

Chapter 5

Sentiment Resources: Lexicons and Datasets

Aditya Joshi, Pushpak Bhattacharyya, and Sagar Ahire

Abstract Sentiment lexicons and datasets represent the knowledge base that lies at the foundation of a SA system. In its simplest form, a sentiment lexicon is a repository of words/phrases labelled with sentiment. Similarly, a sentiment-annotated dataset consists of documents (tweets, sentences or longer documents) labelled with one or more sentiment labels. This chapter explores the philosophy, execution and utility of popular sentiment lexicons and datasets. We describe different labelling schemes that may be used. We then provide a detailed description of existing sentiment and emotion lexicons, and the trends underlying research in lexicon generation. This is followed by a survey of sentiment-annotated datasets and the nuances of labelling involved. We then show how lexicons and datasets created for one language can be transferred to a new language. Finally, we place these sentiment resources in the perspective of their classic applications to sentiment analysis.

Keywords Sentiment lexicons • Sentiment datasets • Evaluation • Transfer learning

The previous chapter shows that sentiment analysis (SA) is indeed more challenging than it seems. The next question that arises is, where does the program ‘learn’ the sentiment from? In other words, where does the knowledge required for a SA system come from? This chapter discusses sentiment resources as means to this requirement of knowledge. We refer to words/phrases and documents as ‘textual units’. In sentiment resources, it is these textual units that are annotated with sentiment information.

A. Joshi (✉)
IITB-Monash Research Academy, Mumbai, India
e-mail: adityaj@cse.iitb.ac.in

P. Bhattacharyya • S. Ahire
IIT Bombay, Mumbai, India
e-mail: pb@cse.iitb.ac.in

5.1 Introduction

Sentiment resources, *i.e.*, lexicons and datasets represent the knowledge base of a SA system. Thus, creation of a sentiment lexicon or a dataset is the fundamental requirement of a SA system. In case of a lexicon, it is in the form of simpler units like words and phrases, whereas in case of datasets, it consists of comparatively longer text. There exists a wide spectrum of such resources that can be used for sentiment/emotion analysis. Before we proceed, we reiterate the definition of sentiment and emotion analysis. We refer to sentiment analysis as a positive/negative/neutral classification task, whereas emotion analysis deals with a wider spectrum of emotions such as angry, excited, etc. A discussion on both sentiment and emotion lexicons is imperative to show how different the philosophy behind construction of the two is.

A sentiment resource is a repository of textual units marked with one or more labels representing a sentiment state. This means that there are two driving components of a sentiment resource: (a) the textual unit, and (b) the labels. We discuss the second component, labels in detail in Sect. 5.2.

In case of a sentiment lexicon, the lexical unit may be a word, a phrase or a concept from a general purpose lexicon like WordNet. What constitutes the labels is also important. The set of labels may be purely functional: task-based. For a simple positive-negative classification, it is often sufficient to have a set of positive and negative words. If the goal is a system that gives ‘magnitude’ (‘The movie was horrible’ is more strongly negative than ‘The movie was bad’), then the lexicon needs to capture that information in terms of a magnitude in addition to positive and negative words.

An annotated dataset consists of documents labelled with one or more output labels. As in the case of sentiment lexicons, the two driving components of a sentiment-annotated dataset are: (a) the textual unit, and (b) the labels. For example, a dataset may consist of a set of movie reviews (the textual units) annotated by human annotators as positive or negative (the labels). Datasets often contain additional annotation in order to enrich the quality of annotation. For example, a dataset of restaurant reviews annotated with sentiment may contain additional annotation in the form of restaurant location. Such annotation may facilitate insights such as: which restaurant is the most popular, what are the issues with respect to this outlet of a restaurant that people complain the most about, etc.

5.2 Labels

A set of labels is the pre-determined set of attributes that each textual unit in a sentiment resource will be annotated with. The process of assigning a label to a textual unit is called annotation, and in case the label pertains to sentiment, the process is called *sentiment annotation*. The goal of sentiment annotation is to assign

labels in one out of three schemes: absolute, overlapping and fuzzy. The first two are shown in Liu (2010).

Absolute labelling is when a textual unit is marked as exactly one out of multiple labels. An example of absolute labelling may be positive versus negative – where each document is annotated as either positive or negative. An additional label ‘neutral’ may be added. A fallback label such as ‘ambiguous’/‘unknown’/‘unsure’ may be introduced. Numeric schemes that allow labels to range between, say, +5 to –5 also fall under this method of labelling.

Labels can be overlapping as well. A typical example of this is emotion labels. Emotions are more complex than sentiment, because there can be more than one emotion at a time. For example, the sentence, “Was happy to bump into my friend at the airport this afternoon.” would be labelled as positive as a sentiment-annotated sentence. On the other hand, an emotion annotation would require two labels to be assigned to this text: happiness and surprise. Emotions can, in fact, be thought of arising from a combination of emotions, and their magnitudes. This means that while positive-negative are mutually exclusive, emotions need not be. In such cases, each one of them must be viewed as a Boolean attribute. This means that the word ‘amazed’ will be marked as ‘happy: yes, surprised: yes’ for an emotion lexicon, whereas the same ‘amazed’ will be marked as ‘positive’ for a sentiment lexicon. By definition, a positive word implies that it is not negative.

Finally, the third scheme of labelling is fuzzy: where a distribution over different labels is assigned to a textual unit. Consider the case where we assign a distribution over ‘positive/negative’ as a label. Such a distribution implies likelihood of the textual unit to belong to the given label. For example, a word with ‘positive:0.8, negative:0.2’ means that the word tends to occur more frequently in a positive sense – however, it is not completely positive and it may still be used in the negative sense to an extent.

Several linguistic studies have explored what constitutes basic labels for a sentiment resource. In the next subsections, we look at three strategies.

5.2.1 *Stand-Alone Labels*

A sentiment resource may use two labels: positive or negative. The granularity can be increased to strongly positive, moderately positive and so on. A positive unit represents a desirable state, whereas a negative unit represents an undesirable state (Liu 2010). Emotion labels are more nuanced. Basic emotions are a list of emotions that are fundamental to human experience. Whether or not there are any basic emotions at all, and whether it is worthwhile to discover these basic emotions has been a matter of disagreement. Ortony and Turner (1990) state that the basic emotion approach (*i.e.*, stating that there are basic emotions and other emotions evolve from them) is flawed, while Ekman (1992) supports the basic emotion theory. Several basic emotions have been suggested. Ekman suggests six basic emotions:

anger, disgust, fear, sadness, happiness and surprise. Plutchik has listed eight basic emotions: six from Ekman's list in addition to anticipation and trust (Plutchik 1980).

5.2.2 *Dimensions*

Sentiment has been defined by Liu (2010) as a 5-tuple: <sentiment-holder, sentiment-target, sentiment-target-aspect, sentiment, sentiment-time>. This means that sentiment in a textual unit can be captured accurately only if information along the five dimensions is obtained. Similarly, emotions can also be looked at in the form of two dimensions: valence and arousal (Mehrabian and Russell 1974). Valence indicates whether an emotion is pleasant or unpleasant. Arousal indicates the magnitude of an emotion. Happy and excited are two forms of a pleasant emotion, but they differ along the arousal axis. Excitement indicates a state where a person is happy, but aroused to a great degree. On the other hand, calm and content, while still being pleasant emotions, represent a deactivated state. Corresponding emotions in the left quadrant (that indicates unpleasant emotions) are sad, stressed, bored and fatigued. In such a case, overlapping labelling must be used. A resource annotated using dimensional structure will assign a value per dimension for each textual unit.

5.2.3 *Structures*

Plutchik wheel of emotions (Plutchik 1982) is a popular structure that represents basic emotions, and emotions that arise as a combination of these emotions. It combines the notion of basic emotions, along with arousal as seen in case of emotion dimensions. The basic emotions according to Plutchik's wheel are joy, trust, fear, surprise, anticipation, sadness, disgust, anger and anticipation. The basic emotions are arranged in a circular manner to indicate antonymy. The opposite of 'joy' is placed diametrically opposite to it: 'sadness'. Similarly, 'anticipation' lies diametrically opposite to 'surprise'. Each 'petal' of the wheel indicates the arousal of the emotion. The emotion 'joy' has 'serenity' above it and 'ecstasy' below it. These emotions indicate a deactivated and activated state of arousal respectively. Similarly, an aroused state of 'anger' becomes 'rage'. Thus, the eight emotions in the central circle are the aroused forms of the basic emotions. These are: rage, loathing, grief, amazement, terror, admiration, ecstasy and vigilance. The wheel also allows combination of emotions to create more nuanced emotions. A resource annotated using a structure such as the Plutchik wheel of emotions will place every textual unit in the space represented by the structure.

5.3 Lexicons

We now discuss sentiment lexicons: we describe them individually first, and then show trends in lexicon generation. Words/phrases have two kinds of sentiment, as given in Liu (2010): absolute and relative. Absolute sentiment means that the sentiment remains the same, given the right word/phrase and meaning. For example, the word ‘beautiful’ is a positive word. Relative sentiment means that the sentiment changes depending on the context. For example, the word ‘increased’ or ‘fuelled’ has a positive/negative sentiment based on what the object of the word is. There exists a third category of sentiment: implicit sentiment. Implicit sentiment is different from absolute sentiment. Implicit sentiment is the sentiment that is commonly invoked in the mind of a reader when he/she reads that word/phrase. Consider the example ‘amusement parks’. A reader typically experiences positive sentiment on reading this word. Similarly, the phrase ‘waking up in the middle of the night’ does involve an implicit negative sentiment.

Currently, most sentiment lexicons limit themselves to absolute sentiment words. Extraction of implicit sentiment in phrases forms a different branch of work. However, there exist word association lexicons that capture implied sentiment in words (Mohammad and Turney 2010). We stick to this definition as well, and discuss sentiment and emotion lexicons that capture absolute sentiment.

5.3.1 *Sentiment Lexicons*

Early development of sentiment lexicons focused on creation of sentiment dictionaries. Stone et al. (1966) present a lexicon called ‘General Inquirer’ that has been widely used for sentiment analysis. Finn (2011) present a lexicon called AFINN. Like General Inquirer, it is also a manually generated lexicon. To show the general methodology underlying sentiment lexicons, we describe some popular sentiment lexicons in the forthcoming subsections.

5.3.1.1 SentiWordNet

SentiWordNet, described first by Esuli and Sebastiani (2006), is a sentiment lexicon which augments WordNet (Miller 1995) with sentiment information. The labelling is fuzzy, and is done by adding three sentiment scores to each synset in the WordNet as follows. Every synsets has three scores:

1. Pos(s): The positive score of synsets
2. Neg(s): The negative score of synsets
3. Obj(s): The objective score of synsets

Thus, in SentiWordNet, sentiment is associated with the meaning of a word rather than the word itself. This representation allows a word to have multiple sentiments corresponding to each meaning. Because there are three scores, each meaning in itself can be both positive and negative, or neither positive nor negative.

The process of SentiWordNet creation is an expansion of the approach used for the three-class sentiment classification to handle graded sentiment values. The algorithm to create SentiWordNet can be summarized as:

1. *Selection of Seed Set:* A seed set L_p and L_n consisting of ‘paradigmatic’ positive and negative synsets respectively was created. Each synset was represented using the TDS. This representation converted words in the synset, its WordNet definition and the sample phrases together with explicit labels for negation into vectors.
2. *Creation of Training Set:* This seed set was expanded for k iterations using the following relations of WordNet: Direct antonymy, Similarity, Derived from, Pertains to, Attribute and Also see. These were the relations hypothesized to preserve or invert the associated sentiment. After k iterations of expansion, this gave rise to the sets Tr_p^k and Tr_n^k . The objective set $L_o = Tr_o^k$ was assumed to consist of all the synsets that did not belong to Tr_p^k or Tr_n^k .
3. *Creation of Classifiers:* A classifier can be defined as a combination of a learning algorithm and a training set. In addition to the two choices of learning algorithms (SVM and Rocchio), four different training sets were constructed with the number of iterations of expansion $k = 0, 2, 4, 6$. The size of the training set increased substantially with an increase in k . As a result, low values of k yielded classifiers with low recall but high precision, while higher k led to high recall but low precision. As a result there were 8 ternary classifiers in total due to all combinations of the 2 learners and 4 training sets. Each ternary classifier was made up of two binary classifiers, positive vs. not positive and negative vs. not negative.
4. *Synset Scoring:* Each synset from the WordNet was vectorized and given to the committee of ternary classifiers as test input. Depending upon the output of the classifiers, each synset was assigned sentiment scores by dividing the count of classifiers that give a label by the total number of classifiers (8).

5.3.1.2 SO-CAL

Sentiment Orientation CALculator (SO-CAL) system (Brooke et al. 2009) is based on a manually constructed low-coverage resource made up of raw words. Unlike SentiWordNet, there is no sense information associated with a word. SO-CAL uses as its basis a lexical sentiment resource consisting of about 5000 words. (In comparison, SentiWordNet has over 38,000 polar words and several other strictly objective words.) Each word in SO-CAL has a sentiment label which is an integer

in $[-5, +5]$ apart from 0 as objective words are simply excluded. The strengths of SO-CAL lie in its accuracy, as it is manually annotated, and the use of detailed features that handle sentiment in various cases in ways conforming to linguistic phenomena.

SO-CAL uses several ‘features’ to model different word categories and the effects they have on sentiment. In addition, a few special features operate outside the scope of the lexicon in order to affect the sentiment on the document level. These are some of the features of SO-CAL:

1. *Adjectives*: A manual dictionary of adjectives was created by manually tagging all adjectives in a 500-document multidomain review corpus, and the terms from the General Inquirer dictionary were annotated added to the list thus obtained.
2. *Nouns, Verbs and Adverbs*: SO-CAL also extended the approach used for adjectives to nouns and verbs. As a result, 1142 nouns and 903 verbs were added to the sentiment lexicon. Adverbs were added by simply adding the -ly suffix to adjectives and then manually altering words whose sentiment was not preserved, such as essentially. In addition multi-word expressions were also added, leading to an addition to 152 multiwords in the lexicon. Thus, while the adjective ‘funny’ has a sentiment of +2, the multiword ‘act funny’ has a sentiment of -1.
3. *Intensifiers and Downtoners*: An Intensifier is a word which increases the intensity of the phrase to which it is applied, while a Downtoner is a word which decreases the intensity of the phrase to which it is applied. For instance the word ‘extraordinarily’ in the phrase ‘extraordinarily good’ is an intensifier while the word somewhat in the phrase ‘somewhat nice’ is a downtoner.

5.3.1.3 Sentiment Treebank & Associated Lexicon

This Treebank was introduced in Socher et al. (2013). In order to do create the Treebank, the work also came up with a lexicon called the Sentiment Treebank, which is a lexicon consisting of partial parse trees annotated with sentiment.

The lexicon was created as follows. A movie review corpus was obtained from www.rottentomatoes.com, consisting of 10,662 sentences. Each sentence was parsed using the Stanford Parser. This gave a parse tree for each sentence. The parse trees were split into phrases, i.e., each parse tree was split into its components, each of which was then output as a phrase. This gave rise to 215,154 phrases. Each of these phrases was tagged for sentiment using Amazon’s Mechanical Turk’s interface. The selection of labels is also described in the original paper. Initially, the granularity of the sentiment values was 25, i.e., 25 possible values could be given for the sentiment, but it was observed from the data from the Mechanical Turks experiment that most responses contained any one of only 5 values. These 5 values were then called ‘very positive’, ‘positive’, ‘neutral’, ‘negative’ and ‘very negative’.

Table 5.1 Summary of sentiment lexicons

	Approach	Lexical unit	Labels	Observation
SO-CAL	Manual	Word	Integer in $[-5, +5]$	Performance can be improved by incorporating linguistic features even with low coverage
SentiWordNet	Automatic	WordNet Synset	3 fractional values Pos, Neg, Obj in $[0, 1]$	WordNet captures senses. Different senses may have different sentiment.
Sentiment Treebank	Manual, Crowdsourced	Phrase	5 labels ranging from “very negative” to “very positive”	Crowdsourcing can be beneficial. Tune labels according to the task.
Macquaire semantic orientation lexicon	Semi-supervised	Words	Positive/ negative	Using links in a thesaurus to discover new words.

5.3.1.4 Summary

Table 5.1 summarizes sentiment lexicons described above, and in addition, also mentions some other sentiment lexicons. We compare along four parameters: the approach used for creation, lexical units, labels and some observations. Mohammad et al. (2009) present Macquaire semantic orientation lexicon. This is a sentiment lexicon that contains 76,400 terms, marked as positive or negative. In terms of obtaining manual annotations, Louviere (1991) present an approach called the MaxDiff approach. In this case, instead of obtaining annotations for one word at a time, an annotator is shown multiple words and asked to identify the least positive and most positive word among them.

5.3.2 Emotion Lexicons

We now describe emotion lexicons. They have been described in this separate subsection so as to highlight challenges and the approaches specific to emotion lexicon generation.

5.3.2.1 LIWC

Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2001) is a popular manually created lexicon. The lexicon consists of 4500 words and word stems

(An example word stem is *happ** which covers adjectival and adverbial forms of the word) arranged in four categories. The four categories of words in LIWC are: Linguistic processes (pronouns, prepositions, conjunctions, etc.), Speaking processes (Interjections, Fillers, etc.), personal concerns (words related to work, home, etc.) and psychological processes. The words in the psychological processes category deal with affect and opinion, and are further classified into cognitive and affective processes. Cognitive processes include words indicating certainty ('definitely'), possibility ('likely') and inhibition ('prevention'), etc. Affective processes include words with positive/negative emotion, words expressing anxiety, anger, sadness. LIWC 2001 has 713 cognitive and 915 affective processes words. LIWC was manually created by three linguistic experts in two steps:

- (a) Define category scales: The judges determined categories and decided how they can be grouped into a hierarchy
- (b) Manual population: The categories were manually populated with words. For each word, three judges manually evaluated whether or not a word should be placed in a category. In addition, they also considered if a word can be moved higher up in the hierarchy.

LIWC now exists in multiple languages, and has been widely used by several applications for analysis of topic as well as sentiment/emotion.

5.3.2.2 ANEW

Affective norms for English words (ANEW) (Bradley and Lang 1999) is a dictionary of around 1000 words where each word is indicated with a three-tuple representation: pleasure, arousal and activation. Pleasure indicates the valence of a word, arousal the intensity while activation indicates whether the emotion expressed in the word is in control or not. Consider the example word 'afraid'. This word is indicated by the tuple (negative, 3, not) indicating that it is a negative emotion, with an arousal of 3, and is a deactivated emotion. ANEW was manually created by 25 annotators separately. Each annotation experiment was conducted in runs of 100–150 words. Annotators are given a sheet called ScanSAM sheet. Each annotator marks values of S, A and M for word. The annotators perform the annotation separately.

5.3.2.3 Emo-Lexicon

Emo-Lexicon (Mohammad and Turney 2013) is a lexicon of 14,000 terms created using crowd-sourcing portals like Amazon Mechanical Turk. Association with positive and negative valence as well as with the eight Plutchik emotions is also available. Although it is manually created, the lexicon is larger than other emotion lexicons – a clear indication that crowdsourcing is indeed a powerful mechanism for large-scale creation of emotion lexicon. However, because the task of lexicon

creation has been opened up to the ‘crowd’, quality control is a key challenge. To mitigate this, the lexicon is created with additional drivers, as follows:

1. A list of words is created from a thesaurus.
2. When an annotator annotates a word with emotion, he/she must first ascertain the sense of the word. The target word is displayed along with four words. The annotator must select one that is closest to the target word.
3. Only if the annotator was able to correctly determine the sense of the word is his/her annotation for emotion label obtained.

5.3.2.4 WordNet-Affect

WordNet-Affect (Strapparava and Valitutti 2004) like SentiWordNet, is a resource that annotates senses in WordNet with emotions. WordNet Affect was created using a semi-supervised method. It consists of 2874 synsets annotated with affective labels (called a-labels). WordNet-Affect was created as follows:

1. A set of core synsets is created. These are synsets whose emotion has been manually labelled in the form of a-labels.
2. These labels are projected to other synsets using WordNet relations.
3. The a-labels are then manually evaluated and corrected, wherever necessary.

5.3.2.5 Chinese Emotion Lexicon

A Chinese emotion lexicon (Xu et al. 2010) was created using a semi-supervised approach, in absence of a graphical structure such as WordNet. There are two steps of creation:

1. Select a core set of labelled words.
2. Expand these words using a similarity matrix. Iterate until convergence.

The similarity matrix takes three kinds of similarity into account:

1. *Syntagmatic similarity*: This includes co-occurrence of two words in a large text corpus.
2. *Paradigmatic similarity*: This includes relations between two words in a semantic dictionary.
3. *Linguistic peculiarity*: This involves syllable overlap, possibly to cover different forms of the same word.

5.3.2.6 SenticNet

SenticNet (The most recent version, being SenticNet 4) by Cambria et al. (2016) is a rich graphical repository of concepts. The resource aims to capture semantic,

Table 5.2 Summary of emotion lexicons

	Approach	Labels	Observation
LIWC	Manual	Hierarchy of categories	Decide hierarchy of categories; have judges interacting with each other
ANEW & ANEW for Spanish	Manual	Valence, arousal, dominance	ScanSAM lists; have a set of annotators annotating in parallel
Emo-Lex	Manual	Eight emotions, two valence categories	Use crowd-sourcing. Attention to quality control.
WordNet affect	Semi-supervised	Affective labels	Annotate a seed set. Expand using WordNet relations.
Chinese emotion lexicon	Semi-supervised	Five emotions	Annotate a seed set. Expand using similarity matrices
NRC Hashtag emotion lexicon	Automatic	Eight emoticons	Use hashtag based supervision of tweets
SenticNet 4	Semi-supervised	A larger structure	Semi-supervised graphical structure, created using techniques such as agglomerative clustering

and sentic properties of words and phrases. The sentic properties are related to connotations of words. A detailed discussion of SenticNet forms a forthcoming chapter of this book.

5.3.2.7 Summary

Table 5.2 shows a summary of emotion lexicons discussed in this section. We observe that manual approaches dominate emotion lexicon creation. Key issues in manual emotion annotation are: ascertaining the quality of the labels, deciding hierarchies if any. Additional useful lexicons are available at: <http://www.saifmohammad.com/WebPages/lexicons.html>. On the other hand, automatic emotion annotation is mostly semi-supervised. To expand a seed set, structures like WordNet may be used, or similarity matrices constructed from large corpora can be employed. Mohammad (2012) present a hashtag emotion lexicon that consists of 16,000+ unigrams annotated with eight emotions. The lexicon is created using emotion-denoting hashtags present in tweets. Mohammad and Turney (2010) is also an emotion lexicon created using a crowdsourcing platform.

5.4 Sentiment-Annotated Datasets

This section describes sentiment-annotated datasets, and is organized as follows. We first describe sources of data, mechanisms of annotation, and then provide a list of some sentiment-annotated datasets.

5.4.1 Sources of Data

The first step is to obtain raw data. The following are candidate sources of raw data:

1. **Social networking websites** like twitter are a rich source of data for sentiment analysis applications. For example, Twitter API (Makice 2009) is a publicly available API that allows you to download tweets based on a lot of interesting search criteria such as keyword-based-search, download-timelines, download-tweet-threads, etc.
2. **Competitions such as SemEval** have been regularly conducting Sentiment analysis related tasks. These competitions release a training dataset followed by a test dataset. These datasets can be used as benchmark datasets.
3. **Discussion forums** are portals where users discuss topics, often in the context of a central theme or an initial question. These discussion forums often arrange posts in a thread-like manner. This allows discourse nature to sentiment. However, this also introduces an additional challenge. A reply to a post could mean one out of three possibilities: (a) The reply is an opinion with respect to the post, offering an agreement or disagreement (example: Well-written post), (b) The reply is an opinion towards the author of the post (example: Why do you always post hateful things?), or (c) The reply is an opinion towards the topics being discussed in the post. (Example: You said that the situation is bad. But do you think that....). Reddit threads have been used as opinion datasets in several past works.
4. **Review websites:** Amazon and other review websites have reviews on different domains. Each kind of reviews has unique challenges of its own. In case of movie reviews, the review often has a portion describing ‘what’ the movie is about. It is possible to create subjective extracts before using them as done by Mukherjee and Bhattacharyya (2012). In case of product reviews, the review often contains sentiment towards different ‘aspects’. (‘Aspects’ of a cell phone are battery, weight, OS, etc.).
5. **Blogs** are often long text describing an opinion with respect to a topic. They can also be crawled and annotated to create a sentiment dataset. Blogs tend to be structured narratives analyzing the topic. They may not always contain the same sentiment throughout but can be useful sources of data that looks at different aspects of the given topic.

5.4.2 Obtaining Labels

Once raw data has been obtained, the second step is to label this data. There are different approaches that can be used for obtaining labels for a dataset:

1. **Manual labelling:** Several datasets have been created by human annotators. The labelling can be done through crowd-sourcing applications like Amazon

Mechanical Turk. They allow obtaining large volumes of annotations by employing the ‘power of the crowds’ (Paolacci et al. 2010). To control the quality of annotation, one way is to use a seed set of gold labels. Human annotators within the controlled setup of the experiment create a set of gold labels. If a crowd-sourced annotator (known as ‘worker’ in the crowd-sourcing parlance) gets a sufficient number of gold labels right, only then is he/she permitted to perform the task of annotation.

2. **Distant supervision:** Distant supervision refers to the situation where the label or the supervision is obtained without an annotator – hence the word ‘distant’. One way to do so is to use annotation provided by the writer themselves. However, the question of reliability arises because not every data unit has been manually verified by a human annotator. This has to be validated using the approach used to obtain distant supervision. Consider the example of Amazon reviews. Each review is often accompanied by star ratings. These star ratings can be used as labels provided by the writer. Since these ratings are out of 5, a review with 1 star is likely to be strongly negative, whereas a review with 5 stars is likely to be strongly positive. To improve the quality of the dataset obtained, Pang and Lee (2005) consider reviews that are definitely positive and definitely negative – i.e. reviews with 5 and 1 stars respectively.

Another technique to obtain distant supervision is the use of hashtags. Twitter provides a reverse index mechanism in the form of hashtags. An example tweet is ‘Just finished writing a 20 page long assignment. #Engineering #Boring’. ‘#Engineering’ and ‘#Boring’ as hashtags – since they are phrases preceded by a hashtag symbol. Note that a hashtag is created by the author of the tweet and hence, can be anything – topical (i.e. identifying what the tweet is about. Engineering, in this case) or emotion-related (i.e. expressing an opinion through a hashtag. In this case, the author of the tweet is bored). Purver and Battersby (2012) emotion-related hashtags to obtain a set of tweets containing emotion-related hashtags. Thus, hashtags such as ‘#happy’, ‘#sad’, etc. are used to download tweets using the Twitter API. The tweets are then labelled as ‘#happy’, ‘#sad’, etc. Since hashtags are user-created, they can be more nuanced than this. For example, consider the hypothetical tweet: ‘Meeting my ex-girlfriend after about three years. #happy #not’. The last hashtag ‘#not’ inverts sentiment expressed by the preceding hashtag ‘#happy’. This unique construct ‘#not’ or ‘#notserious’ or ‘#justkidding’/‘#jk’ is popular in tweets and must be handled properly when hashtag-based supervision is used to create a dataset.

5.4.3 Popular Sentiment-Annotated Datasets

We now discuss some popular sentiment-annotated datasets. We divide them into two categories: sentence-level annotation, discourse-level annotation. The latter points to text longer than a sentence. While tweets may contain more than a sentence, we group them under sentence-level annotation because of limited length of tweets.

Sentence-Level Annotated Datasets

Niek Sanders released a dataset at <http://www.sananalytics.com/lab/twitter-sentiment/>. It consists of 5513 manually labelled tweets, classified as per four topics.

SemEval is a competition that is run for specific tasks. Sentiment analysis and related tasks have featured since 2013 (Nakov et al. 2013; Rosenthal et al. 2014, 2015). The datasets for these tasks are released online, and can be useful for sentiment applications. SemEval 2013 dataset is at: <http://www.cs.york.ac.uk/semEval-2013/semEval2013.tgz> SemEval 2014 dataset is at: <http://alt.qcri.org/semEval2014/task9/> SemEval 2015 dataset is at: <http://alt.qcri.org/semEval2015/task10/index.php?id=subtask-readme>

Darmstadt corpus consists of consumer reviews annotated at sentence and expression level. The dataset is available at: <https://www.ukp.tu-darmstadt.de/data/sentiment-analysis/darmstadt-service-review-corpus/> Sentence annotated polarity dataset from Pang et al. (2002) is also available at: <https://www.cs.cornell.edu/people/pabo/movie-review-data/> Sentiment140 (Go et al. 2009) is a corpus made available by Stanford at <http://help.sentiment140.com/for-students>. The dataset is of tweets and contains additional information such as timestamp, author, tweet id, etc.

Deng et al. (2013) released a *goodFor/badFor corpus* that is available at: <http://mpqa.cs.pitt.edu/corpora/gfbf/>. *goodFor/badFor* indicates positive/negative sentiment respectively. This corpus uses a five-tuple representation for opinion annotation. Consider this example sentence from their user manual: ‘The smell stifled his hunger.’ This sentence is marked as: ‘span: stifled, polarity: badFor, agent: the smell, object: his hunger’.

Discourse-Level Annotated Datasets

Many *movie review* datasets and lexicons are available at: <https://www.cs.cornell.edu/people/pabo/movie-review-data/>. These datasets include: sentiment annotated datasets, subjectivity annotated datasets, and sentiment scale datasets. These have been released in Pang and Lee (2004, 2005), and widely used.

A *Congressional speech dataset* (Thomas et al. 2006) annotated with opinion is available at: <http://www.cs.cornell.edu/home/llee/data/convote.html> The labels indicate whether the speaker supported or opposed a legislation that he/she was talking about.

A corpus consisting of *Amazon reviews* from different domains such as electronics, movies, etc. is available at: <https://snap.stanford.edu/data/web-Amazon.html> (McAuley and Leskovec 2013). This dataset spans a period of 18 years, and contains information such as: product title, author name, star rating, helpful votes, etc.

The Political Debate Corpus by Somasundaran and Wiebe (2009) is a dataset of political debates that is arranged based on different topics. It is available here: http://mpqa.cs.pitt.edu/corpora/product_debates/.

MPQA Opinion Corpus (Wiebe et al. 2005) is a popular dataset that consists of news articles from different sources. Version 2.0 of the corpus is nearly 15,000 sentences. The sentences are annotated with topics and labels. The topics are from different countries around the world. This corpus is available at http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/.

5.5 Bridging the Language Gap

Creation of a sentiment lexicon or a labelled dataset is a time/effort-intensive task. Since English is the dominant language in which SA research has been carried out, it is only natural that many other languages have tried to leverage on resources developed for English by adapting and/or reusing them. Cross-lingual SA refers to use of systems and resources developed for one language to perform SA of another. The first language (where the resources/lexicons/systems have been developed) is called the source language, while the second language (where a new system/resource/lexicon needs to be deployed) is called the target language. The basis of cross-lingual SA is availability of a lexicon or an annotated dataset in the source language. It must be noted that several nuanced methodologies to perform cross-lingual SA exist, but have been left out due to the scope of this chapter. We focus on cross-lingual sentiment resources.

The fundamental requirement is a mapping between the two languages. Let us consider what happens in case we wish to map a lexicon in language X to language Y. For a lexicon, this mapping can be in the form of a parallel dictionary where words of one language are mapped to another. ANEW For Spanish (Redondo et al. 2007) describes the generation of a lexicon called ANEW. Originally created for English words, its parallel Spanish version is created by translating words from English to Spanish, and then manually validating them. It can also be in the form of linked WordNets, in case the lexicons involve concepts like synsets. Hindi SentiWordNet (Joshi et al. 2010) map synsets in English to Hindi using a WordNet linking, and generate a Hindi SentiWordNet from its English variant. Mahyoub et al. (2014) describe a technique to create a sentiment lexicon for Arabic. Based on a seed set of positive and negative words, and Arabic WordNet, they present an expansion algorithm to create a lexicon. The algorithm uses WordNet relations in order to propagate sentiment labels to new words/synsets. The WordNet relations they use are divided into two categories: the ones that preserve the sentiment orientation, and the ones that invert the sentiment orientation.

How is this process of mapping words in one language to another any different for datasets? In case a machine translation (MT) system is available, this task is simple. A dataset in the source language can be translated to the target language. This is a common strategy that has been employed (Mihalcea et al. 2007; Duh

et al. 2011). It follows that translation may introduce additional errors into the system, thus causing a degradation in the quality of the dataset. This is particularly applicable to translation of sentiment-bearing idioms. Salameh et al. (2015) perform their experiments for Arabic where a MT system is used to translate documents, following which sentiment analysis is performed. An interesting observation that the authors make is that although MT may result in a poor translation making it difficult for humans to identify sentiment, a classifier performs reasonably well. However, MT systems may not exist for all language pairs. Balamurali et al. (2012) suggest a naive replacement for a MT system. To translate a corpus from Hindi to Marathi (and vice versa), they obtain sense annotations for words in the dataset. Then, they use a WordNet linking to transfer the word from the source language to the target language.

An immediate question that arises is the hypothesis at the bottom of all cross-lingual approaches: sentiment is retained across languages. This means that if a word has a sentiment s in the source language, the translated word in target language (with appropriate sense recorded) also has sentiment s . How fair is the hypothesis that words in different languages bear the same emotion? This can be seen from linear correlations between ratings for the three affective dimensions, as was done for ANEW for Spanish. ANEW for Spanish (Redondo et al. 2007), as described above, was a lexicon created using ANEW in English. The correlation values for valence, arousal and dominance are 0.916, 0.746 and 0.720 respectively. This means that a positive English word is very likely to be a positive Spanish word. The arousal and dominance values remain the same to a lower extent.

Thus, we have two options now. The first option is cross-lingual SA: use resources generated for the source language and map it to the target language. The second option is in-language SA: create resources for the target language on its own. Balamurali et al. (2013) weighs in-language SA against cross-lingual SA based on Machine Translation. The authors show for English, German, French and Russian that in-language SA does consistently better than cross-lingual SA relying on translation alone.

Cross-lingual SA also benefits from additional corpora in target language:

1. *Unlabeled corpus in target language*: This type of corpus is used in different approaches, the most noteworthy being the co-training-based approach. Wan (2009). The authors assume that a labelled corpus in the source language, unlabeled corpus in target language and a MT system to translate back and forth between the two languages are available.
2. *Labelled corpus in target language*: The size of this dataset is assumed to be much smaller than the training set.
3. *Pseudo-parallel data*: Lu et al. (2011) describe use of pseudo-parallel data for their experiments. Pseudo-parallel data is the set of sentences in the source language that are translated to the target language and used as an additional polarity-labelled data set. This allows the classifier to be trained on a larger number of samples.

5.6 Applications of Sentiment Resources

In the preceding sections, we described sentiment resources in terms of labels, annotation techniques and approaches to creation. We will now see how a sentiment resource (either a lexicon or a dataset) can be used.

A lexicon is useful as a knowledge base for a rule-based SA system. A rule-based SA system takes a textual unit as input, applies a set of pre-determined rules, and produces a prediction. Joshi et al. (2011) present C-Feel-It, a rule-based SA system for tweets. The workflow is as follows:

1. A user types a keyword. Tweets containing the keyword are downloaded using the Twitter API
2. The tweets are pre-processed to correct extended words (e.g. ‘happyyyyy’ is replaced with two occurrences of happy. Two, because the extended form of the word ‘happy’ has a magnified sentiment)
3. The words in a tweet are looked up individually in four *lexical resources*. The sentiment label of a tweet is calculated as a sum of positive and negative words – with rules applied for conjunctions and negation. In case of negation, the sentiment of words within a window is inverted. In case of conjunctions such as ‘but’, the latter part of a tweet is considered.
4. The resultant prediction of a tweet is a weighted sum of prediction made by the four lexical resources. The weights are determined experimentally by considering how well the resources perform on an already labelled dataset of tweets.

The above approach is a common framework for rule-based SA systems. Levallois (2013) also use lexicons and a set of rules to perform sentiment analysis of tweets. The goal, as stated by the authors, is to design it as ‘fast and scalable’. LIWC provides a tool which also uses the lexicon, applies a set of rules to generate a prediction. Typically, systems that use SA as a sub-module of a larger application can benefit greatly from a lexicon and simple hand-crafted rules.

Lexicons have also been used in topic models (Lin and He 2009) to set priors on the word-topic distributions. A topic model takes as input a dataset (labelled or unlabeled) and generates clusters of words called topics, such that a word may belong to more than one topic. A topic model based on LDA (Blei et al. 2003) samples a latent variable called topic, for every word occurrence in a document. This results in two types of distributions over an unlabeled dataset: topic-document distributions (the probability of seeing this topic in this document, given the words and the topic-word assignments), and word-topic distributions (the probability of seeing this word belonging to the topic in the entire corpus, given the words and the topic-word assignments). The word-topic distribution is a multinomial with a Dirichlet prior. Sentiment lexicons have been commonly used as Dirichlet Hyperparameters for the word-topic distribution. Consider the following example. In a typical scenario, all words have symmetric priors over the topics. This means that all words are equally likely to belong to a certain topic. However, if we wish

to have ‘sentiment coherence’ in topics, then, setting Dirichlet Hyperparameters appropriately can adjust priors on topic. Let us assume that we wish to have the first half of topics to represent ‘positive’ topics, and second half of topics to represent ‘negative’ topics. A ‘positive’ topic here means a topic with positive words corresponding to a concept. More complex topic models which model additional latent variables (such as sentiment or switch variables) also use lexicons to set priors (Mukherjee and Bhattacharyya 2012). Lexicons have also been used to train deep learning-based neural networks (Socher et al. 2013). A combination of datasets and lexicons has also been used. Tao et al. (2009) propose a three-pronged factorization method for sentiment classification. They factor in information from sentiment lexicons (in the form of word level polarities), unlabeled datasets (in the form of word co-occurrence) and labelled datasets (to set up the correspondences). Lexicons can also be used to determine values of frequency-based features in a statistical classification system. Kiritchenko et al. (2014) use features derived from a lexicon such as: number of tokens with non-zero sentiment, total and maximal score of sentiment, etc. This work also presents a set of ablation tests to identify value of individual sets of features. When the lexicon-based features are removed from the complete set, the maximum degradation is observed. Such lexicon-based features have been used for related tasks such as sentiment annotation complexity prediction (Joshi et al. 2014), thwarting detection (Ramteke et al. 2013) and sarcasm detection (Joshi et al. 2015).

Let us now look at how sentiment-labelled datasets can be used, especially in machine learning (ML)-based classification systems. ML-based systems model sentiment analysis as a classification problem. A classification model predicts the label of a document as one among different labels. This model is learnt using a labelled dataset as follows. A document is converted to a feature vector. The most common form of a feature vector of a document is the unigram representation with the length equal to the vocabulary size. The vocabulary is the set of unique words in the labelled dataset. A Boolean or numeric feature vector of length equal to the vocabulary size is constructed for each document where the value is set for the words present in the document. The goal of the model is to minimize error on training documents, with appropriate regularization for variance in unseen documents. The labelled documents serve as a building block for a ML-based system. While the unigram representation is common, several features such as word sense based features (Balamurali et al. 2011), qualitative features such as POS sequences (Pang et al. 2002), have been used as features for ML-based systems. The annotated datasets form the basis for creation of feature vectors with the documents acting as observed instances. Melville et al. (2009) combine knowledge from lexicons and labelled datasets in a unique manner. Sentiment lexicon forms the background knowledge about words while labelled datasets provide a domain-specific view of the task, in a typical text classification scenario.

5.7 Conclusion

This chapter described sentiment resources: specifically, sentiment lexicons and sentiment-annotated datasets. Our focus was on the philosophy and trends in the generation and use of sentiment lexicons and datasets. We described creation of several popular sentiment and emotion lexicons. We then discussed different strategies to create annotated datasets, and also presented a list of available datasets. Finally, we add two critical points in the context of sentiment resources: how a resource in one language can be mapped to another, and how these resources are actually deployed in a SA system. The diversity in goals, approaches and uses of sentiment resources highlights the value of good quality sentiment resources to sentiment analysis.

References

- Balamurali, A.R., Aditya Joshi, and Pushpak Bhattacharyya. 2011. Harnessing wordnet senses for supervised sentiment classification. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics.
- . 2012. Cross-lingual sentiment analysis for Indian languages using linked WordNets. In: *COLING*.
- Balamurali, A.R., Mitesh M. Khapra, and Pushpak Bhattacharyya. 2013. Lost in translation: Viability of machine translation for cross language sentiment analysis. In *Computational linguistics and intelligent text processing*. Berlin/Heidelberg: Springer.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3: 993–1022.
- Bradley, M.M., and P.J. Lang. 1999. *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical report C-1. The Center for Research in Psychophysiology, University of Florida.
- Brooke, Julian, Milan Tofiloski, and Maite Taboada. 2009. Cross-linguistic sentiment analysis: From English to Spanish. In: *RANLP*.
- Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. 2016. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *The 26th International conference on computational linguistics (COLING)*, Osaka, 2666–2677.
- Deng, Lingjia, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/Malefactive event and writer attitude annotation. *ACL* (2).
- Duh, Kevin, Akinori Fujino, and Masaaki Nagata. 2011. Is machine translation ripe for cross-lingual sentiment classification?. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies: Short papers-volume 2*. Association for Computational Linguistics.
- Ekman, Paul. 1992. An argument for basic emotions. *Cognition and Emotion* 6 (3–4): 169–200.
- Esuli, Andrea, and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, vol. 6.

- Finn, Arup. 2011. *AFINN*. Informatics and Mathematical Modelling, Technical University of Denmark.
- Go, Alec, Richa Bhayani, and Lei Huang. 2009. *Twitter sentiment classification using distant supervision*. CS224N Project Report, Stanford 1 (2009): 12.
- Joshi, Aditya, A.R. Balamurali, and Pushpak Bhattacharyya. 2010. A fall-back strategy for sentiment analysis in hindi: A case study. In *Proceedings of the 8th ICON*.
- Joshi, Aditya, A.R. Balamurali, and Pushpak Bhattacharyya, and Rajat Mohanty. 2011. C-feel-it: A sentiment analyzer for micro-blogs. In *Proceedings of the 49th annual meeting of the association for computational linguistics*.
- Joshi, Aditya, Abhijit Mishra, and Pushpak Bhattacharyya. 2014. Measuring sentiment annotation complexity of text. In *Conference for association of computational linguistics*.
- Joshi, Aditya, Vinita Sharma, and Pushpak Bhattacharyya. 2015. Harnessing context incongruity for Sarcasm detection. In *Conference for association of computational linguistics*.
- Kiritchenko, Svetlana, Xiaodan Zhu, and Saif M. Mohammad. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50: 723–762.
- Levallois, Clement. 2013. Umigon: Sentiment analysis for tweets based on lexicons and heuristics. In *Proceedings of the international workshop on semantic evaluation*. SemEval, vol. 13.
- Lin, Chenghua, and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM.
- Liu, Bing. 2010. Sentiment analysis and subjectivity. In *Handbook of natural language processing*, vol. 2, 627–666.
- Louviere, Jordan J. 1991. *Best-worst scaling: A model for the largest difference judgments*. Technical report, University of Alberta.
- Lu, Bin, et al. 2011. Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*. Association for Computational Linguistics.
- Mahyoub, Fawaz H.H., Muazzam A. Siddiqui, and Mohamed Y. Dahab. 2014. Building an Arabic sentiment Lexicon using semi-supervised learning. *Journal of King Saud University-Computer and Information Sciences* 26 (4): 417–424.
- Makice, Kevin. 2009. *Twitter API: Up and running: Learn how to build applications with the Twitter API*. Beijing: O'Reilly Media, Inc.
- McAuley, Julian, and Jure Leskovec. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on recommender systems*. ACM.
- Mehrabian, Albert, and James A. Russell. 1974. *An approach to environmental psychology*. Cambridge, MA: MIT Press.
- Melville, Prem, Wojciech Gryc, and Richard D. Lawrence. 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM.
- Mihalcea, Rada, Carmen Banea, and Janyce M. Wiebe. 2007. *Learning multilingual subjective language via cross-lingual projections*.
- Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38 (11): 39–41.
- Mohammad, Saif. 2012. #Emotional tweets. In *Proceedings of the first joint conference on lexical and computational semantics (*Sem)*, June 2012.
- Mohammad, Saif M., and Peter D. Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics.
- . 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29 (3): 436–465.

- Mohammad, S., C. Dunne, and B. Dorr. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *EMNLP*, 599–608.
- Mukherjee, S., and P. Bhattacharyya. 2012. Wikisent: Weakly supervised sentiment analysis through extractive summarization with wikipedia. In *Machine learning and knowledge discovery in databases*, 774–793. Berlin/Heidelberg: Springer.
- Nakov, Preslav, et al. 2013. Semeval-2013 task 2: Sentiment analysis in twitter.
- Ortony, Andrew, and Terence J. Turner. 1990. What's basic about basic emotions? *Psychological Review* 97 (3): 315–331.
- Pang, Bo, and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on association for computational linguistics*. Association for Computational Linguistics.
- . 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing-volume 10*. Association for Computational Linguistics.
- Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5 (5): 411–419.
- Pennebaker, James W., Martha E. Francis, and Roger J. Booth. 2001. *Linguistic inquiry and word count: LIWC 2001*, vol 71. Mahwah: Lawrence Erlbaum Associates.
- Plutchik, Robert. 1980. *Emotion: A psychoevolutionary synthesis*. New York: Harpercollins College Division.
- . 1982. A psychoevolutionary theory of emotions. *Social Science Information/sur les sciences sociales* 21: 529–553.
- Purver, Matthew, and Stuart Battersby. 2012. Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th conference of the European chapter of the association for computational linguistics*. Association for Computational Linguistics.
- Ramteke, Ankit, Pushpak Bhattacharyya, and J. Saketha Nath. 2013. Detecting Turnarounds in sentiment analysis: Thwarting. In *Conference for association of computational linguistics*.
- Redondo, Jaime, et al. 2007. The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods* 39 (3): 600–605.
- Rosenthal, Sara, et al. 2014. Semeval-2014 task 9: Sentiment analysis in twitter. In *Proceedings of SemEval*, 73–80.
- Rosenthal, Sara, Preslav Nakov, Svetlana Kiritchenko, Saif M. Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation*, SemEval.
- Salameh, Mohammad, Saif Mohammad, and Svetlana Kiritchenko. 2015. Sentiment after translation: A case-study on Arabic social media posts. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies*.
- Socher, Richard, et al. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, vol. 1631.
- Somasundaran, Swapna, and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP: Volume 1-volume 1*. Association for Computational Linguistics.
- Stone, Philip J., Dexter C. Dunphy, and Marshall S. Smith. 1966. *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

- Strapparava, Carlo, and Alessandro Valitutti. 2004. WordNet affect: An affective extension of WordNet. LREC, vol. 4.
- Tao, L., Y. Zhang, and V. Sindhvani. 2009. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP Association for Computational Linguistics.
- Thomas, Matt, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*. Association for Computational Linguistics.
- Wan, Xiaojun. 2009. Co-training for cross-lingual sentiment classification. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP: Volume 1-volume 1*. Association for Computational Linguistics.
- Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39 (2–3): 165–210.
- Xu, Ge, Xinfan Meng, and Houfeng Wang. 2010. Build Chinese emotion lexicons using a graph-based algorithm and multiple resources. In *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics.