Accepted Manuscript

Distinguishing Between Facts and Opinions for Sentiment Analysis: Survey and Challenges

Iti Chaturvedi, Erik Cambria, Roy E. Welsch, Francisco Herrera

PII: S1566-2535(17)30390-1
DOI: 10.1016/j.inffus.2017.12.006
Reference: INFFUS 939

To appear in: *Information Fusion*

Received date: 22 June 2017
Revised date: 19 December 2017
Accepted date: 21 December 2017

Please cite this article as: Iti Chaturvedi, Erik Cambria, Roy E. Welsch, Francisco Herrera, Distinguishing Between Facts and Opinions for Sentiment Analysis: Survey and Challenges, *Information Fusion* (2017), doi: 10.1016/j.inffus.2017.12.006

Highlights

- The paper provides a first-of-its-kind survey on subjectivity detection
- The survey contains both traditional methods and recent techniques
- The paper proposes a discussion of current limitations and future directions

Distinguishing Between Facts and Opinions for Sentiment Analysis: Survey and Challenges

Iti Chaturvedi[a], Erik Cambria[a], Roy E. Welsch[b], Francisco Herrera[c]

[a]*School of Computer Engineering, Nanyang Technological University, Singapore.*
[b]*Sloan School of Management, MIT, USA*
[c]*Soft Computing and Intelligent Information Systems, University of Granada*

Abstract

Sentiment analysis requires a lot of information coming from different sources and about different topics to be retrieved and fused. For this reason, one of the most important subtasks of sentiment analysis is subjectivity detection, i.e., the removal of 'factual' or 'neutral' comments that lack sentiment. It is possibly the most essential subtask of sentiment analysis as sentiment classifiers are often optimized to categorize text as either negative or positive and, hence, forcefully fit unopinionated sentences into one of these two categories. This article reviews hand-crafted and automatic models for subjectivity detection in the literature. It highlights the key assumptions these models make, the results they obtain, and the issues that still need to be explored to further our understanding of subjective sentences. Lastly, the advantages and limitations of each approach are compared. The methods can be broadly categorized as hand-crafted, automatic, and multi-modal. Hand-crafted templates work well on strong sentiments, however they are unable to identify weakly subjective sentences. Automatic methods such as deep learning provide a meta-level feature representation that generalizes well on new domains and languages. Multi-modal methods can combine the abundant audio and video forms of social data with text using multiple kernels. We conclude that the high-dimensionality of n-gram features and temporal nature of sentiments in long product reviews are

Email addresses: iti@ntu.edu.sg (Iti Chaturvedi), cambria@ntu.edu.sg (Erik Cambria), rwelsch@mit.edu (Roy E. Welsch), herrera@decsai.ugr.es (Francisco Herrera)

the major challenges in sentiment mining from text.

Keywords: Subjectivity Detection, Convolutional Neural Network, Word Vector Model, Twitter, Multi-lingual Analysis

---

Abbreviations

$$\begin{aligned}
NLP &= \text{Natural Language Processing} \\
SVM &= \text{Support Vector Machines} \\
NB &= \text{Naïve Bayes} \\
MKL &= \text{Multiple Kernel Learning} \\
CRF &= \text{Conditional Random Fields} \\
LSTM &= \text{Long Short-Term Memory} \\
MPQA &= \text{Multi Party Question Answering} \\
BOW &= \text{Bag of Words} \\
CNN &= \text{Convolutional Neural Network} \\
WSD &= \text{Word Sense Disambiguation} \\
LDA &= \text{Latent Dirichlet Allocation} \\
SSWE &= \text{Sentiment-Specific Word Embedding}
\end{aligned} \tag{1}$$

1. Introduction

In recent years, sentiment analysis [1] has become increasingly popular for processing social media data on online communities, blogs, wikis, microblogging platforms, and other online collaborative media. Sentiment analysis is a branch of affective computing research [2] that aims to classify text (but sometimes also audio and video [3]) into either positive or negative. It is a field related to information retrieval and information fusion as it requires data to be collected, integrated, and classified. Most of the literature is on English language but recently an increasing number of publications is tackling the multilinguality issue [4]. Sentiment analysis systems can be broadly categorized into knowledge-based [5] and statistics-based [6]. While most works approach it as a simple categorization problem, sentiment analysis is actually a suitcase research problem [7] that requires tackling many NLP tasks, including named entity recognition [8], concept extraction [9], sarcasm detection [10], aspect extraction [11], and subjectivity detection. Subjectivity detection, in particular, is an essential subtask of sentiment analysis because most polarity detection tools are optimized for distinguishing between positive and negative text. Subjectivity detection, hence, ensures that factual information is filtered out and only opinionated information is passed on to the polarity classifier [12]. Furthermore, subjective extracts are only 60% of the review and produce the same polarity results are full text classification [13].

For example, product reviews on dedicated sites like 'Rotten-tomatoes' or 'Amazon' become difficult to interpret due to presence of several neutral reviews. Another application of subjectivity detection is in determining response of people to different crisis events from text in Twitter and Facebook [14]. This is very useful to analysts in government and political domains [15, 13]. In finance, news has a lot of impact on the psychology of an investor. However, in financial markets, factual information may imply positive or negative sentiment that is not detected by coarse grained methods that only focus on detecting the explicit sentiment [16].

Labeling a sentence as subjective or objective is a challenging task for human annotators [17]. For example, we consider the sentence "Those digging graves for others, get engraved themselves, he [Abdullah] said while citing the example of Afghanistan." Here, there is clearly an objective frame for the writer and a direct subjective frame for Abdullah with the text anchor 'said'. However, it is ambiguous whether the text anchor 'citing' is objective or subjective [18]. Another, bottleneck is the huge computational cost of existing n-gram methods that rely on syntactical representation of text such as part-of-speech (POS) tagging or word-sense disambiguation. For example, in [19], the authors showed that wishful subjective sentences that indicate purchasing interest often contain some modal verbs.

In addition, subjectivity detection in online forums such as Ubuntu[1] is a big data problem. This is because users have different levels of expertise on the topic of discussion and may be from diverse economic and educational backgrounds and often separated by large geographical distances. Threads in forums may be factual questions such as 'What is the resolution of the camera?'. Such forums are also prone to a lot of spam from trolls that may post off-topic messages. In [20], the authors show that subjective sentences in online forums can be identified by 'Dialog Acts' such as 'Question', 'Repeated Question', 'Clarification', etc. They also show that subjective sentences are longer than objective sentences and often contain inappropriate content such as abuses.

Because microblogs such as Twitter have character limitations, abbreviations or short artificial words are often used as hashtags. Furthermore, microblogging posts may have dual meaning due to the context of discussion. For example in [21], the authors show that in political tweets the word 'grun' - 'green' is used for the party called 'Die Grunen' - 'The Greens' but also as the color green. Nate Silver overcame this challenge by using a diffusion model that made forecasting elections a close to real-time experience. Hence, the support or opposition for a candidate could be modeled similar to a disease spreading across a social network via interactions [22]. Diffusion models assess each review by computing the influence of nearby reviews [23].

Subjectivity detection can hence prevent the sentiment classifier from considering irrelevant or potentially misleading text [24]. This is particularly useful in multi-perspective question answering summarization systems that need to summarize different opinions and perspectives and present multiple answers to the user based on opinions derived from different sources. Following our previously proposed model of overlapping NLP curves, we categorize subjectivity detection methods into syntactic and semantic models [25]. Lastly, we discuss

[1] `http://ubuntuforums.org`

models that can be used in new languages and fusion with video data such as YouTube.

Most subjectivity detection methods focus on identification of private states such as emotions and opinions. Personalized advertisement is based on automated generation of expressions with a certain polarity. Expressive conversations in turn require selection and understanding of affective words. The manual annotation of resources is a tedious and costly task. Thus, very few task-specific corpora and dictionaries exist for subjectivity and sentiment analysis. In [26], the authors provide a review of subjectivity detection methods. They conclude that even Naïve Bayes trained on simple uni-grams can lead to good results. However pre-processing is necessary such as removal of re-tweets, translation of abbreviations into original terms, deleting of links, tokenization and POS tagging. For multi-lingual tasks, on the other hand, accuracy depends on the type of machine translation and various features, algorithms and meta-classifiers that are used for polarity detection. Their review is limited to the review of state-of-the-art methods; however, they do not consider recent methods such as word vector model and multi-modal subjectivity detection using video and audio.

Furthermore, the era in which we live is referred to as the 'information age' due to the transformation of the internet into a social web. This challenges us to search and retrieve relevant data and mine them to transform them into knowledge that can be used to take decisions [27].Sentiment analysis is the classification of sentiment containing text into three categories (positive, negative or neutral). This has applications such as opinion summarization and predicting polarity of personal relationships in social network mining. In [28], the authors describe subjectivity analysis as a pre-processing step in sentiment detection. However, they do not discuss word vector based sentences models. Instead, in this paper we describe semantic sentence modeling [25] using deep neural networks and word vectors.

Figure 1 provides the outline of this review. Subjectivity detection methods are broadly discussed under three types, namely: Syntactic (words), Semantic (concepts), and Multi-modal (video and text). This paper discusses only subjectivity detection methods. Some authors have considered positive, neutral and negative sentences. Hence, we had to refer to them in the context of both sentiment and subjectivity analysis. The rest of the paper is organized as follows: Section 2 presents the historical background and the different schools of thought of subjectivity detection; Section 3 discusses past, present and future evolution of subjectivity algorithms; Section 4 describes use of subjectivity detection in traditional syntax-centered NLP methods; Section 5 illustrates subjectivity filtering for emerging semantics-based NLP approaches; Section 6 provides insights on dealing with scarce data in new languages and domains; in Section 7, we compare different baselines on benchmark subjectivity datasets; Section 8 lists open challenges of subjectivity detection; lastly, in Section 9 we provide conclusions.

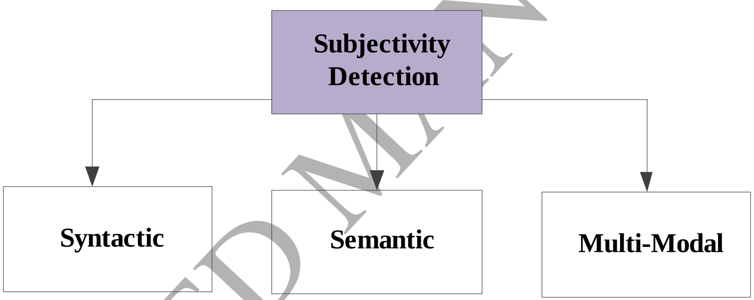

Figure 1: Organization of the review

2. Background

Sentiment analysis has raised growing interest both within the scientific community, leading to many exciting open challenges, as well as in the business world, due to the remarkable benefits to be had from financial [29] and political [30] forecasting, e-health [31] and e-tourism [32], human communication comprehension [33] and dialogue systems [34], etc.

Sentiment analysis requires a lot of information coming from different sources and about different topics to be retrieved and fused. For this reason, a key subtask of sentiment analysis is subjectivity detection. Research on subjectivity detection in the early days used well-established general subjectivity clues to generate training data from un-annotated text [35]. In the next section, we review the timeline of methods for subjectivity detection. Traditional methods looked for templates and patterns that could be easily applied to financial domains however, they did not generalize well to new domains and languages. Here, features such as pronouns, modals, adjectives, cardinal number, and adverbs showed to be effective in subjectivity classification. Several authors published lists of subjective words and several empirical methods tried to automatically identify adjectives, verbs and n-grams that are statistically associated with subjective language. However, several subjective words such as 'un-seemingly' occur infrequently, consequently a large training dataset is necessary to build a broad and comprehensive subjectivity detection system.

While there are several datasets with document and chunk labels available, there is a need to better capture sentiment from short comments, such as Twitter data, which provide less overall signal per document. Hence, in [35], the authors used extraction pattern learning to automatically generate patterns that represent subjective expressions. Subjectivity patterns can be hand-crafted or learnt automatically using software like AutoSlog [35]. For example, the pattern '$< x >$ was asked' would extract 'he was asked to leave the premises' and is strongly subjective. On the other hand, the pattern '$< x >$ was expected' would extract 'he was expected to retire' and is objective as it is a mere fact. Extracted features are used to train state-of-the-art classifiers such as support vector machines (SVM) and naïve Bayes (NB) that assume that the class of a particular feature is independent of the class of other features given the training data [36]. Sentence-level subjectivity detection was integrated into document-level sentiment detection using minimum cuts in graphs over sentences. The contextual constraints between sentences in a graph could lead to significant improvement in polarity classification [37].

On the other hand, bag-of-words (BOW) classifiers represent a document as a multi set of its words disregarding grammar and word order. They can work well on long documents by relying on a few words with strong sentiments like 'awesome'. However, distributional similarities of words such as co-occurrence matrix and context information is unable to capture differences in antonyms such as 'good/bad' since those often have similar contexts [37]. Several works have explored sentiment composition through careful engineering of features or polarity shifting rules on syntactic structures. However, sentiment accuracies for binary positive/negative classification for single sentences has not exceeded 80% for several years. When including a third 'neutral' class the accuracy falls down to only 60%.

A Bayesian network is able to represent subjective degrees of confidence. The representation explicitly explores the role of prior knowledge and combines pieces of evidence of the likelihood of events. In order to compute the joint distribution of the belief network, there is a need to know $p(x|parents(x))$ for each variable x. It is difficult to determine the probability of each variable x and difficult to build a statistical table for large-scale inference. Semantic networks, on the other hand, represent knowledge in patterns of interconnected nodes and arcs. Definitional networks focus on IsA relationships between a concept and a newly defined sub-type. The result of such a structure is called a generalization, which in turn supports the rule of inheritance for copying properties defined for a super-type to all of its sub-type. WordNet is an example of a well-known semantic network [38, 39].

Lexicons generated with automatic methods includes neutral words, introducing noise in the detection of subjectivity. The emotional load of a message is extremely important when it comes to understanding its true meaning. Microblogging platforms such as Twitter, allow real-time sharing of comments and opinions. In [40], the authors labelled a number of words according to Plutchik emotional categories and developed the NRC word-emotion association lexicon. When people are exposed to information regarding a topic or entity, they normally respond to these external stimuli by developing a personal point of view or orientation. In [41], the authors perform lexicon analysis by comparing resources created manually to lexicons that were completely automatically created. It was found that manual lexicons focused on emotional words, while the automatic methods tend to include many neutral words. A deep meta-level feature representation does not depend on the vocabulary size of the collection and, hence, provides considerable dimensionality reduction in comparison to uni-gram or n-gram models.

Recently, in [38] the authors reviewed different techniques for sentiment detection and subjectivity detection. They describe facts as objective elements and the remaining text has subjective characteristics. Sentiment analysis offers many opportunities due to the huge growth of information sources such as blogs and social networks. They conclude that semantic methods such as the use of synonyms and antonyms or relationships from thesaurus such as WordNet may also represent sentiments well. However, their review assumes the same methods for subjectivity and sentiment detection. Instead, in this review, we consider the three-class problem of positive, negative, and neutral reviews. We show that the benchmarks are markedly different from sentiment detection.

Sentence-level analysis is important because it permits a fine-grained view of different opinions expressed. Sentence-level features can be in the form of n-grams, POS tags, or location based features. In [42], the authors study the different types of sentence features and their comparative effectiveness in determining sentiment analysis. They conclude that features such as length or rhetorical segments in text have little effect in accuracy and could sometimes decrease sentiment accuracy. Similarly, bi-grams performed well without the need for POS features.

In the rest of the paper, we review the evolution of algorithms to detect neutral sentences in product reviews and micro-blogs. We discuss the advantages and disadvantages of different methods. Finally, we report the accuracies on benchmark subjectivity datasets.

3. Timeline of Subjectivity Detection

NLP research aims to produce artificially intelligent behavior in text-related tasks. However, most of the existing methods are unable to understand the context of words used. In order to properly extract and manipulate text meaning, a NLP system must have access to a significant amount of knowledge about the world and the domain of discourse. Traditional methods before 2005 used hand-crafted features such as templates and patterns to identify subjective sentences. However, such features did not generalize to new domains and languages. Figure 2 shows the timeline of subjectivity detection methods.

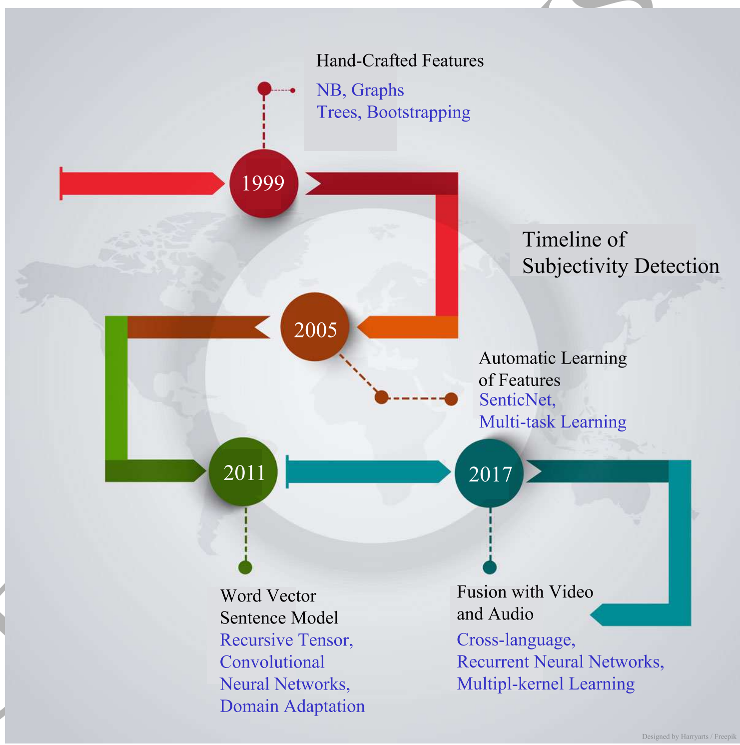

Figure 2: Timeline of subjectivity detection

Relying on arbitrary keywords, punctuation, and word co-occurrence frequencies has worked fairly well so far, but the explosion of user generated content and the outbreak of deceptive phenomena such as web trolling and opinion spam, are causing standard NLP algorithms to be increasing less efficient. In order to properly extract and manipulate text meanings, an NLP system must have access to a significant amount of knowledge about the world and the domain of discourse. NLP research has been interspersed with word-level approaches because, at first glance, the most basic unit of linguistic structure appears to be the word. Single-word expressions, however, are just a subset of concepts, multi-word expressions that carry specific semantics and sentics, that is, the denotative and connotative information commonly associated with real world objects, actions, events, and people.

Word semantic orientation (WSO) can be efficiently done using automatic statistical methods since these have a broad coverage and are able to satisfy a wide range of applications. The most effective WSO is semantic orientation inference using point-wise mutual information proposed by Turney. Here, the semantic orientation of a word can be computed from the strength of its association with predefined positive and negative paradigm words. When dealing with phrases heuristic combination methods are used to generate vectors of semantic orientations. Such a Semantic Hyperspace Analogue (S-HAL) showed a more accurate representation of semantic characteristics [43]. In [44], the authors combine a probabilistic model of opinions and a stochastic mapping model between words to approximate a language model of products. The idea was to apply stochastic co-occurrence mapping between words to capture the hidden latent mapping of the probability model of opinion words and concepts.

Commonsense comprises of things we know about the world but which are usually left unstated in discourse, e.g., that things fall downwards (and not upwards) and people smile when they are happy. The difference between common and commonsense knowledge can be expressed as the difference between knowing the name of an object and understanding the same object's purpose. While the bag-of-concepts model helps to overcome problems such as word-sense disambiguation and semantic role labeling, the bag-of-narratives model will enable tackling NLP issues such as co-reference resolution and textual entailment.

Such models were not suitable for long reviews where class labels keep changing. Hence, in recent years temporal models such as recurrent neural networks are being used to identify subjective sentences. Furthermore, fusion with corresponding video and audio review is also being done using multiple kernel learning that is easily portable to new languages. Several authors proposed use of deep neural networks to automatically learn features from 2005 to 20011. Here, a dictionary of features in the form of convolution kernels is learned simultaneously. The lower layers learn abstract concepts and the higher layers learn complex features for subjective sentences.

When dealing with social media, the content is often diverse and noisy, and the use of a limited number of affect words or a domain-dependent training corpus is not enough. Early works in opinion mining were mainly supervised approaches relying on manually labelled samples such as a movie or product review, where the commentators overall positive or negative attitude was explicitly indicated. Later works have taken down text analysis granularity to the sentence level, for example by using the presence of opinion-bearing lexical items to detect subjective sentences or by using semantic frames for identifying the sentiment topics (or targets). In this work, common and commonsense knowledge were blended together to build a comprehensive resource that can be seen as an attempt to emulate how tacit and explicit knowledge is organized in the human mind.

Well-known common knowledge bases include WordNet, Freebase [45] and YAGO [46]. Here, they apply multidimensional scaling to common and commonsense knowledge base to grasp the semantic and affective similarity between different concepts after plotting them into a multidimensional vector space. Rather than merely determined by a few human annotators and represented as a word-word matrix, the vector space in [47] is built upon a commonsense knowledge base represented as a concept-feature matrix. In the next section, we detail the advantages and limitations of most commonly used subjectivity detection methods.

Statistical NLP methods feed a large training corpus of annotated texts to

Table 1: Summary of subjectivity detection syntactic methods

Method	Model	Advantages	Disadvantages
Hand-crafted Features [35, 36]	Identify frequently occurring templates and patterns in Subjective sentences (Table 3)	• Computationally very fast • Human annotators can design high quality features • Classifiers can be easily trained on the features	• High number of False Positives • Each word is assumed to be independent • Reliant on obvious words
Ontology Model [48, 49, 39, 50]	Comprehensive set of ontologies that define relationships between different classes of words and projects them into a vector space	• Considers the 'affinity' of a word to a sentence • Capture the underlying structure of grammar in a sentence • Scalable to a large vocabulary of concepts	• Works well only on strong emotions

a machine learning algorithm. In this way, they can learn polarities of words in addition to concepts automatically. Natural language data take discrete structures (known as parse tree) hence co-evolution kernels such as sequence and tree kernels are advantageous for high accuracy. In [51] the authors treat sentences as sequence kernels. Significant sub-sequences are selected with a high score computed in a recursive manner. In this way, statistical feature selection can enable us to use large sub-structures effectively. This method however is unable to capture bi-grams such as 'touchscreen' that are easily captured using convolution kernels.

Similarly, in [52] the authors tried to minimize the resources to build a subjectivity lexicon in foreign languages. They used bootstrapping to sample subjective clues from a few manually selected seed words. In each iteration, candidate words with low similarity with the original seed list are discarded.

Table 2: Summary of subjectivity detection syntactic methods

Method	Model	Advantages	Disadvantages
Statistical Model [51, 52]	Annotated corpus used to train a Classifier	• Natural representation of Parse tree as kernels • Portable to new languages and domains	• Unable to capture bigrams • Noise keeps increasing with number of predictions
Latent Dirichlet Model [53, 54]	Word frequency is used to compute posterior distribution	• Classifier trained on one product can be used on another. • Prior for words such as 'awesome' can be included	• Assumes that each word is independent of the others. • Cannot differentiate between antonyms such as 'bad/good'

The method is limited by the fact that suitable seed words may be difficult to determine in some domains or languages. Furthermore, with each bootstrap iteration the noise in the subjectivity dataset keeps increasing. Spanish sentences were first translated to English and then used to train a subjectivity classifier in [55]. However, translation of sentences can lead to loss of lexical information such as word sense resulting low accuracy.

4. Syntactic Methods

Today, syntax-cantered NLP is still the most popular way to manage tasks such as information retrieval and extraction, auto-categorization, topic modeling, etc. Table 1 and Table 2 show a summary of different subjectivity detection methods that are based on syntax or word frequency. Table 5 details the underlying equations that use word frequency to determine subjectivity. Syntax-cantered NLP can be broadly grouped into three main categories: keyword spotting, lexical affinity, and statistical methods.

4.1. Hand-crafted Features : Keyword Spotting

Keyword Spotting is the most Naïve approach and probably also the most popular because of its accessibility and economy. Text is classified into categories based on the presence of unambiguous words. One scheme uses the concept of private state that is a general term for opinions and emotions that are positive or negative [35]. The phrase “Injustice cannot last long” contains a negative private state. Human annotators are used to judge the strength of each private state as low, medium, high, or extreme. A sentence is subjective if it contains a private state and all other sentences are objective.

In [35], the authors ranked patterns using the conditional probability given by the frequency of a pattern in a subjective sentence given the total frequency of each pattern in all training sentences. For example all sentences that contain the verb *asked* in the passive voice are subjective, similarly expressions involving the noun $fact$ are highly correlated with subjective expressions. The drawback with this approach is that it is unable to identify objective patterns effectively resulting in false positives.

Similarly, [36] created a rule-based subjectivity dataset using a list of subjectivity clues and patterns. Next, a Naïve Bayes classifier was trained on patterns as well as pronouns, adjectives, cardinal numbers and adverbs features in subjective and objective sentences. The drawback of their approach was that they assume that low subjectivity score sentences may be objective. However, it is difficult to identify objective sentences since any objective sentence can be made subjective using a subjective modifier. Some authors have shown that in some patterns may be highly correlated with objectivity in a particular domains. For example in Wall Street journals, sentences containing ‘price’ or ‘profit’ are likely to be objective.

The NB algorithm uses word frequency to compute probabilities and makes the naïve assumption that the probability of occurrence of each word is independent of others in a sentence. Consider a document with vocabulary size V and class labels $y = \{-1, 0. +1\}$.We can compute the probability that word w_i

Table 3: POS tags for positive, negative, and neutral subjectivity clues

	Weakly Subjective			Strongly Subjective		
	verb	noun	adj	verb	noun	adj
positive	reform satisfy calm succeed	abundance competence values understanding	well-publicized boost affordable conviction premium	aver urge want understood	will celebration crusader ideal fanfare	hug closeness fashionable willing blithe
negative	manipulate decrease muddle withhold	dictator assault encroachment dictator insignificance	anomalous unsustainable self-defeating cursory urgent	blister mismanage unleash disconcert carp	anti-Israeli stern recklessness pratfall dismissive	ludicrous neglected fainthearted egocentric unwilling
neutral	engross hypnotize touch	transparent legalistic sovereignty notion	un-audited relations quick rare	conjecture metaphorize air surprise theoretize	emotion view foresee baby allusion	adolescents surprise stir considerable view

belongs to class y_j as follows :

$$p(w_i|y_j) = \frac{n_{ij}}{\sum_{w_i \in V} n_{ij}} \tag{2}$$

where n_{ij} is the frequency of word w_i in all documents of class y_j in the training data.

The major weakness of keyword spotting lies in its reliance upon the presence of obvious words which are only on the surface. In text document about ‘dogs’ where the word ‘dog’ is never mentioned, e.g., because ‘dogs’ are addressed according to the specific breeds they belong to, might never be retrieved by a keyword-based search engine. Table 3 shows commonly occurring clue words and corresponding POS tags in positive, negative, and neutral sentences.

4.2. Ontology Model : Lexical Affinity

Lexical Affinity is slightly more sophisticated than keyword spotting as, rather than simply detecting obvious words, it assigns to arbitrary words a probabilistic 'affinity' for a particular category. For example, General Inquirer is a lexicon made of 10,000 words grouped into 180 categories that are used for content analysis [48]. Another limitation of rules or patterns is that they are not scalable, hence Resource Description Framework (RDF) was developed to support the subject-predicate-object model that makes assertions about a resource. RDF-based reasoning engines have been developed to check the semantic consistency, which then helps to improve ontology classification. Next, Ontology Web Language (OWL) [49] was developed as a resource that extends RDF to provide a comprehensive set of ontology representations, such as the definition of classes, relationships between classes and their properties. In general, OWL requires strict definitions of static structures that work well on declarative sentences, and therefore is not suitable for representing knowledge that contains subjective degrees of confidence.

Keyword spotting is unable to capture the underlying structure of grammar in a sentence. Another popular resource is SenticNet [56], which contains 100,000 concepts (single words and multi-word expressions) as nodes and the edges determine the relationships among them. The vector space representation of SenticNet is termed AffectiveSpace [57], where each of the 100 dimensions corresponds to an eigenmood. Lastly, 24 basic emotions in the Hourglass model are used as centroids to cluster AffectiveSpace. A semantic parser breaks a sentence into clauses and then clauses are deconstructed into concepts. For example, the clause 'I went for a walk in the park, would contain the concepts 'go walk and 'go park. Such concepts can be easily classified as positive or negative using AffectiveSpace. A limitation of AffectiveSpace is that it is unable to identify neutral sentiments. Hence, we consider the previously described deep learning model to identify noise or factual comments in the reviews. Figure 3 illustrates two components of AffectiveSpace, we can see that neutral concepts lie closer to each other.

WordNet is a large lexical database of English Nouns, verbs and adjectives that are grouped into sets of cognitive synonyms [39]. SentiWordNet is a resource for opinion mining that is built on top of WordNet, which assigns each synset in WordNet with a score triplet (positive, negative and neutral). Currently, SentiWordNet includes an automatic annotation for all synsets in WordNet, totaling more than 100,000 words [50].

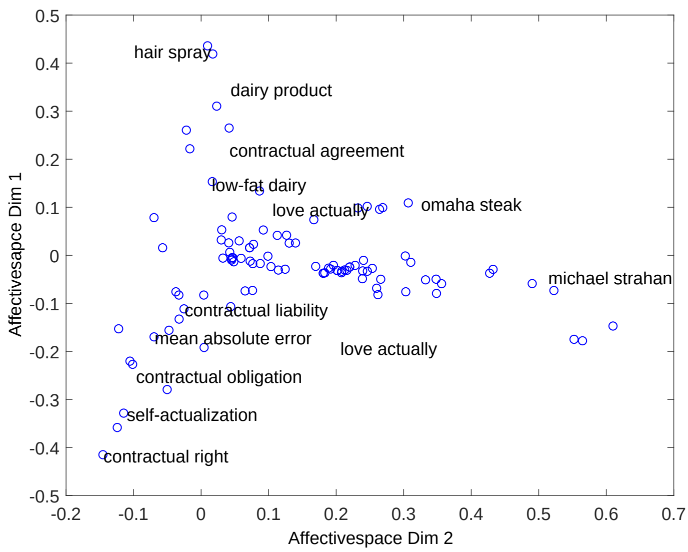

Figure 3: Two eigenmoods of AffectiveSpace. We can see that neutral concepts lie closer to each other.

4.3. Statistical NLP

Statistical NLP methods feed a large training corpus of annotated texts to a machine learning algorithm [58]. In this way, they can learn polarities of words in addition to concepts automatically. Natural language data take discrete structures (known as parse tree) hence co-evolution kernels such as sequence and tree kernels are advantageous for high accuracy. In [51] the authors treat sentences as sequence kernels. Significant sub-sequences are selected with a high score computed in a recursive manner. In this way, statistical feature selection can enable us to use large sub-structures effectively. This method however is unable to capture bi-grams such as 'touchscreen' that are easily captured using convolution kernels [59].

Similarly, in [52] the authors tried to minimize the resources to build a subjectivity lexicon in foreign languages. They used bootstrapping to sample subjective clues from a few manually selected seed words. In each iteration, candidate words with low similarity with the original seed list are discarded. The method is limited by the fact that suitable seed words may be difficult to determine in some domains or languages. Furthermore, with each bootstrap iteration the noise in the subjectivity dataset keeps increasing. Spanish sentences were first translated to English and then used to train a subjectivity classifier in [55]. However, translation of sentences can lead to loss of lexical information such as word sense resulting low accuracy.

4.4. Latent Dirichlet Model

Subjectivity classification is different from review classification task because it uses sentences as opposed to entire documents and the target concept is subjectivity instead of opinion polarity. Latent Dirichlet Allocation (LDA) is a probabilistic model that assumes each document is a mixture of latent topics [60]. Furthermore, we introduce a sentiment component that uses sentiment annotations to constrain words expressing similar sentiment to have similar representations. In particular, [53] uses logistic regression predictor that defines a hyper plane in the word vector space where a word vectors positive sentiment probability depends on where it lies with respect to this hyper plane. The full learning objective maximizes a sum of two objectives presented. Because the maximum a-posterior estimation problem for different documents is independent, they can be solved in parallel. The method clearly outperforms other vector space models and performs best when combined with the original BOW representation.

Similar to the NB algorithm, the LDA algorithm also uses word frequency to compute probabilities. Consider a document with vocabulary size V and class or topic labels $y = \{Topic1, Topic2, Topic3\}$. We can compute the probability that word w_i belongs to class or topic y_j as follows :

$$p(w_i|y_j) = \left(\frac{n_{ij}}{\sum_{w_i \in V} n_{ij}} \right) \times \left(\frac{n_{kj}}{\sum_{y_j \in y} n_{kj}} \right) \quad (3)$$

where n_{ij} is the frequency of word w_i in all documents containing topic y_j and n_{kj} is the number of documents containing topic y_j in the training data. Each topic learned will belong to a distinct sentiment class.

Subjectivity detection is context sensitive, so that classifiers trained on one domain often fail to produce satisfactory performance when shifted to new domains. Previously topic-sentiment mixture models have been proposed to capture mixture of topics and sentiment simultaneously on Weblogs. A small number of seed words with known polarity are used to infer the polarity of a large set of unidentified terms. If a sentence does not match any sentiment words, its prior subjectivity label is randomly sampled. Gibbs sampling is used to estimate the posterior by sequential sampling each variable of interest from the distribution over that variable given the current value of all other variables and the data. It was found that while incorporating words that are more subjective can generally yield better results, the performance gain by employing extra neutral words is less significant [54].

5. Semantic Methods

Semantics-based NLP focuses on the intrinsic meaning associated with natural language text. Rather than simply processing documents at syntax-level, semantics-based approaches rely on implicit denotative features. Table 4 shows a summary of different subjectivity detection methods that model sentence semantics. Table 5 details the underlying equations that use word vectors to determine subjectivity.

Endogenous NLP automatically learn concepts from documents by training state space graphs where nodes are the words and the arc determine causal dependencies among them in large documents. In this way no prior semantic understanding of documents or linguistic databases are needed. Conditional

Table 4: Summary of Subjectivity Detection Semantic Methods.

Method	Model	Advantages	Disadvantages
Conditional Random Fields [61]	Sequence labeling tasks such as POS Tagging and Shallow Parsing	Can capture word order and grammar well in the form of n-grams	Very high dimensionality of features
Semi-supervised learning [37]	A small number of seed words of known polarity are used. Then highly similar words are determined in each iteration.	Since, there is coherence between Subjective sentences it can quickly determine polarity of large set of terms	Graph cuts may not preserve alternating subjective and objective sentences in a review
Deep Learning [62]	Input sentence is processed by several layers that are trained using back-propagation	Meta-level feature representation works well on large vocabularies in contrast to n-gram models	Social media data are very noisy and a domain-dependent corpus is not enough to train the model
Multiple Kernel Learning [63]	Features are organized into groups and each group has its own kernel function	Audio-visual features can be combined with text	Computationally very slow

Random Fields (CRF) is commonly used for sequence labeling tasks such as POS tagging, named-entity recognition, and shallow parsing [64]. Shallow parsing can be used to summarize relevant information from documents by labeling each word as +1 or -1 denoting that it is included or excluded from the summary. Similarly, [61] proposed an isotonic CRF to model sentiment flow in a document in an author-dependent manner. They used ordinal binary features to label sentences in a sequence as negative, neutral, or positive. Such a method is limited as features are only able to capture sentiments across sentences, however it is unable to use features inside a single structure.

To leverage on coherence of subjectivity in nearby sentences in [37], they determined pairwise interactions between pairs of sentences and used graph cuts to determine subjective regions in the document prior to polarity classification. It is easy to see that graph cuts may not preserve alternating subjective and objective sentences in a product review.

5.1. Semantic Sentence Model

In tasks where one is concerned with a specific sentence within the context of the previous discourse, capturing the order of the sequences preceding the one at hand may be particularly crucial. We take as given a sequence of sentences $s(1), s(2), \ldots, s(T)$. Each sentence in turn is a sequence of words so that $s(t) = (x_1(t), x_2(t), \ldots, x_{\mathrm{L}}(t))$, where L is the length of sentence $s(t)$.

Thus, the probability of a word $p(x_i(t))$ follows the distribution :

$$p(x_i(t)) = P(x_i(t)|(x_1(t), x_2(t), \ldots, x_{i-1}(t)), (s(1), s(2), \ldots, s(t-1)) \quad (4)$$

The word-vector model represents each word as a d dimensional vector that is computed from co-occurrence data using Eq. (4). When we concatenate the word vectors of all words in a sentence of length L, it results in a 2D input vector of dimension $\mathrm{L} \times d$.

5.2. Parse Trees

It can be seen that many short n-grams are neutral while longer phrases are well distributed among positive and negative subjective sentence classes. Therefore, matrix representations for long phrases and matrix multiplication to model composition are being used to evaluate sentiment. In such models sentence composition is modeled using deep neural networks such as recursive auto-associated memories [59, 65]. Recursive neural networks predict the sentiment class at each node in the parse tree and try to capture the negation and its scope in the entire sentence. In the standard recursive neural network, each word is represented as a vector and it is first determined which parent already has all its children computed. Next, the parent is computed via a composition function over child nodes. In Matrix Recursive Neural Networks, the composition function for long phrases depends on the words being combined and, hence, is linguistically motivated. However, the number of possible composition functions is exponential, hence in [66] the recursive neural tensor network (RTNN) was introduced that uses a single tensor composition function to define multiple bilinear dependencies between words.

The RTNN model transforms each sentence into word vector representation of dimension d. Consider two word vectors w_1 and w_2, then the parent word vector w_{12} can be computed as follows:

$$w_{12} = f(\mathbf{w} \times [w_1 \ w_2]^T) \tag{5}$$

where $\mathbf{w}$ is a weight matrix of dimension $d \times 2d$ and is learned via back-propagation using class labels in the final layer.

Table 5: Equations for Different Methods

	Naïve Bayes	LDA
BOW	$p(w_i \vert y_j) = \frac{n_{ij}}{\sum_{w_i \in V} n_{ij}}$	$p(w_i \vert y_j) = \left(\frac{n_{ij}}{\sum_{w_i \in V} n_{ij}}\right) \times \left(\frac{n_{kj}}{\sum_{y_j \in y} n_{kj}}\right)$
	RTNN	CNN
Word Vector	$w_{12} = f(\mathbf{w} \times [w_1 \ w_2]^T)$	$\mathrm{E}^l = -\sum_{z=1}^{\mathrm{Z}} \sum_{i,j}^{(\mathrm{L_x}-n_x+1),(\mathrm{L_y}-n_y+1)} \sum_{r,s}^{n_x,n_y} v_{i+r-1,j+s-1} h_{ij}^z w_{rs}^l$

5.3. Convolutional Neural Networks

The simple additive or multiplicative models described in the previous sections do not take into account word order or structure. Another challenge with movie reviews is that real facts are mixed with objective sentences like discussions about the plot in the movie or discussion about the characteristics of the actors and actresses with the opinion sentences about the quality of the movie. Lastly, the system should be able to analyze thousands of reviews efficiently.

Instead of using the parse tree, convolution neural networks (CNN) consider a sentence sub-matrix of words and features as input [59]. Here narrow or wide convolution is achieved by applying filters such as pattern templates across the input features. Each convolution layer is interleaved with a max-pool layer that eliminates redundant values. In this way, filters in higher layers can capture syntactic relations between phrases far apart in the input sentence.

Most existing algorithms for learning continuous word representations typically only model the syntactic context of words but ignore the sentiment of text. It is desirable that learning algorithms are less dependent on extensive feature engineering, The conventional Collobert and Weston model is a neural network based on syntactic context of words where the output is a score [67]. In [15], the authors propose neural networks to integrate the sentiment information of tweets. Sentiment-specific word embedding (SSWE) predicts the sentiment distribution of input text based on n-gram however, it leaves out the context of words. The output layer now has, two neurons corresponding to positive or negative sentiments. A unified model of the C & W and SSWE is used, where the output is a two dimensional vector for the syntactic and semantic scores computed using hinge loss.

In [68], the authors introduce a hierarchical convolution neural network (HCNN) where convolutions are applied feature-wise, across each feature of the word vectors in the sentence. The hierarchy of convolution kernels of increasing sizes, this allows for the composition operation to be applied to sentences of any length, while keeping the model at a depth of roughly $\sqrt{2l}$ where is l is the length of the sentence. The top layer is the value for that feature in the resulting sentence vector. Such a model is sensitive to the structure of the discourse and can capture subtle aspects when coupled with further semantic data and unsupervised pre-training.

5.4. Deep Neural Networks

While CNN are supervised models, deep neural networks are unsupervised with the potential use in dialogue tracking and question answering systems. In [62] the authors advocate a deep neural network architecture, trained in an end-to-end fashion. The input sentence is processed by several layers of feature extraction. As they deal with raw words and not engineered features, the first layer has to map words into real-valued vectors for processing by subsequent layers of the neural network. It is common for the role of a word to depend on words far away in the sentence and, hence, outside of the considered window. The output of the l^{th} layer contains n hidden units that are trained using back propagation.

Deep learning exploits unsupervised learning to discover concepts allowing one to exploit the large amounts of unlabelled data across different domains to learn these intermediate representations. Here, we do not engineer the intermediate concepts but instead use generic learning algorithms to discover them. In [65], the authors used domain adaptation on the Amazon dataset containing 340,000 reviews regarding 22 different product types and for which reviews are labelled as either positive or negative. There is a vast disparity between domains in the total number of instances and in the proportion of negative examples.

However, one sentiment word may imply two opinion polarities in different domains resulting in serious ambiguity in classifying sentences. To disambiguate the sentiment-ambiguous adjectives heuristics can be used. For example, in [69] they consider a Bayesian model of candidate features for word polarity disambiguation. They use features such as opinion target, modifying word and indicative words.

The problem of contextual polarity often causes classification error. For example, the word 'conspiracy' is negative in many domains, but in the mystery novel domain, it is a favorable factor indicating positive sentiment. Previous approaches learned sentiment classifiers using a single domain corpus and then adjusted it to a different domain. This method selectively switches the sentiment polarity of the entry words to adapt to a domain. Hence it compares positive/negative reviews dictionary word occurrence ratios with the positive/negative review ratio itself to determine which entry words to be removed and which entry words sentiment polarity to be switched [70].

The computational speed of deep learning increases with large number of training samples. To overcome this problem, the features learned via deep learning are fed into a single layer neural network called Extreme Learning Machine. In such a model, the input layer of hidden weights is randomly initialized.

Next, the output layer of weights can be computed heuristically using Bayesian networks resulting in high accuracy and low computational cost [71].

6. Adapting to New Domains

In this section, we discuss the models that are portable to resource deficient languages such as 'Arabic' or 'French'. Next, we look at fusion of audio and video data when sufficient training data in the form of text is unavailable.

6.1. Foreign Languages

Rich linguistic resources such as WordNet or polarity tagged words are not available for some languages affecting the performance of word polarity identification. Here, dictionary based approaches for identifying word polarities are commonly used that assume that synonyms convey same orientation and antonym relations convey an inverse sentiment.

In [72], a semi-supervised method is proposed to identify the polarity in resource lean languages such as Persian, which do not have WordNet. Two relatedness graphs are built for the foreign and English language, they are then connected using mappings from foreign senses to the English senses available in WordNet. For cases were mappings are incomplete an online translator is used to connect the two graphs .

The machine learning approach requires a large training data, which is difficult to obtain for many languages, and semantic orientation of words requires large amount of linguistic resources, which depend on the language. Hence, we need to consider a hybrid approach of both methods [73]. Most of the research in opinion mining is in English, hence we can extract information in the target language for example in Spanish or Arabic and translate it into English. Subjectivity detection in foreign languages was proposed by translating English lexicons in [74]. However, translation requires the lemmatized form of words, which can lead to loss of subjective form. For example, the lemma form of 'memories' is 'memory', when translated to Romanian becomes 'memories' objective meaning the power of retaining information.

Subjectivity loss during translation may also be due to word ambiguity in either the source or target language [75]. For example, the word 'fragile' translates into Romanian as 'fragil' that would refer to breakable objects, and it loses it subjectivity meaning of delicate. Another example is the word 'one-sided', completely loses subjectivity once translated, as it becomes in Romanian 'cu o sinura latura', meaning 'with only one side (as of objects)'. In [76], the authors target morphologically rich languages such as Arabic. They replaced words with low frequency by a single word 'unique'. Next, they considered a polarity lexicon from Arabic news articles. In particular, they consider morpheme which are basic linguistic units, for example the word table is made of two morphemes 'table' and 's'. Hence, new words in a language are derived from combining morphemes and semantic drift over time and usage.

6.2. Multi-modal optimization

The majority of state-of-the-art frameworks rely on single modality, i.e., text, audio or video. These systems exhibit limitations in terms of accuracy and can be used in a very restricted way in real applications. For example, Turney achieved 84% accuracy for automobile reviews using only text [77]. Ekman showed that universal facial expressions provide sufficient clues to detect emotions [78]. In [79], the authors have used supervised learning to train three datasets : 1. ISEAR emotion detection from text, 2. CK++ emotion detection from facial expression and 3. eNTERFACE to model emotion extraction from audio.

Similarly, when viewing a photograph, the sequences of eye movements we make are referred to as scan paths. In [80], the authors showed that scan paths were more similar when pictures were described compared to when imagined. Hence, verbalizing a memory of a previously viewed scene leads to increased recognition accuracy. Other modalities such as visual cues can be used to address sentiment analysis. In [81], the authors create a dataset called Institute for Creative Technologies Multi-Modal Movie Opinion (ICT-MMMO) from online social review videos that encompass a strong diversity in how people express opinions about movies and include a real world variability in video recording quality.

For robust sentiment analysis of movie reviews online sources such as WordNet, ConceptNet and General Inquirer are used to determine the semantic relations among words. A challenge with reviews is that movies that originally received a positive review have a greater chance of receiving follow-up reviews because more people will see these movies. They manually annotated YouTube videos with a score of 1 to 5. Visual cues were head gestures such as head nods and shakes. The audio-visual and linguistic information was merged and served as input for bidirectional long short-term memory (BiLSTM) network for sentiment prediction. While SVM generates one prediction for each movie review video, the BiLSTM network outputs a sentiment score for each spoken utterance.

In [82], the authors used the text arising from transcribed speech, that is the phoneme representations of the words in an utterance, and the corresponding acoustic features extracted to predict subjectivity. Each utterance may be classified as subjective because the speaker is expressing an opinion, or because the speaker is discussing someone else's opinion, or because the speaker is eliciting the opinion of someone else with a question. A subjective utterance is a span of words where a private state is expressed. This could also be the private state of someone else. For example, 'Finding them is really a pain, you know'.

6.3. Multiple Kernel Learning

Kernel methods have recently become popular, as non-linear kernels such as radial basis functions, show higher accuracy compared to linear classification models. It is often desirable to use multiple kernels simultaneously as multiple feature representations are derived from the sentences or because different kernels such as RBF or polynomial are used to measure the similarity between two sentences for the same feature representation. Multiple Kernel Learning (MKL) is a feature selection method where features are organized into groups and each group has its own kernel function [83, 84]. However, the choice of kernel coefficients can have significant impact on the classification accuracy and efficiency of MKL [63].

Most previous applications of MKL have been in image and video classification and object recognition. For example in [85], multiple kernel learning was used simultaneously optimize different modalities in Alzheimer disease images since different types of tests may reveal different aspects of the diagnosis. Recently, MKL with Fourier transform on the Gaussian kernels have been applied to Alzheimer disease classification using both sMRI and fMRI images [86]. MKL was also used to detect presence of large lump in images using a convolution kernel on Gaussian features [87].

In [88], higher order kernels are used to enhance the learning of MKL. Here, block co-ordinate Gradient optimization is used that approximates the Hessian matrix of derivatives, as a diagonal resulting is loss of information. Group-sensitive MKL for object recognition in images integrates a global kernel clustering method with MKL for sharing of group-sensitive information [89]. They showed that their method outperformed baseline-grouping strategies on the WikipediaMM data of real-world web images. The drawback of this method is that a looping strategy is used to relabel the groups and may not reach the global optimum solution. In [90], MKL was also used to combine and re-weight multiple features by using structured latent variables during video event detection [90]. Here, two different types of kernels are used to group global features and segments in the test video that are similar to the training videos. The concept of kernel slack variables for each of the base kernels was used to classify YouTube videos in [91]. In order to select good features and discard bad features that may not be useful to the kernel, [92] used a beta prior distribution. Lastly, Online MKL shows good accuracy on object recognition tasks by extending online kernel learning to online MKL, however the time complexity of the methods is dependent on the dataset [93].

In the case of sentiment analysis, MKL was applied to a Polish opinion aggregator service that contains textual opinions of different products in [94], however they did not consider the hierarchical relation of different attributes of products. Video and text multi-modal features were also fused at different levels of fusion for indexing of web data in [95], however they are computationally very slow. It can be seen that the main challenges in using MKL is the computational time and the choice of suitable grouping strategy.

6.4. Annotation of Emotions

Emotions in any brain model emerge as a feedback for any system errors, so that a more accurate estimate of the external world can be achieved [96]. Hence, emotions arise as it tries to minimize its model-prediction errors and other system errors in response to external inputs also known as Hebbian learning. Furthermore, the global error is used as an emotional feedback, and the brain model will not stay in an emotional state forever, but the emotional state will change constantly depending on the environmental conditions. This is consistent with the explanation why happiness or unhappiness does not last forever; it changes over time when the circumstances change.

Annotation of emotions using machine learning is a challenge as the widely used terms of emotional state often differ from the more generic affective state that is adequate from a psychological point of view [97]. There has been increasing interest in understanding how the brain processes emotions and the cognitive and social constituents of emotions. This can be successfully applied to cognitive processes at different levels in human-machine interaction such as generation of embodied conversational agents and automatic extraction of emotional behavior related features for dialog systems. Figure 4 shows emotions in most objective words are almost equal to zero.

Expressiveness of opinion varies widely from person to person. Some people express their opinions more vocally, some more visually and others rely exclusively on logic and express little emotion. The hourglass model [57] describes emotions along four affective dimensions namely Pleasantness, Attention, Sensitivity and Aptitude. Each emotion has six sub-emotions resulting in 24 basic emotions. To determine the value of each sub-emotion for a concept, we cluster AffectiveSpace four times. Each time the centroids are six sub-emotions for that emotion. Furthermore, four basic emotions can be used to compose two complex emotions. The severity of the emotion ranging from -1 to +1 in AffectiveSpace. When the emotion value is close to 0 then it is a neutral concept. For example, in Figure 4 we show the value of Aptitude and Sensitivity for objective concepts such as 'electric_guitar' and 'fire_truck' is almost zero.

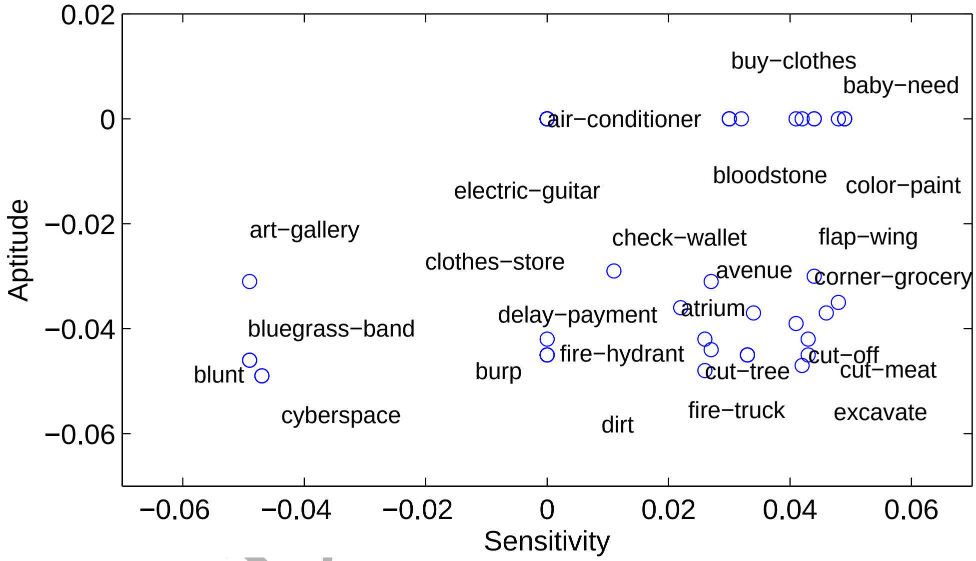

Figure 4: Emotions in Objective Words

7. Benchmarks for Subjectivity Detection

In this section, we provide a summary of subjectivity benchmark datasets. We also compare the accuracies by different methods on each dataset in Table 6.

7.1. MPQA

The Multi Party Question Answering (MPQA) corpus [36] is a collection of 535 English-language news articles from a variety of news sources manually annotated for subjectivity after machine translation from Spanish. There are 9,700 sentences in this corpus, 55% of the sentences are labelled as subjective while the rest are objective. Table 6 provides the accuracies from 4 different baseline algorithms. We can see the convolutional neural networks (CNN) outperform other models by over 20% on this dataset. This is because it uses kernels of different sizes to captures features from the data. Both Rule based methods (Rule) [74] and Bootstrapping (BS) [52] have an accuracy of about 62%. The BS method starts with a set of seed words and includes new words into the lexicon with maximum similarity to the seed words in each iteration of the Bootstrap.

7.2. MPQA Gold

The MPQA Gold corpus has 504 Spanish sentences that are manually annotated for subjectivity by multiple annotators. The annotation resulted in 273 subjective and 231 objective sentences as described in [74]. Table 6 shows that again CNN outperforms other methods by big margin. The accuracy of unsupervised word sense disambiguation (UWSD) is much lower. WSD and rule based classifiers are heavily dependent on templates and do not consider the relative positions between nouns and verbs [98].

7.3. Movie Review

The Movie Review subjectivity dataset contains 5000 movie review snippets (e.g., bold, imaginative, and impossible to resist) from www.rottentomatoes.com. To obtain (mostly) objective data, 5000 sentences were collected from plot summaries available from the Internet Movie Database (www.imdb.com). Sentences or snippets are at least ten words long and drawn from reviews or plot summaries of movies released post-2001, which prevents overlap with the polarity dataset [37]. The best results for this dataset were reported by Phrase Kernels (PhK) [99] and the word vector (WV) model described in [100]. However, even a simple BOW model also showed comparable results [37].

7.4. TASS

The Taller de Anlisis de Sentimientos en la SEPLN (TASS) workshop corpora is a collection of Spanish tweets commonly used for the evaluation of social media analysis tasks [101]. It has a training set of 7219 tweets by 150 public figures coming from politics, sports, or communication. The tweets were collected during the year 2011-2012. Each one is annotated with one of these four categories: *positive*, *neutral*, *negative*, or *without opinion*. Hence, the results reported are for both sentiment and subjectivity analysis that is predicting neutral tweets. A test subset containing 1000 tweets with a similar distribution to the training corpus and manually labelled. The best results were reported by [102]. They propose a framework for subjectivity detection in Spanish by automatically extracting convolution features in Spanish and the translated English form of each sentence. The aligned features of both languages for each sentence are then combined using multiple kernel learning. Another baseline LIF [101], showed 10% results. Their approach was limited as they relied heavily on polarity lexicons that are not available in Spanish.

7.5. Twitter

Twitter dataset is a collection of tweets collected from the time period between April 6, 2009 to June 25, 2009 [38]. It contains 498 tweets manually annotated as positive, negative, or neutral. Table 6 shows that a simple neural network (NN) shows a high accuracy of 91.4%. Another baseline SenticNet showed about 10% lower accuracy. SenticNet contains concepts as nodes and the edges determine the relationships among them. Semantic parser breaks a sentence into concepts and these concepts can be easily classified as positive or negative. Another Twitter dataset of 11,000 manually labeled tweets was developed in [103]. They considered positive, negative and neutral tweets. However, they only evaluated on a simple tree kernel algorithm. Similarly, in [104] the authors developed a Twitter sentiment dataset that was tagged using icons for different emotions.

7.6. Amazon

The Amazon dataset contains product review text and the corresponding rating labels (1-5) for 'Books', 'Dvd', 'Electronics' and 'Kitchen', taken from

Amazon.com. A rating of 1 is strongly negative, 2 is weakly negative, 4 is weakly positive, and 5 is strongly positive. Reviews with rating label 3 are discarded as they are deemed as ambiguous and, hence, indecisive about a product. We can also reason that weak sentiments are due to presence of neutral sentences. Cross-domain tasks evaluate the accuracy of the model when it is trained in one product and tested on another. The highest accuracy was reported by Transfer Deep Network (TDN) [105]. In TDN, the authors considered two parallel deep auto-encoders to learn transferable features and classification features. In Rule3(R3) [106], the authors proposed three rules with hand-crafted features that must be satisfied for cross-domain classification.

Table 6: Summary of Subjectivity Detection Benchmarks and their Accuracy.

Dataset	Size	Method 1	Method 2	Method 3	Method 4
MPQA	9700	62 (SVM) [55]	62 (Rule) [74]	62 (BS) [52]	89.6 (CNN) [107]
MPQA Gold	504	86.3 (NBSVM) [108]	80.35(SWSD) [109]	60(UWSD) [98]	89.4(CNN-MC) [110]
Movie	10000	92.7 (PhK) [99]	87.7 (BOW) [37]	66.6 (LDA) [111]	88.13 (WV) [100]
TASS	10000	63.7 (LYS) [112]	69.2 (LIF) [101]	88.4 (LDNN) [102]	-
Twitter	498	91.4 (NN) [41]	68.2 (SenticNet) [56]	81.8 (SVM) [41]	62.5 [38]
Amazon	8000	71.4 (R3) [106]	80.6 (TDN) [105]	71.5 [38]	-

8. Challenges

Extracting and aggregating opinions from text is a sub-field of information fusion whose aim is to understand the context of the words used. This requires access to large amounts of domain-specific benchmarks that are collectively used to train sentiment classifiers. With the explosion of deceptive activities such as web trolling and opinion spamming, the efficiency of sentiment prediction is declining. For meaningful opinion aggregation, it is critical to discard both opinion spam and unopinionated contents as these can confuse the model trained for positive and negative classification. In this section, we summarize the four main challenges of subjectivity detection.

The first challenge is that subjectivity detection itself is a subjective task, i.e., a piece of text may be neutral to some people but not to others. This is can be caused by a diverse level of expertise on a given topic of discussion but also by a different interpretation of a sentence in multi-lingual settings. In some cases, this is also related to personal preferences, political views or even personality of the user. To this end, user-profiling tasks such as gender detection [113], personality recognition [114], and community detection [115] should be carried out before subjectivity detection is performed.

The second challenge is improving the accuracy of subjectivity detection in short texts. Microblogging data, e.g., Twitter data, are often difficult to classify for the use of microtext, the lack of contextual information, and because they require suitable regularization due to missing data samples. Here, the number of false positives is often high due to lack of features defining neutral sentences. To this end, generative adversarial networks could be used to create additional data samples together with microtext normalization techniques, for converting informal text into plain English.

The third challenge is context dependency. Some words may be objective out of context but could assume subjectivity in a specific context or domain, e.g., the adjective long is neither positive nor negative but it could be positive in some domains, e.g., long battery, or negative in other contexts, e.g., long queue. Often the context of a word is dependent on words far away in the sentence and, hence, outside the window of adjacent words. Parse tree models in conjunction with word vectors are able to identify large sub-structures effectively.

Lastly, the fourth challenge is reducing the computational cost of training features from a large vocabulary of words. Traditional methods used hand-crafted features and templates to identify subjective sentences. Manual annotation of such a large feature set is a very hard and tedious task and it is essential to identify domain-independent affective words in conversations. To this end, unlike previous surveys, this paper reviewed word vector based subjectivity models that can learn features in an unsupervised manner. These employ convolutional kernels and a sliding window mechanism to learn significant features. Furthermore, the dictionary of features learned is portable across new products and languages.

9. Conclusion

Distinguishing between facts and opinions is possibly one of the most important sentiment analysis subtasks, as neutral comments can very negatively affect the information fusion process that enables a polarity classifier to mine and categorize positive and negative opinions. As NLP research is increasingly shifting from syntactic models to semantic models, subjectivity detection becomes a more and more difficult task. Following our previous review of overlapping NLP curves, we categorize subjectivity detection methods into syntactic and semantic models.

Several authors have reviewed sentiment analysis from text. However, the review of subjectivity detection is often overlooked or described as a subsection in the above reviews. In contrast, in this review we have focused only on subjectivity detection methods. Another limitation of previous works on this topic is that they do not cover recent word vector models such as convolutional neural networks. Lastly, multi-lingual subjectivity detection has not been reviewed previously.

The timeline of subjectivity is divided into three phases, the first being handcrafted features such as subjectivity clues, in the next phase automatic deep learning of concepts was dominant, recently the third phase is observed where multimodal models try to combine text with audio and video data. We can see that objective concepts tend to cluster together in AffectiveSpace. Similarly, we can conclude that concepts with high emotional value lack in objective words. Fusion of video and audio with corresponding utterance transcribed as text can significantly help in detecting sentiment in foreign languages and new domains.

References

[1] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, A Practical Guide to Sentiment Analysis, Springer, Cham, Switzerland, 2017.

[2] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective com-

puting: From unimodal analysis to multimodal fusion, Information Fusion 37 (2017) 98–125.

[3] S. Poria, E. Cambria, D. Hazarika, N. Mazumder, A. Zadeh, L.-P. Morency, Context-dependent sentiment analysis in user-generated videos, in: ACL, 2017, pp. 873–883.

[4] S. L. Lo, E. Cambria, R. Chiong, D. Cornforth, Multilingual sentiment analysis: From formal to informal and scarce resource languages, Artificial Intelligence Review 48 (4) (2017) 499–527.

[5] E. Cambria, A. Hussain, Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, Springer, Cham, Switzerland, 2015.

[6] L. Oneto, F. Bisio, E. Cambria, D. Anguita, Statistical learning theory and ELM for big social data analysis, IEEE Computational Intelligence Magazine 11 (3) (2016) 45–55.

[7] E. Cambria, S. Poria, A. Gelbukh, M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intelligent Systems 32 (6) (2017) 74–80.

[8] Y. Ma, E. Cambria, S. Gao, Label embedding for zero-shot fine-grained named entity typing, in: COLING, 2016, pp. 171–180.

[9] E. Cambria, S. Poria, R. Bajpai, B. Schuller, SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives, in: COLING, 2016, pp. 2666–2677.

[10] S. Poria, E. Cambria, D. Hazarika, P. Vij, A deeper look into sarcastic tweets using deep convolutional neural networks, in: COLING, 2016, pp. 1601–1612.

[11] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in: AAAI, 2018.

[12] B. Liu, Sentiment analysis and subjectivity, in: Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition, Chapman and Hall/CRC, 2010, pp. 627–666.

[13] M. Bonzanini, M. Martinez-Alvarez, T. Roelleke, Opinion summarisation through sentence extraction: An investigation with movie reviews, in: SIGIR, SIGIR '12, 2012, pp. 1121–1122.

[14] P. Biyani, S. Bhatia, C. Caragea, P. Mitra, Thread specific features are helpful for identifying subjectivity orientation of online forum threads, in: COLING, 2012, pp. 295–310.

[15] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, B. Qin, Learning sentiment-specificword embedding for twitter sentiment classification, in: ACL, 2014, pp. 1555–1565.

[16] M. V. de Kauter, D. Breesch, V. Hoste, Fine-grained analysis of explicit and implicit sentiment in financial news articles, Expert Systems with Applications 42 (11) (2015) 4999 – 5010.

[17] F. Benamara, B. Chardon, Y. Y. Mathieu, V. Popescu, Towards context-based subjectivity analysis, in: IJCNLP, 2011, pp. 1180–1188.

[18] T. Wilson, Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states, Ph.D. thesis, Intelligent Systems Program, University of Pittsburgh (2007).

[19] S. Vázquez, Ó. Muñoz-García, I. Campanella, M. Poch, B. Fisas, N. Bel, G. Andreu, A classification of user-generated content into consumer decision journey stages, Neural Networks 58 (2014) 68 – 81.

[20] P. Biyani, S. Bhatia, C. Caragea, P. Mitra, Using non-lexical features for identifying factual and opinionative threads in online forums, Knowledge-Based Systems 69 (2014) 170 – 178.

[21] S. Rill, D. Reinel, J. Scheidt, R. V. Zicari, Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis, Knowledge-Based Systems 69 (2014) 24 – 33.

[22] V. Kagan, A. Stevens, V. Subrahmanian, Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election, IEEE Intelligent Systems 30 (1) (2015) 2–5.

[23] S. Krishnamoorthy, Linguistic features for review helpfulness prediction, Expert Systems with Applications 42 (7) (2015) 3751 – 3759.

[24] G. Murray, G. Carenini, Subjectivity detection in spoken and written conversations, Natural Language Engineering 17 (2011) 397–418.

[25] E. Cambria, B. White, Jumping nlp curves: A review of natural language processing research [review article], IEEE Computational Intelligence Magazine 9 (2) (2014) 48–57.

[26] A. Balahur, R. Mihalcea, A. Montoyo, Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications, Computer Speech and Language 28 (1) (2014) 1 – 6.

[27] G. Bello-Orgaz, J. Jung, D. Camacho, Social big data: Recent achievements and new challenges, Information Fusion 28 (2016) 45–59.

[28] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (4) (2012) 675–679.

[29] F. Xing, E. Cambria, R. Welsch, Natural language based financial forecasting: A survey, Artificial Intelligence Review doi:10.1007/s10462-017-9588-9.

[30] M. Ebrahimi, A. Hossein, A. Sheth, Challenges of sentiment analysis for dynamic events, IEEE Intelligent Systems 32 (5).

[31] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, J. Munro, Sentic computing for patient centered application, in: IEEE ICSP, 2010, pp. 1279–1282.

[32] A. Valdivia, V. Luzon, F. Herrera, Sentiment analysis in tripadvisor, IEEE Intelligent Systems 32 (4) (2017) 72–77.

[33] A. Zadeh, P. P. Liang, S. Poria, P. Vij, E. Cambria, L.-P. Morency, Multi-attention recurrent network for human communication comprehension, in: AAAI, 2018.

[34] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, M. Huang, Augmenting end-to-end dialog systems with commonsense knowledge, in: AAAI, 2018.

[35] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: EMNLP, 2003, pp. 105–112.

[36] J. Wiebe, E. Riloff, Creating subjective and objective sentence classifiers from unannotated texts, in: CICLing, 2005, pp. 486–497.

[37] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: ACL, Association for Computational Linguistics, 2004, p. 271.

[38] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, E. Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Information Sciences 311 (2015) 18 – 38.

[39] G. A. Miller, Wordnet: A lexical database for english, Commun. ACM 38 (11) (1995) 39–41.

[40] S. M. Mohammad, P. D. Turney, Crowdsourcing a word-emotion association lexicon, Computational Intelligence 29 (3) (2013) 436–465.

[41] F. Bravo-Marquez, M. Mendoza, B. Poblete, Meta-level sentiment models for big social data analysis, Knowledge-Based Systems 69 (2014) 86 – 99.

[42] J. M. Chenlo, D. E. Losada, An empirical study of sentence features for subjectivity and polarity classification, Information Sciences 280 (2014) 275 – 288.

[43] T. Xu, Q. Peng, Y. Cheng, Identifying the semantic orientation of terms using s-hal for sentiment analysis, Knowledge-Based Systems 35 (2012) 279 – 289.

[44] L. Garcia-Moya, H. Anaya-Sanchez, R. Berlanga-Llavori, Retrieving product features and opinions from customer reviews, IEEE Intelligent Systems 28 (3) (2013) 19–27.

[45] M. Elbattah, M. Roushdy, M. Aref, A. B. M. Salem, Large-scale ontology storage and query using graph database-oriented approach: The case of freebase, in: ICICIS, 2015, pp. 39–43.

[46] A. Ren, X. Du, P. Wang, Ontology-based categorization of web search results using yago, in: CSO, Vol. 1, 2009, pp. 800–804.

[47] E. Cambria, Y. Song, H. Wang, N. Howard, Semantic multidimensional scaling for open-domain sentiment analysis, IEEE Intelligent Systems 29 (2) (2014) 44–51.

[48] G. Psathas, The general inquirer: Useful or not?, Computers and the Humanities 3 (3) (1969) 163–174.

[49] M. Lenzerini, L. Lepore, A. Poggi, Answering metaqueries over hi(owl 2 ql) ontologies, in: IJCAI, IJCAI'16, AAAI Press, 2016, pp. 1174–1180.

[50] A. Esuli, F. Sebastiani, Sentiwordnet: A publicly available lexical resource for opinion mining, in: LREC, 2006, pp. 417–422.

[51] J. Suzuki, H. Isozaki, Sequence and tree kernels with statistical feature mining, in: NIPS, 2006, pp. 1321–1328.

[52] R. M. Carmen Banea, J. Wiebe, A bootstrapping method for building subjectivity lexicons for languages with scarce resources, in: LREC, 2008.

[53] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts, Learning word vectors for sentiment analysis, in: ACL, 2011, pp. 142–150.

[54] C. Lin, Y. He, R. Everson, Sentence subjectivity detection with weakly-supervised learning, in: IJCNLP, 2011, pp. 1153–1161.

[55] C. Banea, R. Mihalcea, J. Wiebe, S. Hassan, Multilingual subjectivity analysis using machine translation, in: EMNLP, 2008, pp. 127–135.

[56] E. Cambria, S. Poria, D. Hazarika, K. Kwok, SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings, in: AAAI, 2018.

[57] E. Cambria, J. Fu, F. Bisio, S. Poria, AffectiveSpace 2: Enabling affective intuition for concept-level sentiment analysis, in: AAAI, 2015, pp. 508–514.

[58] S. Sun, C. Luo, J. Chen, A review of natural language processing techniques for opinion mining systems, Information Fusion 36 (2017) 10–25.

[59] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, CoRR abs/1404.2188.

[60] S. Poria, I. Chaturvedi, E. Cambria, F. Bisio, Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis, in: IJCNN, 2016, pp. 4465–4473.

[61] Y. Mao, G. Lebanon, Isotonic conditional random fields and local sentiment flow, in: B. Schölkopf, J. Platt, T. Hoffman (Eds.), NIPS, MIT Press, 2007, pp. 961–968.

[62] I. Chaturvedi, Y.-S. Ong, I. W. Tsang, R. E. Welsch, E. Cambria, Learning word dependencies in text by means of a deep recurrent belief network, Knowledge-Based Systems 108 (2016) 144–154.

[63] S. Bucak, R. Jin, A. Jain, Multiple kernel learning for visual object recognition: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (7) (2014) 1354–1369.

[64] F. Sha, F. Pereira, Shallow parsing with conditional random fields, in: HLT-NAACL, NAACL '03, 2003, pp. 134–141.

[65] X. Glorot, A. Bordes, Y. Bengio, Domain adaptation for large-scale sentiment classification: A deep learning approach, in: ICML, 2011, pp. 513–520.

[66] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: EMNLP, 2013, p. 1642.

[67] R. Collobert, J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in: ICML, ICML '08, 2008, pp. 160–167.

[68] N. Kalchbrenner, P. Blunsom, Recurrent convolutional neural networks for discourse compositionality, CoRR abs/1306.3584.

[69] Y. Xia, E. Cambria, A. Hussain, H. Zhao, Word polarity disambiguation using bayesian model and opinion-level features, Cognitive Computation 7 (3) (2015) 369–380.

[70] H. Cho, S. Kim, J. Lee, J.-S. Lee, Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews, Knowledge-Based Systems 71 (2014) 61 – 71.

[71] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, E. Cambria, Bayesian network based extreme learning machine for subjectivity detection, Journal of Franklin Institute.

[72] I. Dehdarbehbahani, A. Shakery, H. Faili, Semi-supervised word polarity identification in resource-lean languages, Neural Networks 58 (2014) 50 – 59.

[73] M. D. Molina-González, E. Martínez-Cámara, M.-T. Martín-Valdivia, J. M. Perea-Ortega, Semantic orientation for polarity classification in spanish reviews, Expert Systems with Applications 40 (18) (2013) 7250 – 7257.

[74] R. Mihalcea, C. Banea, J. Wiebe, Learning multilingual subjective language via cross-lingual projections, in: ACL, 2007, pp. 976–983.

[75] R. Mihalcea, C. Banea, J. Wiebe, Multilingual subjectivity and sentiment analysis, in: ACL, Association for Computational Linguistics, 2012, pp. 4–4.

[76] M. Abdul-Mageed, M. Diab, S. Kübler, Samar: Subjectivity and sentiment analysis for arabic social media, Computer Speech and Language 28 (1) (2014) 20 – 37.

[77] P. D. Turney, Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews, in: ACL, ACL '02, Association for Computational Linguistics, 2002, pp. 417–424.

[78] P. Ekman, W. V. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues, Prentice-Hall, Oxford, 1975.

[79] S. Poria, E. Cambria, A. Hussain, G.-B. Huang, Towards an intelligent framework for multimodal affective data analysis, Neural Networks 63 (2015) 104 – 116.

[80] K. Humphrey, G. Underwood, See what i'm saying? expertise and verbalisation in perception and imagery of complex scenes, Cognitive Computation 3 (1) (2011) 64–78.

[81] M. Wollmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, L.-P. Morency, Youtube movie reviews: Sentiment analysis in an audio-visual context, IEEE Intelligent Systems 28 (3) (2013) 46–53.

[82] S. Raaijmakers, K. Truong, T. Wilson, Multimodal subjectivity analysis of multiparty conversation, in: EMNLP, EMNLP '08, Association for Computational Linguistics, 2008, pp. 466–474.

[83] N. Subrahmanya, Y. Shin, Sparse multiple kernel learning for signal processing applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (5) (2010) 788–798.

[84] Z. Zhang, Z.-N. Li, M. Drew, Adamkl: A novel biconvex multiple kernel learning approach, in: ICPR, 2010, pp. 2126–2129.

[85] C. Hinrichs, V. Singh, G. Xu, S. Johnson, Mkl for robust multi-modality ad classification, in: MICCAI, Vol. 5762, 2009, pp. 786–94.

[86] F. Liu, L. Zhou, C. Shen, J. Yin, Multiple kernel learning in the primal for multimodal alzheimers disease classification, IEEE Journal of Biomedical and Health Informatics 18 (3) (2014) 984–990.

[87] S. Nilufar, N. Ray, H. Zhang, Object detection with dog scale-space: A multiple kernel learning approach, IEEE Transactions on Image Processing 21 (8) (2012) 3744–3756.

[88] S. Wang, S. Jiang, Q. Huang, Q. Tian, Multiple kernel learning with high order kernels, ICPR 0 (2010) 2138–2141.

[89] J. Yang, Y. Tian, L.-Y. Duan, T. Huang, W. Gao, Group-sensitive multiple kernel learning for object recognition, IEEE Transactions on Image Processing 21 (5) (2012) 2838–2852.

[90] A. Vahdat, K. Cannons, G. Mori, S. Oh, I. Kim, Compositional models for video event detection: A multiple kernel learning latent variable approach, in: ICCV, 2013, pp. 1185–1192.

[91] X. Xu, I. Tsang, D. Xu, Soft margin multiple kernel learning, IEEE Transactions on Neural Networks and Learning Systems 24 (5) (2013) 749–761.

[92] B. Ni, T. Li, P. Moulin, Beta process multiple kernel learning, in: CVPR, 2014, pp. 963–970.

[93] H. Xia, S. Hoi, R. Jin, P. Zhao, Online multiple kernel similarity learning for visual search, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (3) (2014) 536–549.

[94] A. Wawer, Mining opinion attributes from texts using multiple kernel learning, in: ICDMW, 2011, pp. 123–128.

[95] U. Niaz, B. Merialdo, Fusion methods for multi-modal indexing of web data, in: WIAMIS, 2013, pp. 1–4.

[96] D. Tam, Computation in emotional processing: Quantitative confirmation of proportionality hypothesis for angry unhappy emotional intensity to perceived loss, Cognitive Computation 3 (2) (2011) 394–415.

[97] L. Devillers, L. Vidrascu, L. Lamel, Challenges in real-life emotion annotation and machine learning based detection, Neural Networks 18 (4) (2005) 407 – 422.

[98] C. Akkaya, J. Wiebe, R. Mihalcea, Subjectivity word sense disambiguation, in: EMNLP, 2009, pp. 190–199.

[99] J. Kim, F. Rousseau, M. Vazirgiannis, Convolutional sentence kernel from word embeddings for short text categorization., in: EMNLP, The Association for Computational Linguistics, 2015, pp. 775–780.

[100] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts, Learning word vectors for sentiment analysis, in: ACL, HLT '11, Association for Computational Linguistics, 2011, pp. 142–150.

[101] J. Villena-Román, J. García-Morera, M. García-Cumbreras, E. Martínez-Cámara, M. Martín-Valdivia, L. Ureña-López, Overview of tass 2015, in: TASS Workshop, 2015.

[102] I. Chaturvedi, E. Cambria, D. Vilares, Lyapunov filtering of objectivity for spanish sentiment model, in: IJCNN, 2016, pp. 4474–4481.

[103] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, in: LSM, LSM '11, Association for Computational Linguistics, 2011, pp. 30–38.

[104] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: The good the bad and the omg!, in: ICWSM, AAAI Press, 2011, pp. 538–541.

[105] M. Long, J. Wang, Y. Cao, J. Sun, P. S. Yu, Deep learning of transferable representation for scalable domain adaptation, IEEE Transactions on Knowledge and Data Engineering 28 (8) (2016) 2027–2040.

[106] D. Bollegala, T. Mu, J. Y. Goulermas, Cross-domain sentiment classification using sentiment sensitive embeddings, IEEE Transactions on Knowledge and Data Engineering 28 (2) (2016) 398–410.

[107] Y. Kim, Convolutional neural networks for sentence classification, in: EMNLP, 2014, pp. 1746–1751.

[108] S. Wang, C. Manning, Fast dropout training, in: ICML, 2013, pp. 118–126.

[109] R. Ortega, A. Fonseca, Y. Gutiérrez, A. Montoyo, Improving subjectivity detection using unsupervised subjectivity word sense disambiguation., Procesamiento del Lenguaje Natural 51 (2013) 179–186.

[110] Y. Kim, Convolutional neural networks for sentence classification, CoRR abs/1408.5882.

[111] H. M. Wallach, D. M. Mimno, A. McCallum, Rethinking lda: Why priors matter., in: Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, A. Culotta (Eds.), NIPS, Curran Associates, Inc., 2009, pp. 1973–1981.

[112] D. Vilares, Y. Doval, M. A. Alonso, C. Gómez-Rodríguez, Lys at tass 2014: A prototype for extracting and analysing aspects from spanish tweets, in: TASS workshop, 2014.

[113] R. Mihalcea, A. Garimella, What men say, what women hear: Finding gender-specific meaning shades, IEEE Intelligent Systems 31 (4) (2016) 62–67.

[114] N. Majumder, S. Poria, A. Gelbukh, E. Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intelligent Systems 32 (2) (2017) 74–79.

[115] S. Cavallari, V. Zheng, H. Cai, K. Chang, E. Cambria, Learning community embedding with community detection and node embedding on graphs, in: CIKM, 2017, pp. 377–386.