)) are extended by Kang and Zhou (2017) in an approach that differentiates subjective and objective aspects depending on the existence of dependency relations with a sentiment-bearing word. The extraction of subjective aspects leverage on the Qiu et al.'s patterns, and on some new patterns that accept more than two intermediate nodes for indirect relations. As regards the extraction of objective features, concrete terms and part-whole relations (like Zhang et al. (2010)) are deemed. To improve the quality of the extracted aspects, three pruning methods are proposed. The first method filters out the candidates that appear only independently, not inside a phrase candidate. The second method treats all the candidates that appear together in a sentence without any conjunction connectors as a singular aspect (the first candidate is kept and the remaining ones are removed). The last method filters out the infrequent candidates and the ones with lower semantic similarity with respect to the given domain topic.

Different from the previous extensions, Liu et al. (2016) aims to assess the effectiveness of Qiu et al.'s patterns. Starting from the presumption that multiple dependencies between the nodes of patterns together with the double propagation algorithm may lead to an erroneous set of aspects, the authors propose two ranking methods. The Greedy algorithm is the first employed method according to which only the patterns that improve the overall fitness should be kept. The disadvantage of the Greedy algorithm is that it could lead to only a local optimum. To prevent this effect, the second method utilises the simulated annealing technique (Kirkpatrick et al. (1983)). The proposed method is packed as an iterative process that accepts the selection of a subset of rules at the iteration level that may be less effective than the recommended one as long as the global optimum is reached.

Similar to the work of Liu et al., Tubishat et al. (2021) adapt the Whale Optimization Algorithm (WOA) (Mirjalili and Lewis (2016)) to evaluate the fitness of the candidate rules by means of two improvements. The first extension is focused on balancing the two major states of the WOA algorithm (exploitation - attacking phase and exploration - searching for a pray phase) using the Cauchy mutation. The second improvement is similar to the above mentioned simulated annealing technique aiming to help the algorithm to reach the global optimum instead of a local one. In order to achieve this task, a heuristic method is used to pollute the set of effective patterns with a small subset of less effective patterns in terms of recall or precision. In addition to the presented method for the selection of patterns, the authors also provide a list of 126 rules for the extraction of candidate aspects that might be helpful for future analyses.

Instead of exploiting the relations between aspects and their opinion words like most of the previously presented methods, the method presented by Popescu and Etzioni (2007) leverages only on the co-occurrence of aspects with different granularities (similar to the aspect-aspect relations defined by Somasundaran and Wiebe (2009), Zhang et al. (2010) and Kang and Zhou (2017)). Namely, the authors define a set of lexical patterns and assess the quality of the candidate aspects using a Point-wise Mutual Information score computed based on their co-occurrence with the given patterns (associated with a high-level aspect).

According to the work proposed by Weichselbraun et al. (2017), the patterns are replaced by a set of relations that connect structured data in the common sense knowledge sources as DBPedia and ConceptNet used to provide information about companies and products. The expected output is a set of aspects that together with some affective knowledge (a sentiment lexicon and SenticNet emotional categories defined by Cambria et al. (2016)[2]) are used to detect sentiment-aspect pairs.

Wu et al. (2009) notice that aspects may not be represented by only single-terms, but also by multi-term expressions usually modeled as noun phrases. Their solution consists in creating a phrase dependency parser that generates high-level relations between different phrases within the input sentence. Namely, a sentence can be seen like a nested hierarchical structure where the outer relations (at the sentence-level) are given by a shallow parser and the inner relations (at the phrase-level) are indicated by a lexical dependency parser. The resulted noun phrases are considering candidate aspects.

The idea of multi-term aspects is also considered by Rana and Cheah (2017), where three noun-based patterns are introduced to detect multi-term aspects. The first pattern identifies non-opinion adjectives directly associated with the head noun and embodies them in the noun phrase. The second two patterns confer the aspect label to all pairs of nouns that share either a common preposition (like "of") or an opinion expression. Next, the set of candidate aspects is pruned based on the frequency and the normalised Google distance (Cilibrasi and Vitanyi (2007)). Additionally, the method is refined by taking into account concepts defined by SenticNet 4 (Cambria et al. (2016)) as irregular opinion expressions, and negations as sentiment indicators.

A much more recent unsupervised pattern-based method is proposed by Dragoni et al. (2019), and despite its simplicity, it has competing results with the most recent state-of-art methods developed to detect aspects. The method follows the traditional pattern-based approach that models the relation between aspects and sentiment terms. Stanford CoreNLP is the employed framework with the Coref Annotator, POS tagger, and Dependency Parser as the main components.

The majority of the above methods comprise two steps. First, candidate aspects are detected, and then, the non-aspect expressions are removed. Along with the above pruning methods, numerous approaches make use of tf-idf scores, the number of related adjectives, or different frequency-based approaches. The constraint employed by Moghaddam and Ester (2010) consists in filtering out the less frequent aspect patterns (the list of POS tags associated with the words between the aspect and nearest adjective). The most frequent patterns are found using the Generalized Sequential Pattern algorithm (Srikant and Agrawal (1996)). On the

[2] A more recent version (SenticNet 6) is already available (Cambria et al. (2020)).

other hand, Hai et al. (2013) proposed a more laborious score for domain relevance derived from deviation and dispersion measures. Its major downside is the incapacity to detect generic aspects due to the required strong correlation between aspects and the given domain topic. Likewise, Manek et al. (2017) consider the Gini index as an alternative to the tf-idf score. The authors slightly adjust the Gini index to measure the purity instead of the impurity and select only the terms with the highest scores.

Supervised Learning. Currently, the number of purely supervised pattern-based methods for the detection of explicit aspects is smaller compared with the number of unsupervised approaches. Liu et al. (2005) employ association rule mining (like Hu and Liu (2004)) to generate aspect patterns that do not need to entirely match the input sentences in terms of word ordering and distance. However, the method's flexibility could lead to a larger number of candidate aspects per pattern. The method deals with this problem using a heuristic method based on the frequencies of the candidate aspects inside the corpus. Like the previous work, Liu (2010) detect aspects by using label sequential rules (Liu (2007)), a sub-class of the association rules that take into account the order of the words when generating new patterns. However, both methods (Liu et al. (2005); Liu (2010)) are applied mainly for short reviews (with Pros and Cons) that usually comprise telegraphic sentences. For the common reviews with full sentences, both works recommend the unsupervised method proposed by Hu and Liu (2004).

Rana and Cheah (2019) propose a partially supervised approach, where sequential patterns are used to determine candidate syntactic expressions suitable for both aspect and opinion words. Initially, the approach requires aspect and opinion labels in order to prune out irrelevant patterns. Later on, the newly generated expressions are generalized in a set of rules used to extract unknown aspects.

Combined Learning. Instead of presenting the list of aspects separately for either supervised or unsupervised learning, Blair-Goldensohn et al. (2008) combine the two approaches in a hybrid and detect aspects accordingly. The unsupervised learning conventionally considers that any noun is a candidate aspect if it appears in a syntactic pattern together with an opinion word, and its global frequency exceeds a given threshold. In addition, all aspects can be found only within sentiment-bearing sentences. This approach is convenient as it allows the extraction of the aspects with a low-level of abstraction. Besides this dynamic approach, the authors propose a static supervised solution that employs a binary maximum entropy method for multi-label classification of the high-level aspects (previously annotated by the authors).

2.2 Machine Learning Methods

As was already mentioned above, machine learning methods are categorised as discriminative and generative based on how the input-output relation is modeled. Discriminative neural networks are the most numerous machine learning methods applied for the extraction of aspects, but given their particularities and the special attention they have received recently, we choose to introduce them separately. All machine learning methods mentioned in this section are supervised. Sections 2.2.1 and 2.2.2 introduce SVM and CRF discriminative classifiers, respectively, and Sect. 2.2.3 is reserved for the generative HMM. Table 2 presents the effectiveness

Table 2 Machine learning methods for the detection of explicit aspects

References	Method	Dataset	Language	Domain	Performance
Toh and Wang (2014)	DLIREC	SemEval 2014 (Task 4)	English	Restaurant	Recall: 82.72% Precision: 85.35% F_1: **84.01%**
Kiritchenko et al. (2014)	NRC-Canada	SemEval 2014 (Task 4)	English	Restaurant	Recall: 76.37% Precision: 84.41% F_1: **80.19%**
Chernyshevich (2014)	IHS RD	SemEval 2014 (Task 4)	English	Restaurant	F_1: **79.62%**
Chernyshevich (2014)	IHS RD	SemEval 2014 (Task 4)	English	Laptop	F_1: **74.55%**
Toh and Wang (2014)	DLIREC	SemEval 2014 (Task 4)	English	Laptop	Recall: 67.13% Precision: 81.90% F_1: **73.78%**
Kiritchenko et al. (2014)	NRC-Canada	SemEval 2014 (Task 4)	English	Laptop	Recall: 60.70% Precision: 78.77% F_1: **68.57%**
Hamdan et al. (2015)	LSISLIF	SemEval 2015 (Task 12)	English	Restaurant	Recall: 55.0% Precision: 72.0% F_1: **62.0**%
Jin et al. (2009)	L-HMM +POS+VE +BS	Hu and Liu (2004)	English	Digital Camera	Recall: 86.0% Precision: 67.7% F_1: **75.1%**
Shu et al. (2017)[1]	L-CRF	Hu and Liu (2004)	English	Mixed	Recall: 76.0% Precision: 81.3% F_1: **78.6**%
Mitchell et al. (2013)[2]	Discrete CRF (pipeline)	Etter et al. (2013)	English/ Spanish	Twitter	Recall: 63.0% Precision: 62.5% F_1: **62.75%**
Li et al. (2010)	SkipTreeCRFs	Self-defined	English	Movie	Recall: 76.2% Precision: 82.6% F_1: **79.3%**
Li et al. (2010)	SkipTreeCRFs	Self-defined	English	Mixed (Products)	Recall: 69.3% Precision: 86.6% F_1: **77.0%**
Yu et al. (2011)[3]		Self-defined	English	Mixed	F_1: **73.12%**%

[1] The result is reported for the aspects *computer*, *camera*, *router*, *phone*, *speaker*, *DVD*, and *mp3*. The independent data is provided by Chen and Liu (2014).
[2] The result is reported as the average for the *English*, and *Spanish* datasets (the number of aspects per dataset is unknown).
[3] The result is reported as the average for the aspects *laptop*, *camera*, *phone*, and *mp3*.

of the introduced machine learning methods together with the details about the employed datasets

2.2.1 Support Vector Machines

Using a set of candidate aspects selected based on noun tags,Yu et al. (2011) apply the one-class SVM model (Manevitz and Yousef (2001)) to distinguish between aspects and non-aspect terms. In addition to the extraction of aspects, this work aims to generate a set of weights that reveal the relevance of aspects in order to compute the final opinion rating at the document-level. Specifically, the method tries to learn an optimal set of weights that follow a multivariate normal distribution whose joint probability distribution of the mean and variance is defined by taking into account the Kullback-Leibler divergence.

2.2.2 Conditional Random Fields

CRF assumes supervised learning based on a sequence of labels, usually generated by the OBI tagging scheme that follows a process similar to the Named Entity Recognition (NER) systems. According to the OBI scheme, each token of a sentence gets a tag that represents either a non-aspect word (O), or a first or middle position term of an aspect (B and I). Using the tagging scheme, aspect detection turns out to be a classification problem applied sequentially to predict a label for each word of a sentence.

While the majority of CRFs for the extraction of aspects are trained on the OBI tagging scheme, Chernyshevich (2014) proposes a slightly different annotation scheme with four possible tags. Knowing that the aspects usually are represented by nouns, the annotation scheme always allocates the same label to the head noun of a noun phrase, no matter its position. Then, two different labels mark the left and right positions around the head noun. The fourth label is for the remaining words. The reason behind this update of the sequence labeling is based on the author's claim that aspects are easier extracted if their terms get the same label. For example, a word like "phone" has always the same label in cases like "phone" and "mobile phone".

The typical CRF predicts a label for a word at each time step based on a set of emission and transition features. Emission features couple the particularities of each word in an input sequence with the output labels. On the other hand, transition features couple consecutive output labels. The point where the majority of works dedicated to aspect identification try to enhance the effectiveness of CRF is at the level of the emission features by exploiting different word particularities. The generic word features employed by most of CRFs are POS tags, dependency relations, lemmas, gazetteers, or word clusters. Other binary information that indicates if the given word is a digit or a stop word, or has uppercase letters are extracted as well. As aspects are sensitive to opinion words, Hamdan et al. (2015) also considers polarity scores found in sentiment lexicons as extra word features. Similarly, Toh and Wang (2014) rely on the pattern defined by Qiu et al. (2009) to enrich the word-level information.

Usually, the emission features differentiate CRFs, while the transition features model only the relation between consecutive labels. Along with this general approach, transition features defined by Li et al. (2010) couple not only consecutive words but also words involved in dependencies relations. The extra links between labels are determined by the conjunctions that connect head words (determined by dependencies trees) or words with the same POS tag.

Shu et al. (2017) employ a lifelong machine learning method (Chen and Liu (2018)) packed as a CRF to enrich aspect detection of a given domain based on some prior knowledge from other independent domains. The employed CRF is a typical method with emission features defined based on dependency relations. Regarding the lifelong nature of the model, the authors simply introduced a new emission feature that marks all words involved in a dependency relation with an already known aspect.

Even if it is admitted that the two tasks of ABSA are highly related, most of the works address them separately. This shortcoming is remedied by Mitchell et al. (2013) by turning the standard pipeline approach into a joint or collapsed method. The joint method is the result of a multi-learning approach, while the collapsed method simply uses unified or joint labels. To implement these approaches, a CRF model is developed based on surface, linguistic, clustering, and sentiment word features.

2.2.3 Hidden Markov Models

HMM is one of the few generative machine learning methods we found for modeling the extraction of aspects. HMM resembles CRF as both are designed for sequential data based on emission and transition features. However, HMM is not as popular as CRF for the extraction of aspects due to its generative nature that requires an extra Bayesian operation to compute the final output probability conditioned by the input. Currently, we found only the work of Jin et al. (2009) that assesses the HMM model for aspect detection. The system is dubbed OpinionMiner and extracts both aspects and opinion words. The solution incorporates the words and their POS tags as observed features and extends the training vocabulary using antonyms, synonyms, and related words suggested by Microsoft Word's thesaurus.

In addition to the above methods, we should also mention the work of Kiritchenko et al. (2014) that adapt the above emission and transaction features to a structured Passive-Aggressive algorithm (Crammer et al. (2006)) for sequential detection.

2.3 Deep Learning Methods

Neural networks are a subset of the machine learning methods, but since they have been extensively studied for aspect detection recently, we choose to present them separately. While the first neural networks were integrated with other pattern-based or machine learning methods, the most recent works focus on a complete deep learning approach to tackle not only aspect detection but also the sentiment identification in an end-to-end manner. Currently, most of the neural networks developed for the detection of explicit aspects utilise supervised learning. This fact might be the result of the coincidence between the debut of deep learning with the introduction of ABSA as a task at the SemEval workshops, which has led to numerous supervised works that exploited the SemEval labeled data. The performance results and the details about the employed datasets of all deep learning methods presented below are listed in Table 3.

The remaining of this section is organised as follows. Section 2.3.1 presents RBMs and Sect. 2.3.2 introduces the feed-forward neural networks (FFNNs) for aspect extraction. Sections 2.3.3 and 2.3.4 give an overview of the most relevant methods based on the recurrent neural networks (RNNs) and convolutional neural networks (CNNs). Section 2.3.5 is dedicated to the attentive neural networks (ANNs) and the Transformer-based models exploited for the extraction of aspects.

2.3.1 Restricted Boltzmann Machines

The RBM model was introduced in the field of aspect detection by Wang et al. (2015) as an unsupervised neural network with three types of hidden units that correspond to the aspects, opinion, and background terms. As RBM has a generative nature, a final posterior distribution is necessary to generate the population of potential aspects, opinions, and background words. To avoid the poor selection, the objective function of Wang et al.'s RBM integrates some prior information related to nouns and adjectives, dubbed as candidate aspects and opinion words, respectively. The information consists of tf-idf and topic similarities scores (for nouns) and sentiment-bearing probabilities based on a sentiment lexicon (for adjectives). We consider that Wang et al.'s RBM is suitable for the extraction of explicit aspects despite its unsupervised nature, due to its configuration (the prior information and the organisation of the hidden layer).

Table 3 Deep learning methods for the detection of explicit aspects

References	Method	Dataset	Language	Domain	Performance
Karimi et al. (2021)	BAT	SemEval 2014 (Task 4)	English	Laptop	F_1: **85.57%**
Xu et al. (2019)[1]	BERT-PT	SemEval 2014 (task 4)	English	Laptop	F_1: **84.26%** F_1: **85.33%**
Xu et al. (2020)	DomBERT	SemEval 2014 (Task 4)	English	Laptop	F_1: **83.89%**
Yan et al. (2021)	BART-ABSA BART-ABSA[2]	SemEval 2014 (Task 4)	English	Laptop	F_1: **83.52%** F_1: **67.37%**
Hu et al. (2019)	SPAN	SemEval 2014 (Task 4)	English	Laptop	F_1: **83.35%**
Mao et al. (2021)	Dual-MRC Dual-MRC[2]	SemEval 2014 (Task 4)	English	Laptop	F_1: **82.51%** F_1: **65.94%**
Poria et al. (2016)	CNN-LP	SemEval 2014 (task 4)	English	Laptop	Recall: 78.35% Precision: 86.72% F_1: **82.32%**
Wei et al. (2020)	BiSELF-CRF	SemEval 2014 (Task 4)	English	Laptop	F_1: **81.90%**
Xu et al. (2018)	DE-CNN	SemEval 2014 (Task 4)	English	Laptop	F_1: **81.59%**
Ma et al. (2019)	Seq2Seq4ATE	SemEval 2014 (Task 4)	English	Laptop	F_1: **81.59%**
Li et al. (2018)	THA+STN	SemEval 2014 (Task 4)	English	Laptop	F_1: **80.31%**
He et al. (2019)[3, 4]	IMN	SemEval 2014 (task 4)	English	Laptop	F_1: **78.46%**
Wang et al. (2016)	RNCRF+F	SemEval 2014 (Task 4)	English	Laptop	F_1: **78.42%**
Peters et al. (2018)	SpanMlt	SemEval 2014 (Task 4)	English	Laptop	F_1: **77.87%**
Wang et al. (2017)	CMLA	SemEval 2014 (Task 4)	English	Laptop	F_1: **77.80%**
Liu et al. (2019)	CSAE	SemEval 2014 (Task 4)	English	Laptop	F_1: **77.65%**
Li and Lam (2017)	MIN	SemEval 2014 (Task 4)	English	Laptop	F_1: **77.58%**
Yin et al. (2016)	W+L+D+B	SemEval 2014 (Task 4)	English	Laptop	F_1: **75.16%**
Liu et al. (2015)	LSTM+Feat	SemEval 2014 (Task 4)	English	Laptop	F_1: **75.00%**
Chen et al. (2020)	D-GCN	SemEval 2014 (Task 4)	English	Laptop	F_1: **68.53%**
Xu et al. (2020)[2]	DomBERT	SemEval 2014 (Task4)	English	Laptop	Recall: 65.58% Precision: 66.96% F_1: **66.21%**
Peng et al. (2020)[2]	BG+SC+OE (extended)	SemEval 2014 (Task4)	English	Laptop	Recall: 61.55% Precision: 63.15% F_1: **62.34%**
Li et al. (2019b)	BERT-GRU	SemEval 2014 (Task4)	English	Laptop	Recall: 60.47% Precision: 61.88% F_1: **61.12%**
Wu et al. (2018)		SemEval 2014 (Task 4)	English	Laptop	Recall: 66.51% Precision: 55.91% F_1: **60.75%**
Luo et al. (2019a)	DOER	SemEval 2014 (Task4)	English	Laptop	F_1: **60.35%**
Li et al. (2019a)[2]	BG+SC+OE	SemEval 2014 (Task4)	English	Laptop	Recall: 54.89% Precision: 61.27% F_1: **57.90%**
Poria et al. (2016)	CNN-LP	SemEval 2014 (task 4)	English	Restaurant	Recall: 86.10% Precision: 88.27% F_1: **87.17%**
Yan et al. (2021)	BART-ABSA BART-ABSA[4]	SemEval 2014 (Task 4)	English	Restaurant	F_1: **87.07%** F_1: **73.56%**
Liu et al. (2019)	CSAE	SemEval 2014 (Task 4)	English	Restaurant	F_1: **86.65%**
Mao et al. (2021)	Dual-MRC Dual-MRC[2]	SemEval 2014 (Task 4)	English	Restaurant	F_1: **86.60%** F_1: **75.95%**
Wei et al. (2020)	BiSELF-CRF	SemEval 2014 (Task 4)	English	Restaurant	F_1: **86.58%**
Patel and Ezeife (2021)	BERT-MTL-GRU	SemEval 2014 (Task 4)	English	Restaurant	F_1: **86.19%**
Li et al. (2018)	THA+STN	SemEval 2014 (Task 4)	English	Restaurant	F_1: **85.61%**
Peters et al. (2018)	SpanMlt	SemEval 2014 (task 4)	English	Restaurant	F_1: **85.54%**
Wang et al. (2017)	CMLA	SemEval 2014 (task 4)	English	Restaurant	F_1: **85.29%**
Yin et al. (2016)	W+L+D+B	SemEval 2014 (task 4)	English	Restaurant	F_1: **84.97%**
Wang et al. (2016)	RNCRF+F	SemEval 2014 (task 4)	English	Restaurant	F_1: **84.93%**
He et al. (2019)[3, 4]	IMN	SemEval 2014 (task 4)	English	Restaurant	F_1: **84.01%**
Liu et al. (2015)	BiElman+Feat	SemEval 2014 (task 4)	English	Restaurant	F_1: **82.06%**
Wu et al. (2018)		SemEval 2014 (Task 4)	English	Restaurant	Recall: 79.81% Precision: 72.81% F_1: **76.15%**
Peng et al. (2020)[2]	BG+SC+OE (extended)	SemEval 2014 (Task4)	English	Restaurant	Recall: 67.84% Precision: 76.60% F_1: **71.95%**
Yan et al. (2021)	BART-ABSA BART-ABSA[2]	SemEval 2015 (Task 12)	English	Restaurant	F_1: **75.48%** F_1: **66.01%**
Mao et al. (2021)	Dual-MRC Dual-MRC[2]	SemEval 2015 (Task 12)	English	Restaurant	F_1: **75.08%** F_1: **65.08%**
Wei et al. (2020)	BiSELF-CRF	SemEval 2015 (Task 12)	English	Restaurant	F_1: **71.72%**
Li et al. (2018)	THA+STN	SemEval 2015 (Task 12)	English	Restaurant	F_1: **71.46%**
Peters et al. (2018)	SpanMlt	SemEval 2015 (task 12)	English	Restaurant	F_1: **71.07%**
Wang et al. (2017)	CMLA	SemEval 2015 (task 12)	English	Restaurant	F_1: **70.73%**
San Vicente et al. (2015)	EliXa	SemEval 2015 (Task 12)	English	Restaurant	Recall: 71.22% Precision: 68.93% F_1: **70.05%**
He et al. (2019)[4]	IMN	SemEval 2015 (task 12)	English	Restaurant	F_1: **70.04%**

[1] The second result is reported by Li et al. (2020) using data augmentation.
[2] The result is reported for the unified tag setting (aspect+sentiment label).
[3] We ignore the conflict sentiment labels that reduces the overall performance of aspect detection.
[4] The result is reported for the case without document-level (pre-)training.

References	Method	Dataset	Language	Domain	Performance
Peng et al. (2020)[2]	BG+SC+OE (extended)	SemEval 2015 (task 12)	English	Restaurant	Recall: 64.02% Precision: 67.65% F_1: **65.79%**
Wu et al. (2018)		SemEval 2015 (Task 12)	English	Restaurant	Recall: 65.13% Precision: 62.26% F_1: **63.36%**
Karimi et al. (2021)	BAT	SemEval 2016 (Task 5)	English	Restaurant	F_1: **81.50%**
Toh and Su (2016)		SemEval 2016 (Task 5)	English	Restaurant	Recall: 77.09% Precision: 81.02% F_1: **79.01%**
Xu et al. (2019)[1]	BERT-PT	SemEval 2016 (task 5)	English	Restaurant	F_1: **77.97%** F_1: **80.29%**
Xu et al. (2020)	DomBERT	SemEval 2016 (task 5)	English	Restaurant	F_1: **77.21%**
Wei et al. (2020)	BiSELF-CRF	SemEval 2016 (task 5)	English	Restaurant	F_1: **75.56%**
Ma et al. (2019)	Seq2Seq4ATE	SemEval 2016 (task 5)	English	Restaurant	F_1: **75.14%**
Xu et al. (2018)[1]	DE-CNN	SemEval 2016 (Task 5)	English	Restaurant	F_1: **74.37%** F_1: **75.19%**
Li et al. (2018)	THA+STN	SemEval 2016 (Task 5)	English	Restaurant	F_1: **73.61%**
Li and Lam (2017)	MIN	SemEval 2016 (Task 5)	English	Restaurant	F_1: **73.44%**
Chen et al. (2017)	BiLSTM-CRF	SemEval 2016 (Task 5)	English	Restaurant	F_1: **72.44%**
Toh and Su (2016)	NLANGP	SemEval 2016 (Task 5)	English	Restaurant	Recall: 69.44% Precision: 75.49% F_1: **72.34%**
Peng et al. (2020)[2]	BG+SC+OE (extended)	SemEval 2016 (Task 5)	English	Restaurant	Recall: 72.30% Precision: 71.18% F_1: **71.73%**
Xenos et al. (2016)		SemEval 2016 (Task 5)	English	Restaurant	Recall: 69.12% Precision: 71.82% F_1: **70.44%**
Wu et al. (2018)		SemEval 2016 (Task 5)	English	Restaurant	Recall: 70.59% Precision: 58.94% F_1: **64.24%**
Hu et al. (2019)[5]	SPAN	SemEval	English	Restaurant	F_1: **82.38%**
Chen et al. (2020)[5]	D-GCN	SemEval	English	Restaurant	F_1: **77.81%**
Li et al. (2019b)[5]	BERT-SAN	SemEval	English	Restaurant	Recall: 76.72% Precision: 72.92% F_1: **74.72%**
Xu et al. (2020)[2, 5]	DomBERT	SemEval	English	Restaurant	Recall: 74.96% Precision: 72.17% F_1: **73.45%**
Luo et al. (2019a)[5]	DOER	SemEval	English	Restaurant	F_1: **72.78%**
Li et al. (2019a)[2, 5]	BG+SC+OE	SemEval	English	Restaurant	Recall: 71.01% Precision: 68.64% F_1: **69.80%**
Hu et al. (2019)	SPAN	Etter et al. (2013)	English	Twitter	F_1: **75.28%**
Zhang et al. (2015)[6]	Integrated Neural CRF (pipeline)	Etter et al. (2013)	English/ Spanish	Twitter	Recall: 58.57% Precision: 67.08% F_1: **62.42%**
Chen et al. (2020)	DOER	Etter et al. (2013)	English	Twitter	F_1: **62.26%**
Luo et al. (2019a)	DOER	Etter et al. (2013)	English	Twitter	F_1: **51.37%**
Li et al. (2019a)[2]	BG+SC+OE	Etter et al. (2013)	English/ Spanish	Twitter	Recall: 43.56% Precision: 53.08% F_1: **48.01%**
Wang et al. (2015)[7]	SERBM	Ganu et al. (2009) CitySearch Corpus	English	Restaurant	Recall: 67.6% Precision: 83.83% F_1: **74.47%**
Da'u et al. (2020)	MCNN	Dataset developed by	English	Electronic	F_1: **88.55%**
Poria et al. (2016)	CNN-LP	Hu and Liu (2004)	English	Electronic Products	Recall: 86.22% Precision: 90.05% F_1: **88.03%**
Gong et al. (2020)[8]	BERT_E-UDA BERT_E-UDA[2]	Multiple Datasets	English	Mixed	F_1: **45.53%** F_1: **40.63%**
Li et al. (2019c)[3,8]	AD-SAL ADS-SAL[2]	Multiple Datasets	English	Mixed	F_1: **43.91%** F_1: **33.71%**

[5] The result is reported for the union set of Restaurant datasets proposed for the SemEval workshops that addressed ABSA (2014, 2015, 2016).
[6] The result is reported as the weighted average for the *English*, and *Spanish* datasets.
[7] The result is reported as the average for the aspects *Staff*, *Food*, and *Ambiance*.
[8] The result is the average of multiple cross-domain training between four datasets: Laptop dataset (SemEval 2014), Restaurant dataset (SemEval 2014-2016), and the datasets defined by Hu and Liu (2004) and Toprak et al. (2010).

2.3.2 Feed-Forward Neural Networks

FFNN is one of the first neural networks adjusted to find aspects, following only supervised approaches. The simplest FFNN is the perceptron (Collins (2002)), used by San Vicente et al. (2015) to detect aspects as it is indicated in the ixa-pipe-nerc system (Agerri et al. (2014)) proposed for the Named Entity Recognition and Classification. The input is represented by a set of local features at the word-level (e.g., punctuation, previous predictions, suffixes or prefixes) and by the classes of words suggested by three clustering solutions. Despite the simplicity of the method, it is shown to be quite accurate and efficient.

Zhang et al. (2015) choose a Neural CRF (Artieres et al. (2010)) in an approach inspired by Mitchell et al. (2013) to extract both aspects and their sentiment polarities. While in the work of Mitchell et al. (2013), CRF employs a discrete set of features (surface, linguistic, clustering, and sentiment features) defined manually, the features defined by Zhang et al. are continuous word embeddings. According to Zhang et al., the word representations are gleaned by concatenating discrete word features with contextual word embeddings, obtained after summing up all the word2vec word embeddings (Mikolov et al. (2013)) within a given window size. A much straightforward approach is proposed by Xenos et al. (2016) that simply considers word2vec word embeddings as CRF features.

2.3.3 Recurrent Neural Networks

The sequential nature specific to RNNs is the main reason for their extensive use in aspect detection. In this work, the presented RNN-based methods are supervised using either real or noisy labels. Besides the Long Short-Term Memory (LSTM) neural network (Hochreiter and Schmidhuber (1997)), which is the most employed RNN for aspect detection, some RNN variants like Gated Recurrent Unit (GRU) (Cho et al. (2014)), Jordan (Jordan (1997)), or Elman (Elman (1990)) neural networks are also explored.

Supervised Learning. The first RNNs widely applied the idea of using them as input for CRFs, similar to Zhang et al. (2015). A simple example related to this approach is introduced by Toh and Su (2016), where the output of a bidirectional Elman RNN together with other word features like word clusters and word embeddings are used as input for a CRF. Wang et al. (2016) replace the output of an RNN model with a dependency tree that works like a recursive neural network (a generalisation of RNNs). Given a set of dependency relations between words, the representation of each parent word within the sentence is computed as a linear combination between its associated word embedding and the ones associated with its children. When compared with the simple word embeddings (without the representation of children words), the new word embeddings have more syntactic information and are more related to other words of the sentence that act as modifiers and determine the final sentiment polarity. In addition to the extraction of aspects, the method is also designed to detect opinion words by using a tagging scheme that sequentially checks if a word is an aspect, an opinion or a background word.

According to Yin et al. (2016), the computational process for the generation of CRF's input implies the concatenations of the generic word embeddings (i.e., word2vec) with context-aware word embeddings (the sum of representations associated with the neighbouring words), and some word embeddings sensitive to dependency relations. The last type of word embeddings employs a dependency parser, but unlike Wang et al. (2016), the final vectors are not computed based on the parent-child relation but using the shortest path between words. As a result, the representation of a word is not limited only to some dependency relations but takes into account the entire sentence. Namely, given a sentence with n words, the method assigns to a given word a set of $n-1$ word representations based on the shortest path between it and the remaining words. The final word representation is computed as an average of the $n-1$ representations.

In addition to the previously presented methods, the RNN-CRF architecture introduced above is extended by Wei et al. (2020) with a module designed to deal with the boundary errors generated by the incomplete overlapping between the learnt multi-term aspects and real aspects. Given an encoder-decoder system, the input of the module is represented by the candidate aspects extracted by RNN-CRF, while the output indicates the boundaries of aspects. The main advantage of the introduced module is its capacity to increase the performance of models design to find aspects that are affected by learning the aspects only partially.

According to Majumder et al. (2020), the architecture RNN-CRF might also facilitate the aspect-level sentiment classification. Precisely, the hidden states of a bidirectional GRU neural network obtained after learning to identify aspects by using a bidirectional GRU-CRF model can enrich the Glove word embeddings Pennington et al. (2014) used as input for a model design to detect sentiment polarities of aspects. According to the authors' findings, extended word representations boost the performance of the state-of-the-art models for sentiment classification even when different datasets are employed.

Instead of an input-output relation between RNN and CRF, Chen et al. (2017) applies the two methods simultaneously. However, the found aspects only serve for a more-refined sentiment sentence classification. First, aspects are extracted using a neural network with a layer that independently runs both a bidirectional LSTM, and a CRF, and sums up their final results. Further, sentences are classified as non-aspect, one-aspect, and multi-aspect based on the number of generated aspects. The final sentiment prediction is achieved using a CNN trained for each sentence type.

Another direction proposed by Mitchell et al. (2013) and Zhang et al. (2015), is to couple the detection and sentiment classification of aspects as an integrated method using collapsed labels or joint learning. Li et al. (2019a) focused on the collapsed approach for double-label prediction while keeping the multi-task learning for the multiple objectives the method has to meet. First, an LSTM is trained to learn the sequence of aspect terms and used as a guideline for a second LSTM that predicts unified (sentiment, target) labels. The method embodies two refinements. The first one works as a gate mechanism that assures sentiment consistency inside the boundaries of an aspect. The second refinement is used to improve the method convergence by training an additional binary prediction that finds if a word is an aspect term. This task is achieved by a dense layer with a softmax activation function that employs the noisy labels produced by the first LSTM. Li et al.'s work is extended by Peng et al. (2020) by including a new module dedicated to the opinion term extraction. The module is both addressed as an auxiliary task and also integrated into the inner working of Li et al.'s method, affecting the final performance of aspect detection.

The stacked LSTMs proposed by Li et al. (2019a) (without the two refinements) are also adopted by Li et al. (2019c) as a baseline to predict the unified aspect-sentiment labels. To enrich the prediction quality, a multi-hop Dual Memory Interaction (DMI) system is inserted between LSTMs to capture the connections between aspects and opinion words. The DMI system is an iterative system whose output is represented by two vectors for aspects and opinion words that comprises both intraclass and interclass correlations with respect to the words of the input sentences. Given that the first LSTM layer is trained for the extraction of aspects, only the correlation vector associated with aspects is used as input for the second LSTM layer that predicts the final joint label. The second correlation vector for opinion words is used as input for an auxiliary opinion detection task based on a fully connected layer with a softmax activation function. The method has a secondary aim of easing the knowledge transfer from a source domain to another target domain. This aim is achieved using selective adversarial learning via a Gradient Reversal Layer (Ganin et al. (2016)) that maximises the loss function associated with the identification of the domain label at the word level.

The second option for the end-to-end ABSA approaches based on joint learning is exploited by Luo et al. (2019a). The proposed method encapsulates the dual word embeddings introduced by Xu et al. (2018) as input for a stack with two RNN layers designed to predict aspect labels. Simultaneously, the first RNN is applied for the prediction of sentiment polarities. The RNN is actually a novel Residual Gated Unit (ReGU), specially proposed for the given work. Like LSTM, ReGU has both memory cells and hidden states as inner representations. However, the information flow is controlled using only two gates (forget and residual gates) instead of using three gates like LSTM. Given that the two ABSA tasks are strongly correlated, a shared-task attention layer is inserted between the first and the second layer of ReGUs to transfer knowledge between tasks. In the end, the prediction of each task is done by a CRF. However, the main aim of the paper is to provide joint aspect-sentiment labels. This task is achieved by simply aggregating the two task-specific labels. To meet the sentiment consistency requirement, the most frequent or the first sentiment label is chosen inside the aspect boundaries. For the enhancement of the method, three additional tasks are learned to focus on the prediction of the length of aspects and sentiments (the number of consecutive and unique aspect and sentiment labels), and on the prediction of the sentiment label of a word (using a sentiment lexicon).

Opinion words represent another direction approached by deep learning methods. For example, in the work of Li and Lam (2017), a self-defined Memory Interaction Network (MIN) extracts aspects terms using an LSTM layer with memory vectors of opinion words weighted based on a clustering method and positional indices. Then, a sentiment sentence classification works like a filter that enforces the aspects to be found only in the sentiment-bearing sentences. Li et al. take the idea of memory further and claim that the long-term aspect-opinion word dependencies are more effective than the local ones based on positional indices. Besides the opinion words-specific memory, the proposed method caches the aspects' history useful not only to reduce the error of the next predicted tag (i.e., the OBI tag I can not follow O tag) but also to capture infrequent aspects found in coordinate structures. The method employs multi-task learning for predicting both aspect and opinion words via two LSTMs trained separately for the two tasks. However, the extraction of opinion words works more like an auxiliary task that together with the aspects' history enhance the quality of the model.

While the above methods extract the opinion and aspect terms using multi-task learning, Wang et al. (2017) replace the dual approach with a tagging scheme (used also by Wang et al. (2016)) that detects whether a word is an aspect, opinion or background word. The presented method exploits the idea of replacing the relations indicated by a dependency tree with relations pointed out by attention scores. This makes sense especially when the raw data is represented by reviews with short sentences and written in an informal style. The introduced process is iterative, with the hope that a higher number of iterations will increase the capacity of the method to learn a broader range of dependencies between aspects and opinion words. First, the method generates contextual word representations using a GRU neural network. Then, each sentence is encoded in two prototype vectors representative for its opinion and aspects words. The two prototype vectors are used to compute high-level representations for all words of the input sentence with respect to their context (via another GRU network) and a set of attention scores that ranks the likelihood of a word to be an aspect or opinion term. Further, the attention scores and the new word representation are used to learn the prototype vectors of the next iteration.

In addition to the above RNN-based methods that mostly rely on LSTMs or bidirectional LSTMs, Liu et al. (2015) also assess the performance of (bidirectional) Elman, and Jorden RNNs in terms of aspect detection. Using as input word embeddings expanded with linguistic features like POS tags, Liu et al.'s results proved that (bidirectional) LSTM is not always the best option and bidirectional Elman RNN, can surprisingly lead to better results, depending on the employed dataset.

Weakly supervised learning. While the above methods use real labels, the method introduced by Wu et al. (2018) is trained using inexact or noisy labels. The extraction of training labels follows the typical pattern-based approach for the selection of aspects. First, the nouns or noun phrases extracted by a POS tagger are deemed candidate aspects if they satisfy a set of patterns and POS constraints. Then, candidate aspects are evaluated in terms of similarity with the given domain and filtered. In the end, all the found aspects considered noisy labels are used to sequentially train a bidirectional GRU that employs expanded word embeddings with POS word representations. A similar approach was also implemented by Chauhan et al. (2020). The major difference stands in the replacement of the bidirectional GRU with an attention-based LSTM model.

2.3.4 Convolutional Neural Networks

CNNs (LeCun et al. (1999)) are used as well to identify aspects, but, while for the aspect classification of a word, RNN gathers information from the entire input data, the prediction of CNN is sensitive only to the neighbouring words over which the inner filters convolves. Currently, all the CNN-based methods for aspect identification are supervised.

The first deep neural network with multiple convolutional layers was proposed by Poria et al. (2016) The method is applied only over a small chunk of text that corresponds to a word (the one for which the method checks whether it is an aspect) and to its given context (the left and right sides). Sequential data processing is assured by splitting the input sentence into n-grams and applying CNN over each one. In addition to CNN, a set of lexical patterns is employed.

A set of terms is considered an aspect if at least the CNN-based method or the patterns indicate so.

A simpler method is proposed by Xu et al. (2018), where a neural network with four convolutional layers followed by a perceptron utilise dual word embeddings (general word embeddings - trained on large corpora, and domain word embeddings - trained on small datasets) as input. While the network is applied over the entire input sentence, the final dense layer with a softmax activation function is applied sequentially, for each word of the sentence. Therefore, Xu et al.'s CNN has a lower computational complexity compared to Poria et al.'s method. Instead of concatenating the input word embeddings like Xu et al. (2018), the CNN neural network proposed by Da'u et al. (2020) exploits a multi-channel approach with different inputs. As the aim of the paper is to provide a recommender system, the proposed framework also includes a module for aspect categorization based on the Latent Dirichlet Allocation (Blei et al. (2003)).

Different from the previous methods designed exclusively for the extraction of aspects, He et al. (2019) propose an end-to-end approach to carry out both ABSA tasks. The method runs in two steps for each ABSA task, each step being represented by a CNN-based encoder. The output of the first encoder is a shared set of latent word representations used as input for the second task-specific encoder. The method has two auxiliary classification tasks that assign domain (Restaurant and Laptop for the case of SemEval datasets) and sentiment labels at the document-level. These two tasks are jointly trained with aspect-level tasks and mitigate the problem of small training sets. The interesting part of the network is the interactive multi-task learning which assumes a pre-training of the method exclusively on the document-level tasks, and then an alternate training on aspect and document-level tasks.

2.3.5 Attentive Neural Networks and Transformer-Based Models

ANNs use attention weighting, a mechanism widely applied in the development of deep neural networks. The attention mechanism has control over the entire input data like RNN, but it does not behave like a black box and is more comprehensible. Like the majority of deep learning methods, ANNs methods presented in this section employ supervision as a learning approach. Sequence to sequence learning (seq2seq), one of the most common used ANN-based methods, is introduced in the field of aspect detection by Ma et al. (2019) by means of two bidirectional GRUs for both the encoder and the decoder modules. While the seq2seq framework stores the entire meaning of the input text due to the encoder module and can control the dependencies between consecutive labels (like Li et al. (2018)), it faces some limitations for which some improvements are introduced. The first improvement enhances the standard context vectors with position-aware attention scores that penalise the long-distance words. Secondly, the framework also embodies gated unit networks that help the model to learn particularities shared by aspects.

More recently, the attention mechanism has evolved into the Transformer encoder-decoder (Vaswani et al. (2017)) that benefits of both the parallelisation specific to CNN and the control over the long-term dependencies specific to RNN. Since the beginning, the Transformer architecture has evolved into a wide range of models including GPT (Radford et al. (2018)) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. (2018)). Given that GPT represents a left-to-right Transformer decoder and BERT is a bidirectional Transformer encoder, Lewis et al. (2020) pack all together and obtain Bidirectional and Auto-Regressive Transformers (BART). While not as well-known as GPT and BERT, the BART model is still used by Yan et al. (2021) to generate sequential predictions. The proposed method is designed in such a way to cover all tasks associated with ABSA providing predictions not only for aspect terms but also for opinion words and sentiment polarities.

Despite the wide range of Transformer-based models, in terms of ABSA and especially aspect extraction, the majority of works make use only of BERT, usually required to generate contextual word embeddings. The reason behind the wide applicability of BERT is its effectiveness so that even the most simple method with its word embeddings and a dense layer with a softmax activation function proves to be competitive with other state-of-the-art methods for aspect detection (Li et al. (2019b)). Similar to the work of Li et al. (2019b), Phan and Ogunbona (2020) also identify aspects by running a simple fully connected layer over word representations gleaned from POS and dependency information, and over word embeddings produced by the enhanced version of BERT (RoBERTa) introduced by Liu et al. (2019).

Further on, the work of Karimi et al. (2021) aims to demonstrate the effectiveness of considering extra adversarial input over the method's performance. The adversarial examples are generated based on the perturbed input sentences that should minimise the loss function. The method has a simple architecture with BERT word embeddings and a fully connected layer, but given that it is trained separately for the two inputs (real and adversarial inputs), it resembles the multi-task learning approach.

Besides the linear approach (with a fully connected layer with a softmax activation function), Li et al. (2019b) evaluate the effectiveness of BERT word embeddings in terms of GRU, self-attention, TFM (a variant of BERT), and CRF layers. While GRU provides the best results for the Laptop dataset of the SemEval 2014 workshop, self-attention is the best option for the combined Restaurant datasets proposed for the three SemEval workshops that approached ABSA. A similar architecture embodying a BERT encoder and additional FFNN, LSTM and GRU layers is also employed by Patel and Ezeife (2021). The difference lies in the multi-task learning approach used to predict not only aspect labels but also high-level aspects (aspect categories). Since the extraction of coarser aspects is executed as a classification at the sentence-level, we consider this task to be similar to the extraction of implicit aspects. Therefore, the result of this task is presented in the section dedicated to the implicit aspects.

The typical BERT word embeddings are pre-trained using BookCorpus and Wikipedia dumps. However, these two datasets might not be flexible enough to fit all the NLP tasks the BERT word embeddings are used for. The solution proposed by Xu et al. (2019) is the post-training of BERT on new datasets to confer task and domain awareness. Given that the method has to fit the SemEval Laptop and Restaurant datasets, the domain post-training is done using unlabeled Yelp and Amazon laptop reviews corpora (McAuley and Yang (2016)). For task awareness, the authors use the SQuAD dataset (Rajpurkar et al. (2016)), designed for question answering which might facilitate the aspect detection task. The reason behind the choice of the SQuAD dataset is justified by the authors' decision to consider the extraction of aspects as an adjacent task of a question answering process focused on product reviews.

Starting from the idea of Xu et al. (2019), Xu et al. (2020) develop a new domain-orientated language model as an extension of BERT (DomBERT). Removing the next sentence prediction task of the traditional BERT, the new model keeps the masked-language modeling and add a new task that learns domain labels per sentence. Knowing that DomBERT is trained on a large number of domain-specific corpora (Yelp dataset and Amazon review datasets (Xu et al. (2020))), a domain sampler is included to select the next most similar domain with respect to the current one. Next, all instances of the new domain-specific corpus feed the DomBERT model, and a new domain is again selected.

Given that the input representation plays an important role in obtaining effective models, Chen et al. (2020) propose a more elaborate solution than the previous ones relying on BERT word embeddings as input for a graph convolutional network (GCN) defined with the help of a dependency tree of the input. The method improves the traditional GCN by learning different weight matrices with respect to the positional relationships between words of a sentence, and attention scores that neglect words without dependencies. In the end, the model jointly predicts not only aspects, but also sentiment labels.

Another method that embodies BERT word embeddings is proposed by Hu et al. (2019) as a streamlined alternative to the widely used sequential labelling solution. Specifically, two one-layer FFNNs assess the probabilities that a word of a sentence represents either the starting or the ending token of an aspect. As multiple aspects might occur in the same sentence, the method leverage a heuristic algorithm that keeps only the shortest aspects and removes the overlapping cases. In addition, this work also integrates a module for the detection of the sentiment polarities of aspects. According to the reported results, the pipeline solution that analyses the aspect extraction and sentiment classification separately works better than the multi-task learning and collapsed labels. Further on, the Hu et al. (2019)'s prediction approach is extended by Mao et al. (2021). Using the BERT word embeddings, the proposed method employs the multi-task learning approach and predicts the starting and ending position of the aspects and opinion expressions, and the sentiment polarities using the [CLS] token learnt by BERT. The result of the model is a set of triples (aspect terms, opinion expressions, sentiment polarity) at the sentence-level.

In addition to the prediction of OBI labels or the starting or ending aspect positions as Hu et al. (2019) and Mao et al. (2021), Zhao et al. (2020) split the input sentence into chunks of a length smaller than a given threshold and classify them as aspects or non-aspects. Instead of using binary classification, the model is also trained to detect whether the given chunk contains opinion words. Besides learning aspects and opinion words, the second task the model is prepared to learn is related to the detection of the pair-wise dependencies between them. In the end, the model follows a multi-task learning approach that runs on a Bidirectional LSTM encoder with ELMo word embeddings (Peters et al. (2018)). Alternatively, the model embodies a BERT encoder, but the approach is less effective than the LSTM-based encoder.

The problem of label unavailability is tackled by Gong et al. (2020) by defining two components to solve the discrepancies between two domains in terms of features and instances. Considering that, despite the particularities specific to each domain, the aspect-related language structures are universal, the feature-based component aims to learn POS tags and dependency relations of the target and source domains. To learn the POS tags, the BERT encoder is adjusted to predict masked POS tags. Unlike the traditional BERT model (Devlin et al. (2018)), the input representations of each word result after summing not only context-independent word embeddings, position representations, and segment embeddings, but also POS tag embeddings. The newly obtained context-aware BERT word embeddings are used to learn dependency relations and to provide the input needed to run the instance-based component. Considering that aspects are more related to their domain than other words, the instance-based component learns the domain-related word distribution first, and then, predicts the final aspect labels.

Differently from most of the presented papers, Li et al. (2020) address the problem of data augmentation for the aspect detection task. Given the particularities of the task, the novel approach should yield new instances that are not only opinionated texts, but also share the same labelling sequence with the original instances. According to Li et al. (2020), this problem is solved using a Transformer-based encoder-decoder system that generates new instances following a masked sequence-to-sequence strategy.

3 Implicit Aspects

Implicit aspects do not benefit from a positional index within the sentences and are extracted only conceptually. As long as there is no constraint about the presence of aspects inside the input sentences, aspect category detection (ACD) introduced in the SemEval workshops can be considered a subtask of the extraction of implicit aspects. Therefore, all the presented works of this section that refer to the detection of implicit aspects as a task of class labeling are good candidate solutions for the ACD task. All methods suitable for ACD are specified in Tables 4-6.

With regards to the aspect detection methods designed for both implicit and explicit aspects, their approaches are already introduced in the previous sections, and we only present the adjustments for the detection of implicit aspects in the current section. Sections 3.1, 3.2, and 3.3 are dedicated to the pattern-based, machine learning and deep learning methods introduced for the detection of implicit aspects. The performances results of the three approaches together with the details about the employed datasets are listed in Tables 4, 5, and 6, respectively.

3.1 Pattern-Based Methods

Unlike for explicit aspects, the pattern-based methods for implicit aspects are mainly supervised. This might be the result of the similarity between the extraction of implicit and high-level aspects that usually requires the injection of prior knowledge.

Unsupervised Learning. Following the idea of creating aspect-opinion rules, extensively exploited in the section dedicated to explicit aspects, Poria et al. (2014) also propose a set of rules that match both implicit and explicit aspects. Along with employing patterns, the solution found by Schouten et al. (2017) for the detection of implicit aspects consists in looking for their descriptors in the sentences. Using the Spreading Activation method (Anderson et al. (1983)), the main goal is to map the sentence words to a set of seed words representative for the given aspects in terms of similarity. Based on a chain reaction, the more a word co-occurs with a seed word or a similar word in a sentence, the higher the similarity score is. In the end, one or more high-level aspects are allocated to a sentence based on their related words found inside the sentence.

Supervised Learning. Some of the pattern-based methods can detect both explicit and implicit aspects. As the methods do not set a clear delimitation between these two types of aspects, we present these methods in both sections. Two such examples are the methods proposed by Liu et al. (2005) and Liu (2010), where the implicit aspects are extracted using association and sequential rules, respectively. Since all the aspect references are annotated in the training data, the methods can detect all types of aspects.

According to Dosoula et al. (2016), an implicit aspect is allocated to a sentence if its aggregated co-occurrence score for all the words of the sentence computed based on the training data is higher than a given threshold. Additionally, a binary logistic regression is used to indicate whether the sentence is allowed to have multiple implicit aspects. The binary classifier has as input a set of sentence characteristics like the number of nouns, adjectives, commas, and "and" conjunction. If the result of regression is positive, then all the aspect labels with a co-occurrence score higher than the given threshold are allocated to the sentence. The main disadvantage of this approach is the need for a large training data.

Along with the co-occurrence between words and high-level aspects (Dosoula et al. (2016)), the second proposed method by Schouten et al. (2017) also leverage on the co-occurrence of aspects with multiple dependency form sets. The solution is a supervised variant of the Spreading Activation method mentioned above and dubbed by the authors as the Probabilistic Activation method. The final aspect label of an out-of-training sentence is determined by the highest co-occurrence score of an aspect with the words and dependency relations presented in the sentence.

Semi-supervised Learning. In addition to the above supervised methods, Zhai et al. (2010) introduce a semi-supervised method to expand the training data by means of the Expectation-Maximization algorithm (Dempster et al. (1977)). The method maps aspect terms or aspects of a smaller level of abstraction (already extracted) to high-level aspects or entities. For the starting point of the method, a set of seed aspect expressions is allocated to each high-level aspect. Additionally, to the seed aspect words used as hard or inflexible labels, the method appends some soft labels used as a guideline and available only in the first iteration of the Expectation-Maximization algorithm. The soft labels are an extension of the hard labels and their extraction is derived from the presence of a shared word or synonym (found using WordNet) with the seed expressions. One should note that the method proposed by Zhai et al. (2010) is similar to the Spreading Activation method employed by Schouten et al. (2017). However, since the final aim of Zhai et al. is to expand the training set via an Expectation-Maximization approach, we present the two methods in different sections.

Table 4 Pattern-based methods for the detection of implicit aspects

References	Method	Dataset	Language	Domain	Performance
Schouten et al. (2017)[1]	Supervised Method	SemEval 2014 (Task 4)	English	Restaurant	Recall: 83.1% Precision: 84.4% F_1: **83.8%**
Schouten et al. (2017)[1]	Unsupervised Method	SemEval 2014 (Task 4)	English	Restaurant	Recall: 64.7% Precision: 69.5% F_1: **67.0%**
Dosoula et al. (2016)[1]		Dataset developed by Ganu et al. (2009) Ganu et al. (2009) CitySearch Corpus	English	Restaurant	F_1: **64.5%**
Zhai et al. (2010)[1, 2]	SC-EM	Self-defined Dataset	English	Mixed	Accuracy: 68% Purity: 69% Entropy: 1.24

[1] The method is also suitable for the ACD task.
[2] The result is reported for the aspects *hometheater*, *car*, *insurance*, *vacuum*, and *mattress*. The percentage of labeled aspect terms (hard labels) is 30%.

3.2 Machine Learning Methods

Similar to the pattern-based methods, machine learning methods currently detect implicit aspects using either their descriptors or via a multi-label classification at the level of a text unit. SVM and Maximum Entropy discriminative methods are used especially for the classification of high-level aspects and are always supervised. On the other hand, the generative class is represented only by topic modeling techniques that provide clustering solutions, and are useful especially for unsupervised learning. In terms of topic modeling, a topic represents an aspect (usually high-level), and its word distribution may comprise both explicit aspects and descriptors.

However, along with the key words like aspects and descriptors, the topic word-distributions include a wide range of irrelevant word expressions. As a result, numerous topic modeling techniques proposed for aspect identification come with a set of adjustments to alleviate this problem. For example, Mei et al. (2007) use a sentiment topic mixture model to filter out the sentiment-bearing words inside the topics. Similarly, Titov and McDonald (2008) propose the Multi-Grain LDA to capture aspects with different levels of abstraction.

According to Mukherjee and Liu (2012), the problem of selecting the right aspect terms inside topics is simply enforced by a set of seed words used as a guideline. Besides the problem of inferring the right aspect terms within topic word-distributions, the topic modeling techniques have to cope with the problem of selecting a correct number of clusters. This problem is solved by Brody and Elhadad (2010), by running a consistency test (Niu et al. (2007)) that measures the robustness of the clustering solution against resampling. Even if there are plenty of topic modeling techniques adjusted for aspect detection, we choose not to present them in the current survey due to the lack of comparative results. The remaining two sections present the SVM and Maximum Entropy classifiers, respectively, used for the classification of high-level aspects.

3.2.1 Support Vector Machines

The standard implementation of SVM is straightforward, and according to Kiritchenko et al. (2014), it simply relies on a one-vs-all approach. Likewise, the method proposed by Xenos et al. (2016) generates the aspect labels using a linear combination between two SVMs. The first SVM is constrained to use only the

Table 5 Machine learning methods for the detection of implicit aspects

References	Method	Dataset	Language	Domain	Performance
Kiritchenko et al. (2014)[1]	NRC-Canada	SemEval 2014 (Task 4)	English	Restaurant	Recall: 86.24% Precision: 91.04% F_1: **88.58%**
Xenos et al. (2016)[1]		SemEval 2016 (Task 5)	English	Restaurant	Recall: 69.12% Precision: 71.82% F_1: **70.44%**
Xenos et al. (2016)[1]		SemEval 2016 (Task 5)	English	Laptop	Recall: 53.19% Precision: 45.60% F_1: **49.10%**
Kim and Hovy (2006)[2, 3]		FrameNet Dataset	English	Mixed	Recall: 69.53% Precision: 68.16% F_1: **68.74%**
Kim and Hovy (2006)[1, 3]		Self-defined Dataset	English	News Media	Recall: 13.95% Precision: 61.75% F_1: **22.1%**

[1] The method is also suitable for the ACD task.
[2] The result is reported as the weighted average for verb and adjective opinion words.
[3] The result is reported as the average by taking into account the opinion words (verbs and adjectives) labeled by two annotators.

local training data (SemEval data) and runs a binary classification for each aspect at the sentence-level using a set of unigram and bigram features. The input of the second SVM is the centroid of a sentence expanded with its similarity scores with respect to the given aspects. The centroid is determined by the sum of word embeddings trained on an external dataset and weighted with their tf-idf scores.

3.2.2 Maximum Entropy Methods

Maximum Entropy classifier is used as an alternative for SVM by Kim and Hovy (2006) in a weakly supervised learning approach, that relies on labels provided by a lexical corpus dubbed FrameNet II (Baker and Sato (2003)). The first step consists in the recognition of sentiment-bearing words (adjectives and verbs). Then, the method uses the opinion words as lexical units or roots for different semantic roles (the target aspects) in a broader frame indicated by the FrameNet II corpus. If an opinion word is out of the list of lexical units, the method attaches to it the list of semantic roles belonging to the most similar unit found based on the Clustering by Committee algorithm (Pantel and Lin (2002)). The method is employed to predict not only the target aspects but also the holders (the opinion's owner).

3.3 Deep Learning Methods

Neural networks, presented again separately from the machine learning methods, leverage both on supervised and unsupervised learning for the detection of implicit aspects. The introduced methods relies on ANNs and Transformer-based models (Sect. 3.3.1), FFNNs (Sect. 3.3.2), and RNNs (Sect. 3.3.3).

3.3.1 Attentive Neural Networks and Transformer-Based Models

The current section is dedicated to the models developed based on the attention mechanism. While most of the models are unsupervised, Patel and Ezeife (2021) present a supervised method where a BERT encoder part of a Transformer architecture is used to assign coarser aspect labels per sentence and to detect aspects explicitly mentioned. Since the classification of sentences in terms of aspects is similar to the extraction of implicit aspects, this method is mentioned again in this section. Besides Patel and Ezeife (2021)'s work, all the remaining neural networks are unsupervised. As we can not be sure if the found terms are only explicit aspects, we decided to consider these methods as designed for the extraction of implicit aspects.

Using an encoder-decoder architecture with multiple iterations, the method of He et al. (2017) can capture both explicit and implicit aspects. At each iteration, the system generates a latent variable that intermediately represents the input sentence and shows the probabilities that the sentence refers to the different aspects (the encoder part). The aspects are unknown, except for their number. Next, the latent variable reconstructs the sentence representation used as input for the next iteration (the decoder part). Besides the sentence representations, the method also generates a set of aspect embeddings. The process is trained to minimise the difference between consecutive sentence representations and to maximise the difference between the sentence representation and other randomly chosen sentences in the corpus. The method is trained using multi-task learning, focused not only on the sentence representations but also on the diversity of aspects. The final aspect terms are found in the embedding space using cosine similarity between their representations and the newly computed aspect embeddings.

Starting with He et al.'s work, Luo et al. (2019b) aim to refine the model by implementing a different initial sentence representation for the first iteration of the encoder-decoder system. While He et al. merely use self-attention for input sentence embeddings, Luo et al. generate sentence representations as linear combinations of all sememes (groups of words with similar meanings) associated with the sentence words. Sememes are extracted using WordNet (Miller (1995)). To prevent the tendency of the method to forget the context of the sentence, the initial sentence representation is enriched with the final hidden state of an RNN-based structure.

Angelidis and Lapata (2018) notice that the dynamic approach of He et al.'s method could lead to the extraction of both high-context and low-context aspects (which might not be independent of each other, but into a hierarchical relation) simultaneously. In their work, the authors choose to focus only on the detection of high-context aspects and to leverage on their descriptors (low-level aspects) to generate the aspect matrix. While the aspect matrix defined by He et al. (2017) is the output of the proposed encoder-decoder system, according to Angelidis and Lapata, each row of the aspect matrix refers to a high-level aspect and is computed as a weighted average of its descriptors, a priori extracted. As a result, the new encoder-decoder system is used only to reconstruct the input sentences. The neural network is still trained in a multi-task fashion but instead of trying to improve the diversity of aspects, the model predicts the high-context aspect per sentence.

A much more simple, but effective unsupervised approach for the detection of high-context aspects is provided by Tulkens and van Cranenburgh (2020). Initially, the set of most frequent nouns is detected as candidate aspects. Next, the Radial Basis Function kernel is used to calculate attention scores for each sentence word with respect to the candidate aspects. Finally, the new probability distribution over words provides a new document representation for which a new aspect label is assigned based on cosine similarities with aspect representations.

Table 6 Deep learning methods for the detection of implicit aspects

References	Method	Dataset	Language	Domain	Performance
Patel and Ezeife (2021)[1]	BERT-MTL	SemEval 2014 (Task 4)	English	Restaurant	F_1: **90.18%**
Toh and Su (2016)[1]	NLANGP	SemEval 2016 (Task 5)	English	Restaurant	Recall: 73.62% Precision: 72.45% F_1: **73.03%**
Toh and Su (2016)[1]	NLANGP	SemEval 2016 (Task 5)	English	Laptop	Recall: 47.81% Precision: 56.85% F_1: **51.94%**
Ma et al. (2018b)[1]	Hybrid Sentic LSTM + TA + SA	SentiHood	English	Locations - London	Accuray: 67.52% Macro F_1: **78.10%** Micro F_1: **77.87%**
Ma et al. (2018a)[1]	Sentic LSTM + TA + SA	SentiHood	English	Locations - London	Accuray: 67.43% Macro F_1: **78.18%** Micro F_1: **77.66%**
Tulkens and van Cranenburgh (2020)[2]	CAt	Ganu et al. (2009) CitySearch Corpus	English	Restaurant	Recall: 86.4% Precision: 86.5% F_1: **86.4%**
Luo et al. (2019b)[2]	AE-CSA	Ganu et al. (2009) CitySearch Corpus	English	Restaurant	Recall: 81.83% Precision: 82.5% F_1: **82.1%**
He et al. (2017)[2]	ABAE	Ganu et al. (2009) CitySearch Corpus	English	Restaurant	Recall: 72.23% Precision: 85.66% F_1: **77.5%**
Luo et al. (2019b)[3]	AE-CSA	McAuley et al. (2012) Beer Advocate	English	Beer	Recall: 81.83% Precision: 82.5% F_1: **82.1%**
He et al. (2017)[3]	ABAE	McAuley et al. (2012) Beer Advocate	English	Beer	Recall: 72.72% Precision: 71.16% F_1: **69.54%**
Angelidis and Lapata (2018)[4]	Mate+MT	Self-defined	English	Mixed	F_1: **49.1%**

[1] The method is also suitable for the ACD task.
[2] The result is reported as the average for the aspects *Staff*, *Food*, and *Ambiance*.
[3] The result is reported as the average for the aspects: *Feel*, *Smell*, *Look*, *Taste*, and *Overall*.
[4] The result is reported for the aspects *laptop Bag*, *boot*, *bluetooth headset*, *vacuum*, *keyboard*, and *television*. The dataset is a subset extracted from the Amazon Product Dataset (McAuley et al. (2015)).

3.3.2 Feed-Forward Neural Networks

The method introduced by Toh and Su (2016) determines high-level aspects using a supervised FFNN with a singular layer. The network works like a binary classifier that decides whether an aspect conceptually exists in a sentence. The input representation is similar to the one introduced for the extraction of explicit aspects. The only difference lays in the replacement of the bidirectional Elman neural network output with a CNN output. While the bidirectional Elman neural network is trained to predict the OBI tags of words, CNN assigns high-level aspects at the sentence-level.

3.3.3 Recurrent Neural Networks

The supervised approach proposed by Ma et al. (2018a) uses as a basis a variation of the LSTM neural network dubbed Sentic LSTM that incorporates affective concepts suggested by AffectiveSpace (Cambria et al. (2015)) to better control the information filtering. Further, a hierarchical attention mechanism is applied over the hidden states of the Sentic LSTM to generate two representations of the aspect terms (low-level aspects explicitly mentioned in the sentences), and of the input sentences. In the end, a softmax layer applied over the new encoded sentence representation evaluates the presence of high-level aspects within the sentence. The Sentic LSTM is refined by Ma et al. (2018b), in an approach inspired by the Recurrent Additive Networks (Lee et al. (2017)). The novel neural network reduces the complexity of the Sentic LSTM and slightly outperforms it in terms of aspect detection due to its additive nature.

4 Discussion

Aspect-based sentiment analysis is a process in two steps that identifies the topics within the documents and assigns them sentiment scores. In the current work, we concentrate on the first step of the analysis and present the main research directions with a special focus on the more recent works. Sections 4.1, 4.2, and 4.3 discuss the surveyed pattern-based systems, machine learning methods, and neural networks designed for the extraction of aspects.

4.1 Pattern-Based Methods

Currently, the learning approach of the pattern-based systems is related to the nature of aspects. Regarding explicit aspects, their extraction mainly leverages on unsupervised learning. As a result, many works usually require a two-step process that first selects candidate aspect expressions and then revises their quality. On the other hand, implicit aspects are usually deemed as high-level aspects or categories of aspects. Based on this perception, many pattern-based systems for implicit aspects require previously annotated training data.

The most common idea that defines the unsupervised learning of explicit aspects is the assumption that aspects are usually nouns (Liu (2011)). Following this idea, many authors introduce patterns for modeling the relation between nouns and opinion words, other aspects (already extracted), and different structures that might suggest the presence of multi-term aspects. The next step that follows the generation of the recommended list of aspects is the elimination of aspects with a low quality. The current pruning methods vary from using tf-idf indicators, domain-related frequencies, or Gini scores to the counting of related sentiment words. As the patterns might detect wrong aspects, some methods are concerned with the pattern ranking, keeping only the most effective patterns.

Regarding implicit aspects, unsupervised learning is not applied to such a great extent as in the case of explicit aspects. The few works in this direction are focused on the aspect's descriptors found using patterns defined either by the experts or by means of the search graphs that detect similarities between a set of seed words and the words of a sentence.

Supervised pattern-based systems are less common for aspect detection and usually make use of the co-occurrence of the high-level aspects with the words or POS structures specific to a sentence. As supervised learning draw inferences from a set of labeled instances, another approach is to annotate all aspect references (explicit aspects and descriptors) and to learn new aspects using association mining.

Even if some supervised methods have been proposed for the pattern-based systems, they are conventionally designed for unsupervised learning. Their major asset is domain independence, which means that each collection of reviews can be a candidate input for the method, regardless of the availability of the annotated data. However, this advantage is often traded off for a poorer performance. To capitalise on the benefits of pattern-based systems, probably the best solution is to use the patterns as a guideline for a more effective machine learning method that can predict an aspect only if a pattern validates it. As a result, the reinforced method might be not only more effective but also more portable between domains. Another direction to be explored by future works might be the use of pattern-based methods as generators of aspect labels. Given the new labels, one might consider them as the gold standard for a more effective method in a weakly supervised approach. Also, the new labels can be utilised to learn automatically new aspect patterns that are not only more efficient than the manual ones but might be even more effective (Mangnoesing et al. (2020)).

4.2 Machine Learning Methods

Excepting the neural networks presented in a different section and the topic modeling techniques that are not introduced in this paper because of the unavailability of the results, the current machine learning methods are exclusively supervised. CRFs and HMMs are the main machine learning methods employed for the sequential detection of explicit aspects. However, despite their similarities, only CRFs are extensively used due to their discriminative nature (HMMs have a generative nature). Reagrding implicit aspects, the current methods are limited only to SVM and Maximum Entropy classifiers (the only SVM presented in the section of explicit aspects is actually used as a pruning method).

Similar to the pattern-based systems, we consider that future works should mainly take into account hybrid methods that combine classical machine learning models with deep neural networks or with prior information provided by a pattern-based method. In doing this, one might consider the encapsulation of different emission and transition features like sentiment polarities, dependencies relations, or word similarities in the future more effective solutions. Or, the final conventional fully connected layer might be replaced with a CRF or HMM, more effective for sequential tasks with highly related tags (Reimers and Gurevych (2017)). Along with these design particularities, another important direction for future work is the lifelong approach required for the knowledge transfer between different domains. Besides the explicit aspects, we also consider that the detection of implicit aspects needs much more attention. Hence, it would be interesting to see a wider variety of classifiers applied in this field and enriched with knowledge about aspects' descriptors.

4.3 Deep Learning Methods

Even if deep learning methods have high flexibility for any learning type, a great extent of methods are created only for supervised learning. The reason behind this choice might be the desire of authors to promote their works using the widespread SemEval labeled data. While the supervised detection of explicit aspects tries to reproduce the sequentiality of the input data, the implicit aspects are treated as high-level aspects via a multi-label supervised classification.

Unsupervised learning is used for both types of aspects, but its applicability is limited. The unsupervised methods are inspired by the ANN model proposed by He et al. (2017) and built as an encoder-decoder system whose output is a set of aspect embeddings. As the embeddings may not refer only to explicit aspects but also descriptors or even background words, we consider this baseline and its variants more suitable for the extraction of implicit aspects. As regards the unsupervised detection of explicit aspects, we found only the RBM neural network proposed by Wang et al. (2015). Despite the unsupervised nature of the method that infers a higher level of uncertainty about the type of the detected terms than the supervised methods, we consider that the configuration of the method together with the injection of the prior linguistic information make the proposed RBM more suitable for the extraction of explicit aspects.

FFNNs with at most one hidden layer were among the first supervised neural networks used to identify aspects. Soon after FFNNs, CNNs were introduced but only for the detection of explicit aspects. Like FFNN, CNN is not a sequential method, which means that the sequential dense layer with s softmax activation function is required. However, another proposed option is to apply CNN only over a chunk of text and to slide the entire neural network over the remaining words. The downside of this method is the higher computational complexity.

The most widespread methods for the detection of aspects are the sequential RNNs. While the detection of implicit aspects relies only on the Sentic LSTM and its variant for the multi-label classification, RNNs adjusted for explicit aspects are more diverse. Some of the first RNNs introduced for the latter direction are combined with CRFs either in an input-output relation or applied simultaneously. Another RNN solution for explicit aspects is to deliver an end-to-end approach for both ABSA tasks using unified labels or multi-task learning. The last common idea developed using as basis RNNs suggests the use of neighbouring opinion terms in order to refine the extraction of explicit aspects, similar to the pattern-based systems.

Numerous neural networks combine an attention weighting layer with additional layers based on RNNs, CNNs, or FFNNs. We find only a few neural networks exclusively built on attention and used only for the unsupervised detection of implicit aspects. More recently, Transformer and its NLP variant, BERT, have been introduced as attention-based systems. Even the simplest method with BERT word embeddings and a dense layer applied sequentially proved to be effective, a fact that probably will encourage future works in this direction.

Deep learning has proved to be an effective framework for the detection of both types of aspects by exploiting a wide range of approaches, including state-of-the-art methods in NLP. Given the no-reliance on annotated labels and domain flexibility, unsupervised learning is probably one of the most promising directions for future works. However, the major downside of this approach is the uncertainty of methods about the extracted terms, which might be explicit aspects, descriptors of the implicit aspects, or background words. A relevant work for this research direction is proposed by Wang et al. (2015), where an unsupervised method is guided by some prior information (provided by a pattern-based system) to extract explicit aspects. However, the solution can be improved both in terms of the architecture (e.g., using the attention mechanism) and prior information (e.g., using manual or generated patterns as a guideline). Besides unsupervised learning, end-to-end approaches used to jointly extract aspects and their sentiment polarities represent another important research direction that has not yet been fully exploited, and that proved to be effective, especially because it addresses the ABSA as a singular task instead of considering multiple separate subtasks.

5 Conclusions

In sentiment mining, aspects represent the subjects to which the opinions are expressed in text. In this survey, we define a taxonomy for the extraction of implicit and explicit aspects and present the most relevant works accordingly. The first proposed solutions for the extraction of aspects rely on syntactic and lexical patterns with embedded concepts like POS tags and dependency relations. Later on, the detection of aspects was introduced as a NER problem with highly correlated consecutive labels. The problem was first modeled by classical machine learning methods and then by deep learning. While the pattern-based methods relied more on unsupervised learning to detect aspects, the more recent methods mostly adopt a supervised approach.

When considering the future of aspect extraction, we think that domain portability is probably the most important research direction. This idea is justified by the limited availability of the annotated data that needs to be exploited not only in other domains but also over the time variations. Conventionally, the domain independence is addressed by unsupervised learning that was mainly represented by the pattern-based methods, so far. However, we consider that a plain solution that relies either on some prior knowledge or on a simple machine learning method is not enough to control the rich linguistic and syntactic environment. Therefore, a hybrid method should be developed to reinforce the performance of a machine learning approach with the prior knowledge required for the detection of either explicit or implicit aspects. Along with the unsupervised learning, another future step in the field of aspect detection is related to the end-to-end (joint) approaches proposed recently and with the strong capability to treat ABSA as a unified task.

6 Declarations

6.1 Funding

Not applicable.

6.2 Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

6.3 Availability of data and material

Not applicable.

6.4 Code availability

Not applicable.

References

Agerri R, Bermudez J, Rigau G (2014) IXA pipeline: Efficient and ready to use multilingual NLP tools. In: 9th International Conference on Language Resources and Evaluation (LREC 2014), ELRA, vol 2014, pp 3823–3828, URL http://www.lrec-conf.org/proceedings/lrec2014/summaries/775.html

Anderson JR, et al. (1983) A spreading activation theory of memory. Journal of Verbal Learning and Verbal Behavior 22(3):261–295, URL https://doi.org/10.1016/S0022-5371(83)90201-3

Angelidis S, Lapata M (2018) Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In: 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), ACL, pp 3675–3686, URL https://doi.org/10.18653/v1/D18-1403

Artieres T, et al. (2010) Neural conditional random fields. In: 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010), JMLR.org, JMLR Proceedings, vol 9, pp 177–184

Baker CF, Sato H (2003) The framenet data and software. In: 41st Annual Meeting on Association for Computational Linguistics (ACL 2003), ACL, pp 161–164, URL https://doi.org/10.3115/1075178.1075206

Berger AL, Pietra VJD, Pietra SAD (1996) A maximum entropy approach to natural language processing. Computational Linguistics 22(1):39–71

Blair-Goldensohn S, Hannan K, McDonald R, Neylon T, Reis G, Reynar J (2008) Building a sentiment summarizer for local service reviews. In: Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2008)

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of Machine Learning Research 3:993–1022, URL http://jmlr.org/papers/v3/blei03a.html

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: 5th Annual Conference on Computational Learning Theory (COLT 1992), ACM, pp 144–152, URL https://doi.org/10.1145/130385.130401

Brody S, Elhadad N (2010) An unsupervised aspect-sentiment model for online reviews. In: 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2010), ACL, pp 804–812, URL https://aclanthology.org/N10-1122

Cambria E, Fu J, Bisio F, Poria S (2015) AffectiveSpace 2: Enabling affective intuition for concept-level sentiment analysis. In: 29th Conference on Artificial Intelligence (AAAI 2015), AAAI Press, pp 508–514, URL http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9911

Cambria E, Poria S, Bajpai R, Schuller BW (2016) SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: 26th International Conference on Computational Linguistics (COLING 2016), ACL, pp 2666–2677, URL https://aclanthology.org/C16-1251/

Cambria E, Li Y, Xing FZ, Poria S, Kwok K (2020) SenticNet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: 29th ACM International Conference on Information & Knowledge Management (CIKM 2020), ACM, pp 105–114, URL https://doi.org/10.1145/3340531.3412003

Chauhan GS, Meena YK, Gopalani D, Nahta R (2020) A two-step hybrid unsupervised model with attention mechanism for aspect extraction. Expert Systems with Applications 161:113673, URL https://doi.org/10.1016/j.eswa.2020.113673

Chen G, Tian Y, Song Y (2020) Joint aspect extraction and sentiment analysis with directional graph convolutional networks. In: 28th International Conference on Computational Linguistics (COLING 2020), ICCL, pp 272–279, URL https://doi.org/10.18653/v1/2020.coling-main.24

Chen T, Xu R, He Y, Wang X (2017) Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications 72:221–230, URL https://doi.org/10.1016/j.eswa.2016.10.065

Chen Z, Liu B (2014) Topic modeling using topics from many domains, lifelong learning and big data. In: 31th International conference on machine learning (ICML 2014), JMLR.org, pp 703–711, URL http://proceedings.mlr.press/v32/chenf14.html

Chen Z, Liu B (2018) Lifelong machine learning. Synthesis Lectures on Artificial Intelligence and Machine Learning 12(3):1–207, URL https://doi.org/10.2200/S00737ED1V01Y201610AIM033

Chernyshevich M (2014) IHS R&D Belarus: Cross-domain extraction of product features using conditional random fields. In: 8th International Workshop on Semantic Evaluation (SemEval@COLING 2014), ACL, pp 309–313, URL https://doi.org/10.3115/v1/S14-2051

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: 2014 Conference on Empirical Methods in Natural (EMNLP 2014), ACL, pp 1724–1734, URL https://doi.org/10.3115/v1/D14-1179

Cilibrasi RL, Vitanyi PM (2007) The Google similarity distance. IEEE Transactions on Knowledge and Data Engineering 19(3):370–383, URL https://doi.org/10.1109/TKDE.2007.48

Collins M (2002) Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In: 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), ACL, pp 1–8

Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y (2006) Online passive-aggressive algorithms. Journal of Machine Learning Research 7(Mar):551–585, URL http://jmlr.org/papers/v7/crammer06a.html

Da'u A, Salim N, Rabiu I, Osman A (2020) Weighted aspect-based opinion mining using deep learning for recommender system. Expert Systems with Applications 140:112871, URL https://doi.org/10.1016/j.eswa.2019.112871

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39(1):1–22, URL https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. In: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), ACL, pp 4171–4186, URL https://doi.org/10.18653/v1/N19-1423

Do HH, Prasad P, Maag A, Alsadoon A (2019) Deep learning for aspect-based sentiment analysis: A comparative review. Expert Systems with Applications 118:272–299, URL https://doi.org/10.1016/j.eswa.2018.10.003

Dosoula N, Griep R, Den Ridder R, Slangen R, Schouten K, Frasincar F (2016) Detection of multiple implicit features per sentence in consumer review data. In: in 12th International Baltic Conference on Databases and Information Systems (DB&IS 2016), Springer, pp 289–303, URL https://doi.org/10.1007/978-3-319-40180-5_20

Dragoni M, Federici M, Rexha A (2019) An unsupervised aspect extraction strategy for monitoring real-time reviews stream. Information Processing & Management 56(3):1103–1118, URL https://doi.org/10.1016/j.ipm.2018.04.010

Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. Computational Linguistics 19(1):61–74

Elman JL (1990) Finding structure in time. Cognitive Science 14(2):179–211, URL https://doi.org/10.1207/s15516709cog1402_1

Etter D, Ferraro F, Cotterell R, Buzek O, Van Durme B (2013) NERIT: Named entity recognition for informal text. Human Language Technology Center of Excellence, Johns Hopkins Technical Report 11

Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V (2016) Domain-adversarial training of neural networks. The Journal of Machine Learning Research 17(1):2096–2030

Ganu G, Elhadad N, Marian A (2009) Beyond the stars: Improving rating predictions using review text content. In: 12th International Workshop on the Web and Databases (WebDB 2009), vol 9, pp 1–6, URL http://webdb09.cse.buffalo.edu/papers/Paper9/WebDB.pdf

Gong C, Yu J, Xia R (2020) Unified feature and instance based domain adaptation for end-to-end aspect-based sentiment analysis. In: 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), ACL, pp 7035–7045, URL https://doi.org/10.18653/v1/2020.emnlp-main.572

Hai Z, Chang K, Kim JJ, Yang CC (2013) Identifying features in opinion mining via intrinsic and extrinsic domain relevance. IEEE Transactions on Knowledge and Data Engineering 26(3):623–634, URL https://doi.org/10.1109/TKDE.2013.26

Hamdan H, Bellot P, Bechet F (2015) LSISLIF: CRF and logistic regression for opinion target extraction and sentiment polarity analysis. In: 9th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2015), ACL, pp 753–758, URL https://doi.org/10.18653/v1/S15-2128

He R, Lee WS, Ng HT, Dahlmeier D (2017) An unsupervised neural attention model for aspect extraction. In: 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), ACL, pp 388–397, URL https://doi.org/10.18653/v1/P17-1036

He R, Lee WS, Ng HT, Dahlmeier D (2019) An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: 57th Conference of the Association for Computational Linguistics (ACL 2019), ACL, pp 504–515, URL https://doi.org/10.18653/v1/p19-1048

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780, URL https://doi.org/10.1162/neco.1997.9.8.1735

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp 168–177, URL https://doi.org/10.1145/1014052.1014073

Hu M, Peng Y, Huang Z, Li D, Lv Y (2019) Open-domain targeted sentiment analysis via span-based extraction and classification. In: 57th Conference of the Association for Computational Linguistics (ACL 2019), ACL, URL https://doi.org/10.18653/v1/p19-1051

Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011), ACL, vol 1, pp 151–160, URL https://aclanthology.org/P11-1016/

Jin W, Ho HH, Srihari RK (2009) OpinionMiner: A novel machine learning system for web opinion mining and extraction. In: 15th International Conference on Knowledge Discovery and Data Mining (SIGKDD 2009), ACM, pp 1195–1204, URL https://doi.org/10.1145/1557019.1557148

Jindal N, Liu B (2008) Opinion spam and analysis. In: 1st International Conference on Web Search and Data Mining (WSDM 2008), ACL, pp 219–230, URL https://doi.org/10.1145/1341531.1341560

Jordan MI (1997) Serial order: A parallel distributed processing approach. In: Advances in Psychology, vol 121, Elsevier, pp 471–495, URL https://doi.org/10.1016/S0166-4115(97)80111-2

Kang Y, Zhou L (2017) RubE: Rule-based methods for extracting product features from online consumer reviews. Information & Management 54(2):166–176, URL https://doi.org/10.1016/j.im.2016.05.007

Karimi A, Rossi L, Prati A (2021) Adversarial training for aspect-based sentiment analysis with BERT. In: 25th International Conference on Pattern Recognition (ICPR 2020), pp 8797–8803, URL https://doi.org/10.1109/ICPR48806.2021.9412167

Kim SM, Hovy E (2006) Extracting opinions, opinion holders, and topics expressed in online news media text. In: Workshop on Sentiment and Subjectivity in Text 2006 (SST 2006, ACL, pp 1–8, URL https://aclanthology.org/W06-0301

Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014) NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In: 8th International Workshop on Semantic Evaluation (SemEval@COLING 2014), ACL, pp 437–442, URL https://doi.org/10.3115/v1/S14-2076

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680, URL https://doi.org/10.1126/science.220.4598.671

Lafferty J, McCallum A, Pereira FC (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: 18th International Conference on Machine Learning (ICML 2001), Morgan Kaufmann, pp 282–289

LeCun Y, Haffner P, Bottou L, Bengio Y (1999) Object recognition with gradient-based learning. In: Shape, Contour and Grouping in Computer Vision, Springer, LNCS, vol 1681, pp 319–345, URL https://doi.org/10.1007/3-540-46805-6_19

Lee K, Levy O, Zettlemoyer L (2017) Recurrent additive networks. arXiv preprint arXiv:170507393 URL http://arxiv.org/abs/1705.07393

Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2020) BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), ACL, pp 7871–7880, URL https://doi.org/10.18653/v1/2020.acl-main.703

Li F, Han C, Huang M, Zhu X, Xia YJ, Zhang S, Yu H (2010) Structure-aware review mining and summarization. In: 23rd International Conference on Computational Linguistics (COLING 2010), ACL, pp 653–661, URL https://aclanthology.org/C10-1074/

Li K, Chen C, Quan X, Ling Q, Song Y (2020) Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. In: Jurafsky D, Chai J, Schluter N, Tetreault JR (eds) 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), ACL, pp 7056–7066, URL https://doi.org/10.18653/v1/2020.acl-main.631

Li X, Lam W (2017) Deep multi-task learning for aspect term extraction with memory interaction. In: 2017 Conference on Empirical Methods In Natural Language Processing (EMNLP 2017), ACL, pp 2886–2892, URL https://doi.org/10.18653/v1/d17-1310

Li X, Bing L, Li P, Lam W, Yang Z (2018) Aspect term extraction with history attention and selective transformation. In: 27th International Joint Conference on Artificial Intelligence (IJCAI 2018), ijcai.org, pp 4194–4200, URL https://doi.org/10.24963/ijcai.2018/583

Li X, Bing L, Li P, Lam W (2019a) A unified model for opinion target extraction and target sentiment prediction. In: 33rd Conference on Artificial Intelligence (AAAI 2019), AAAI Press, vol 33, pp 6714–6721, URL https://doi.org/10.1609/aaai.v33i01.33016714

Li X, Bing L, Zhang W, Lam W (2019b) Exploiting BERT for end-to-end aspect-based sentiment analysis. In: 5th Workshop on Noisy User-generated Text (W-NUT@EMNLP 2019), ACL, pp 34–41, URL https://doi.org/10.18653/v1/D19-5505

Li Z, Li X, Wei Y, Bing L, Zhang Y, Yang Q (2019c) Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In: 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019) and the 9th International Joint Conference on Natural Language Processing (IJCNLP 2019), ACL, pp 4589–4599, URL https://doi.org/10.18653/v1/D19-1466

Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge-Based Systems 235:107643, URL https://doi.org/10.1016/j.knosys.2021.107643

Liu B (2007) Web data mining: Exploring hyperlinks, contents, and usage data. DCSA, Springer, URL https://doi.org/10.1007/978-3-642-19460-3

Liu B (2010) Sentiment analysis and subjectivity. In: Indurkhya N, Damerau FJ (eds) Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC, pp 627–666, URL http://www.crcnetbase.com/doi/abs/10.1201/9781420085938-c26

Liu B (2011) Web Data Mining: Exploring hyperlinks, contents, and usage data. Data-Centric Systems and Applications, Springer, URL https://doi.org/10.1007/978-3-642-19460-3

Liu B (2015) Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, URL https://doi.org/10.1017/CBO9781139084789

Liu B, Hsu W, Ma Y, et al. (1998) Integrating classification and association rule mining. In: 4th International Conference on Knowledge Discovery and Data Mining (KDD 1998), AAAI Press, vol 98, pp 80–86, URL http://www.aaai.org/Library/KDD/1998/kdd98-012.php

Liu B, Hu M, Cheng J (2005) Opinion Observer: Analyzing and comparing opinions on the web. In: 14th International Conference on World Wide Web (WWW 2005), ACM, pp 342–351, URL https://doi.org/10.1145/1060745.1060797

Liu P, Joty S, Meng H (2015) Fine-grained opinion mining with recurrent neural networks and word embeddings. In: 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), ACL, pp 1433–1443, URL https://doi.org/10.18653/v1/D15-1168

Liu Q, Gao Z, Liu B, Zhang Y (2016) Automated rule selection for opinion target extraction. Knowledge-Based Systems 104:74–88, URL https://doi.org/10.1016/j.knosys.2016.04.010

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:190711692 URL http://arxiv.org/abs/1907.11692

Luo H, Li T, Liu B, Zhang J (2019a) DOER: Dual cross-shared RNN for aspect term-polarity co-extraction. In: 57th Conference of the Association for Computational Linguistics (ACL 2019), ACL, pp 591–601, URL https://doi.org/10.18653/v1/P19-1056

Luo L, Ao X, Song Y, Li J, Yang X, He Q, Yu D (2019b) Unsupervised neural aspect extraction with sememes. In: 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), AAAI Press, pp 5123–5129, URL https://doi.org/10.24963/ijcai.2019/712

Ma D, Li S, Wu F, Xie X, Wang H (2019) Exploring sequence-to-sequence learning in aspect term extraction. In: 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), ACL, pp 3538–3547, URL https://doi.org/10.18653/v1/P19-1344

Ma Y, Peng H, Cambria E (2018a) Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: 32th Conference on Artificial Intelligence (AAAI 2018), AAAI Press, pp 5876–5883, URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16541

Ma Y, Peng H, Khan T, Cambria E, Hussain A (2018b) Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis. Cognitive Computation 10(4):639–650, URL https://doi.org/10.1007/s12559-018-9549-x

Maitama JZ, Idris N, Abdi A, Shuib L, Fauzi R (2020a) A systematic review on implicit and explicit aspect extraction in sentiment analysis. IEEE Access 8:194166–194191, URL https://doi.org/10.1109/ACCESS.2020.3031217

Maitama JZ, Idris N, Zakari A (2020b) A systematic mapping study of the empirical explicit aspect extractions in sentiment analysis. IEEE Access 8:113878–113899, URL https://doi.org/10.1109/ACCESS.2020.3003625

Majumder N, Bhardwaj R, Poria S, Gelbukh A, Hussain A (2020) Improving aspect-level sentiment analysis with aspect extraction. Neural Computing and Applications URL https://arxiv.org/abs/2005.06607

Manek AS, Shenoy PD, Mohan MC, Venugopal K (2017) Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier. World Wide Web 20(2):135–154, URL https://doi.org/10.1007/s11280-015-0381-x

Manevitz LM, Yousef M (2001) One-class SVMs for document classification. Journal of Machine Learning Research 2:139–154, URL http://jmlr.org/papers/v2/manevitz01a.html

Mangnoesing GV, Trusca MM, Frasincar F (2020) Pattern Learning for Detecting Defect Reports and Improvement Requests in App Reviews. In: 25th International Conference on Natural Language Information Systems (NLDB 2020), Springer, pp 126–134, URL https://doi.org/10.1007/978-3-030-51310-8_12

Mao Y, Shen Y, Yu C, Cai L (2021) A joint training dual-mrc framework for aspect based sentiment analysis. In: 35th Conference on Artificial Intelligence (AAAI 2021) 33rd Conference on Innovative Applications of Artificial Intelligence (IAAI 2021), 11th Symposium on Educational Advances in Artificial Intelligence (EAAI 2021), AAAI Press, pp 13543–13551, URL https://ojs.aaai.org/index.php/AAAI/article/view/17597

McAuley J, Yang A (2016) Addressing complex and subjective product-related queries with customer reviews. In: 25th International Conference on World Wide Web (WWW 2016), ACM, pp 625–635, URL https://doi.org/10.1145/2872427.2883044

McAuley J, Leskovec J, Jurafsky D (2012) Learning attitudes and attributes from multi-aspect reviews. In: 12th IEEE International Conference on Data Mining (ICDM 2012), IEEE, pp 1020–1025, URL https://doi.org/10.1109/ICDM.2012.110

McAuley J, Targett C, Shi Q, Van Den Hengel A (2015) Image-based recommendations on styles and substitutes. In: 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015), ACM, pp 43–52, URL https://doi.org/10.1145/2766462.2767755

Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal 5(4):1093–1113, URL https://doi.org/10.1016/j.asej.2014.04.011

Mei Q, Ling X, Wondra M, Su H, Zhai C (2007) Topic sentiment mixture: Modeling facets and opinions in weblogs. In: 16th international conference on World Wide Web (WWW 2007), ACM, pp 171–180, URL https://doi.org/10.1145/1242572.1242596

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: 27th Annual Conference in Neural Information Processing Systems (NIPS 2013), pp 3111–3119, URL https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Miller GA (1995) WordNet: a lexical database for English. Communications of the ACM 38(11):39–41, URL https://doi.org/10.1145/219717.219748

Mirjalili S, Lewis A (2016) The whale optimization algorithm. Advances in engineering software 95:51–67, URL https://doi.org/10.1016/j.advengsoft.2016.01.008

Mitchell M, Aguilar J, Wilson T, Van Durme B (2013) Open domain targeted sentiment. In: 2013 Conference on Empirical Methods in Natural Language Pro-

cessing (EMNLP 2013), ACL, pp 1643–1654, URL https://doi.org/10.18653/v1/P19-1051
Moghaddam S, Ester M (2010) Opinion Digger: An unsupervised opinion miner from unstructured product reviews. In: 19th International Conference on Information and Knowledge Management (CIKM 2010), ACM, pp 1825–1828, URL https://doi.org/10.1145/1871437.1871739
Mukherjee A, Liu B (2012) Aspect extraction through semi-supervised modeling. In: 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), ACL, pp 339–348, URL https://aclanthology.org/P12-1036
Niu ZY, Ji DH, Tan CL (2007) I2r: Three systems for word sense discrimination, chinese word sense disambiguation, and english word sense disambiguation. In: 4th International Workshop on Semantic Evaluations (SemEval@ACL 2007), ACL, pp 177–182, URL https://aclanthology.org/S07-1037
Pantel P, Lin D (2002) Discovering word senses from text. In: 8thh International Conference on Knowledge Discovery and Data Mining (SIGKDD 2002), ACM, pp 613–619, URL https://doi.org/10.1145/775047.775138
Patel M, Ezeife CI (2021) BERT-based multi-task learning for aspect-based opinion mining. In: 32nd International Conference on Database and Expert Systems Applications (DEXA 2021), Springer, pp 192–204, URL https://doi.org/10.1007/978-3-030-86472-9_18
Peng H, Xu L, Bing L, Huang F, Lu W, Si L (2020) Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: 34th AAAI Conference on Artificial Intelligence, (AAAI 2020), 32nd Innovative Applications of Artificial Intelligence Conference (IAAI 2020), 10th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2020), AAAI Press, pp 8600–8607, URL https://doi.org/10.1609/aaai.v34i05.6383
Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. In: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), ACL, pp 1532–1543, URL https://doi.org/10.3115/v1/d14-1162
Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologie (NAACL-HLT 2018), ACL, pp 2227–2237, URL https://doi.org/10.18653/v1/n18-1202
Phan MH, Ogunbona PO (2020) Modelling context and syntactical features for aspect-based sentiment analysis. In: 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), ACL, pp 3211–3220, URL https://doi.org/10.18653/v1/2020.acl-main.293
Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S (2014) SemEval-2014 task 4: Aspect based sentiment analysis. In: 8th International Workshop on Semantic Evaluation (SemEval@COLING 2014), ACL, Dublin, Ireland, pp 27–35, URL https://doi.org/10.3115/v1/S14-2004
Pontiki M, Galanis D, Papageorgiou H, Manandhar S, Androutsopoulos I (2015) Semeval-2015 task 12: Aspect based sentiment analysis. In: 9th international Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2015), ACL, pp 486–495, URL https://doi.org/10.18653/v1/S15-2082
Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Al-Smadi M, Al-Ayyoub M, Zhao Y, Qin B, De Clercq O, et al. (2016) Semeval-

2016 task 5: Aspect based sentiment analysis. In: 10th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2016), ACL, pp 19–30

Popescu AM, Etzioni O (2007) Extracting product features and opinions from reviews. In: Natural language processing and text mining, Springer, pp 9–28, URL https://doi.org/10.1007/978-1-84628-754-1_2

Poria S, Cambria E, Ku LW, Gui C, Gelbukh A (2014) A rule-based approach to aspect extraction from product reviews. In: 2nd Workshop on Natural Language Processing for Social Media (SocialNLP 2014), ACL, pp 28–37, URL https://doi.org/10.3115/v1/W14-5905

Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems 108:42–49, URL https://doi.org/10.1016/j.knosys.2016.06.009

Qiu G, Liu B, Bu J, Chen C (2009) Expanding domain sentiment lexicon through double propagation. In: 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), Morgan Kaufmann Publishers Inc., URL http://ijcai.org/Proceedings/09/Papers/202.pdf

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2):257–286, URL https://doi.org/10.1109/5.18626

Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. OpenAI

Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) SQUAD: 100,000+ questions for machine comprehension of text. In: 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), ACL, pp 2383–2392, URL https://doi.org/10.18653/v1/D16-1264

Rana T, Cheah YN (2019) Sequential patterns rule-based approach for opinion target extraction from customer reviews. Journal of Information Science 45(5):643–655, URL https://doi.org/10.1177/0165551518808195

Rana TA, Cheah YN (2016) Aspect extraction in sentiment analysis: Comparative analysis and survey. Artificial Intelligence Review 46(4):459–483, URL https://doi.org/10.1007/s10462-016-9472-z

Rana TA, Cheah YN (2017) A two-fold rule-based model for aspect extraction. Expert Systems with Applications 89:273–285, URL https://doi.org/10.1016/j.eswa.2017.07.047

Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems 89:14–46, URL https://doi.org/10.1016/j.knosys.2015.06.015

Reimers N, Gurevych I (2017) Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging. In: 2017 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2017), ACL, pp 338–348, URL https://doi.org/10.18653/v1/D17-1035

Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: 24th international Conference on Machine Learning (ICML 2007), ACM, vol 227, pp 791–798, URL https://doi.org/10.1145/1273496.1273596

San Vicente I, Saralegi X, Agerri R (2015) EliXa: A modular and flexible ABSA platform. In: 9th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2015), ACL, pp 748–752, URL https://doi.org/10.18653/v1/S15-2127

Schouten K, Frasincar F (2015) Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering 28(3):813–830, URL https://doi.org/10.1109/TKDE.2015.2485209

Schouten K, Van Der Weijde O, Frasincar F, Dekker R (2017) Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE Transactions on Cybernetics 48(4):1263–1275, URL https://doi.org/10.1109/TCYB.2017.2688801

Shu L, Xu H, Liu B (2017) Lifelong learning CRF for supervised aspect extraction. In: 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), ACL, URL https://doi.org/10.18653/v1/P17-2023

Somasundaran S, Wiebe J (2009) Recognizing stances in online debates. In: 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL 2009), ACL, pp 226–234, URL https://aclanthology.org/P09-1026

Srikant R, Agrawal R (1996) Mining sequential patterns: Generalizations and performance improvements. In: 5th International Conference on Extending Database Technology (EDBT 1996), Springer, LNCS, pp 1–17, URL https://doi.org/10.1007/BFb0014140

Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. Expert Systems with Applications 36(7):10760–10773, URL https://doi.org/10.1016/j.eswa.2009.02.063

Titov I, McDonald R (2008) Modeling online reviews with multi-grain topic models. In: 17th international conference on World Wide Web (WWW 2008), ACM, pp 111–120, URL https://doi.org/10.1145/1367497.1367513

Toh Z, Su J (2016) NLANGP at SemeVal-2016 task 5: Improving aspect based sentiment analysis using neural network features. In: 10th international Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2016), ACL, pp 282–288, URL https://doi.org/10.18653/v1/S16-1045

Toh Z, Wang W (2014) DLIREC: Aspect term extraction and term polarity classification system. In: 8th International Workshop on Semantic Evaluation (SemEval@COLING 2014), ACL, pp 235–240, URL https://doi.org/10.3115/v1/s14-2038

Toprak C, Jakob N, Gurevych I (2010) Sentence and expression level annotation of opinions in user-generated discourse. In: 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), ACL, pp 575–584, URL https://aclanthology.org/P10-1059

Tubishat M, Idris N, Abushariah M (2021) Explicit aspects extraction in sentiment analysis using optimal rules combination. Future Generation Computer Systems 114:448–480, URL https://doi.org/10.1016/j.future.2020.08.019

Tulkens S, van Cranenburgh A (2020) Embarrassingly simple unsupervised aspect extraction. In: 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), ACL, pp 3182–3187, URL https://doi.org/10.18653/v1/2020.acl-main.290

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: 31st Annual Conference on Neural Information Processing Systems (NIPS 2017), pp 5998–6008, URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Wang L, Liu K, Cao Z, Zhao J, De Melo G (2015) Sentiment-aspect extraction based on restricted Boltzmann machines. In: 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015) and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL 2015), ACL, pp 616–625, URL https://doi.org/10.3115/v1/P15-1060

Wang W, Pan SJ, Dahlmeier D, Xiao X (2016) Recursive neural conditional random fields for aspect-based sentiment analysis. In: 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), ACL, pp 616–626, URL https://doi.org/10.18653/v1/D16-1059

Wang W, Pan SJ, Dahlmeier D, Xiao X (2017) Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: 31st Conference on Artificial Intelligence (AAAI 2017), AAAI Press, pp 3316–3322, URL http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14441

Wei Z, Hong Y, Zou B, Cheng M, Jianmin Y (2020) Don't eclipse your arts due to small discrepancies: Boundary repositioning with a pointer network for aspect extraction. In: 58th Annual Meeting of the Association for Computational Linguistics, ACL, pp 3678–3684, URL https://doi.org/10.18653/v1/2020.acl-main.339

Weichselbraun A, Gindl S, Fischer F, Vakulenko S, Scharl A (2017) Aspect-based extraction and analysis of affective knowledge from social media streams. IEEE Intelligent Systems 32(3):80–88, URL https://doi.org/10.1109/MIS.2017.57

Wu C, Wu F, Wu S, Yuan Z, Huang Y (2018) A hybrid unsupervised method for aspect term and opinion target extraction. Knowledge-Based Systems 148:66–73, URL https://doi.org/10.1016/j.knosys.2018.01.019

Wu Y, Zhang Q, Huang X, Wu L (2009) Phrase dependency parsing for opinion mining. In: 2009 Conference on Empirical Methods in Natural Language Processing, ACL, pp 1533–1541

Xenos D, Theodorakakos P, Pavlopoulos J, Malakasiotis P, Androutsopoulos I (2016) AUEB-ABSA at SemEval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In: 10th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT 2016), ACL, pp 312–317, URL https://doi.org/10.18653/v1/S16-1050

Xu H, Liu B, Shu L, Yu PS (2018) Double embeddings and CNN-based sequence labeling for aspect extraction. In: 56th Annual Meeting of the Association for Computational (ACL 2018), ACL, pp 592–598, URL https://doi.org/10.18653/v1/P18-2094

Xu H, Liu B, Shu L, Yu PS (2019) BERT post-training for review reading comprehension and aspect-based sentiment analysis. In: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), ACL, pp 2324–2335, URL https://doi.org/10.18653/v1/N19-1242

Xu H, Liu B, Shu L, Yu PS (2020) DomBERT: Domain-oriented language model for aspect-based sentiment analysis. In: 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), ACL, pp 1725–1731, URL https://doi.org/10.18653/v1/2020.findings-emnlp.156

Yan H, Dai J, Ji T, Qiu X, Zhang Z (2021) A unified generative framework for aspect-based sentiment analysis. In: 59th Annual Meeting of the Association for Computational Linguistics and 11th International Joint Conference on Natural

Language Processing (ACL/IJCNLP 2021), ACL, pp 2416–2429, URL https://doi.org/10.18653/v1/2021.acl-long.188

Yi J, Nasukawa T, Bunescu R, Niblack W (2003) Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: 3rd IEEE International Conference on Data Mining (ICDM 2003), IEEE, pp 427–434, URL https://doi.org/10.1109/ICDM.2003.1250949

Yin Y, Wei F, Dong L, Xu K, Zhang M, Zhou M (2016) Unsupervised word and dependency path embeddings for aspect term extraction. In: 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), AAAI Press, pp 2979–2985, URL http://www.ijcai.org/Abstract/16/423

Yu J, Zha ZJ, Wang M, Chua TS (2011) Aspect ranking: Identifying important product aspects from online consumer reviews. In: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011), ACL, pp 1496–1505, URL https://aclanthology.org/P11-1150

Zhai C, Lafferty J (2001) Model-based feedback in the language modeling approach to information retrieval. In: 10th International Conference on Information and knowledge Management (CIKM 2001), ACM, pp 403–410, URL https://doi.org/10.1145/502585.502654

Zhai Z, Liu B, Xu H, Jia P (2010) Grouping product features using semi-supervised learning with soft-constraints. In: 23rd international Conference on Computational Linguistics (COLING 2010), ACL, pp 1272–1280, URL https://aclanthology.org/C10-1143

Zhang L, Liu B, Lim SH, O'Brien-Strain E (2010) Extracting and ranking product features in opinion documents. In: 23rd International Conference on Computational Linguistics (COLING 2010), ACL, pp 1462–1470, URL https://aclanthology.org/C10-2167/

Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4), URL https://doi.org/10.1002/widm.1253

Zhang M, Zhang Y, Vo DT (2015) Neural networks for open domain targeted sentiment. In: 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), ACL, pp 612–621, URL https://doi.org/10.18653/v1/d15-1073

Zhang Z (2013) urCF: An approach to integrating user reviews into memory-based collaborative filtering. PhD thesis, University of Maryland

Zhao H, Huang L, Zhang R, Lu Q, Xue H (2020) SpanMlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In: 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), ACL, pp 3239–3248, URL https://doi.org/10.18653/v1/2020.acl-main.296