Chapter 3

Concept-Level Domain Sentiment Discovery

3.1 Introduction

Most of the early sentiment-analysis systems took a *lexicon-based* approach to a document sentiment classification task. This approach is based on the so-called *lexicon design*, having domain-specific sentiment lexicons as the main sentiment information source [92, 111, 149]. Later, the focus of research shifted more to learning-based approaches [39, 53]. Sentiment-analysis systems based on supervised machine-learning techniques usually achieve the best performance in sentiment detection. However in many cases, they are black boxes in the sense that no explanation or justification can be provided to users.

Another concern in sentiment analysis is the domain dependency problem. With a large enough training corpus, a supervised learning-based method can perfectly fit a target domain and achieve a high sentiment classification accuracy. Unfortunately, this comes at a cost, such as the domain overfitting or dependency issue. Domain dependency is not unique to learning-based methods. Other approaches also have difficulties dealing with documents outside of their domain boundaries. However, learning-based methods are more susceptible to this problem and have a higher sensitivity to crossing domain boundaries. On the one hand, bag-of-words learning solutions have superior performance, but they suffer a significant loss of accuracy if domain boundaries are crossed.

To address domain adaptation, researchers have proposed various methods. Almost all of the domain-adaptation experiments have been done on synthetic datasets, which have clearly defined domain boundaries. Yet real-world information sources typically contain a mixture of cross-domain documents and have different characteristics from static experimental datasets.

Customer Review => { (Aspect$_1$: View$_1$), (Aspect$_2$: View$_2$), ..., (Aspect$_k$: View$_k$) }, e.g., a user comment on Google Chrome => { (Appearance: +0.8), (Plugins: +0.6), ..., (Speed: +0.9) }.

Fig. 3.1: An example of *pSenti*'s aspect-oriented output.

Moreover, the domain-adaptation process does not make the underlying sentiment-analysis model less agnostic to domain boundaries. In other words, even after domain adaptation, bag-of-words learning-based sentiment-analysis methods retain their sensitivity to crossing domain boundaries and typically have difficulties dealing with noisy sentiment sources. As we will demonstrate later, lexicon-based systems are less sensitive near domain boundaries. Thus, there is a need for a concept-level sentiment-analysis system that could seamlessly integrate lexicon-based and learning-based approaches to get the best of both.

3.2 Contribution

To overcome the above lexicon and bag-of-words learning limitations, we have developed a novel sentiment-analysis method which is less sensitive to crossing domain boundaries and has similar performance to pure learning-based methods. In this chapter, we present the anatomy of *pSenti*[1 2] — **a concept-level sentiment-analysis system** that seamlessly integrates lexicon-based and learning-based approaches to acquire **adaptive sentiment analysis**.

The main advantage of our *hybrid* approach using a lexicon/learning symbiosis is to get the best of both worlds — the stability and readability of a carefully hand-picked lexicon, and the high accuracy from a powerful supervised learning algorithm. Thanks to the built-in sentiment lexicon and numerous linguistic rules, *pSenti* can detect and measure sentiments at the concept level, providing structured and readable aspect-oriented outputs, as illustrated in Figure 3.1.

The main idea of *pSenti* is to generate feature vectors for supervised machine learning in the same fashion as lexicon-based sentiment-analysis systems see it. In a sense, *pSenti* is a lexicon-based sentiment-analysis system with an integrated learning-based domain-adaptation module. Our experimental results confirmed that such a two-step design is less prone to domain overfitting and less sensitive to a change of topic or writing style. Compared to pure lexicon-based systems, it achieves significantly higher accuracy in sentiment-polarity classification and sentiment-strength detection. Compared to pure learning-based systems,

[1]https://github.com/AndMu/Wikiled.Sentiment
[2]http://www.dcs.bbk.ac.uk/~andrius/psenti/

our method offers more structured and readable results with aspect-oriented explanation and justification, while being less sensitive to the writing style of a text. Moreover, contrary to a bag-of-words design, it can be modified and further adjusted after a learning phase (i.e. we can introduce new linguistic rules or expand a sentiment lexicon at any time to further improve the system's performance). Also, in the case of insufficient labelled training data, it can fall back to a lexicon-based component and perform sentiment analysis of unseen examples.

The ability to perform cross-style sentiment analysis is significant, as it implies that we can train the system using formal professional reviews as training examples and then apply the system for sentiment analysis on informal customer reviews. We cover the anatomy of our proposed approach in **Section 3.4**. The extensive experiments we have carried out on two real-world datasets are reported in **Section 3.5**. Both datasets, CNET software reviews and IMDB movie reviews, confirm the superiority of the proposed composite approach over state-of-the-art systems such as *SentiStrength* [149, 111].

In addition to a single-dimensional sentiment output, it is also able to calculate the eight Plutchik [7] mood dimensions — *anger, anticipation, disgust, fear, joy, sadness, surprise and trust*. Mood dimensions are extracted using the NRC sentiment lexicon [160] and are generated as an XML output for each document and a whole dataset (see Listing 3.1).

```xml
<MoodData>
   <Mood name="Anger" value="0.028"/>
   <Mood name="Anticipation" value="0.054"/>
   <Mood name="Disgust" value="0.011" />
   <Mood name="Fear" value="0.038" />
   <Mood name="Joy" value="0.028" />
   <Mood name="Sadness" value="0.028" />
   <Mood name="Surprise" value="0.015" />
   <Mood name="Trust" value="0.108" />
</MoodData>
```

Listing 3.1: Mood information XML output

pSenti can also resolve four sentiment temporal orientations: *past, present, future*, and *undefined* as well as present results with the greater granularity (see Figure 3.2). A temporal orientation is calculated using two different methods: using the *SuTime* temporal tagger [140] and by finding tense of a sentence.

In later chapters we will make more extensive use of mood and temporal sentiment information.

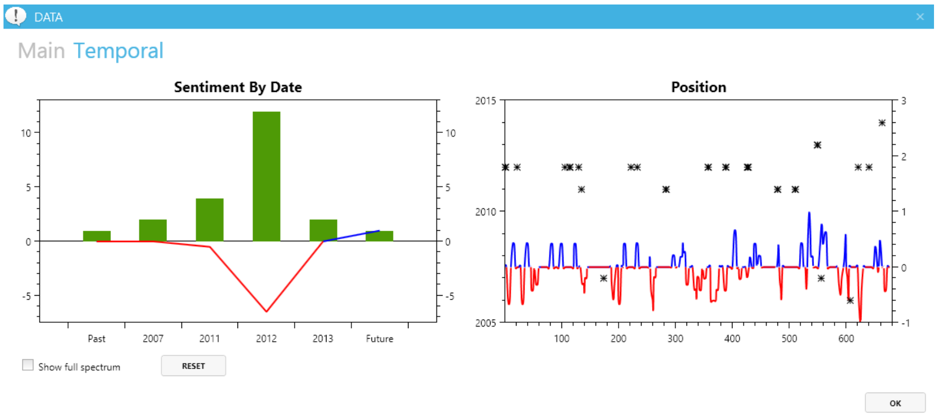


Fig. 3.2: *pSenti* article temporal sentiment-analysis output

3.3 Datasets

To empirically evaluate our *pSenti* system, we conducted experiments on two real-world datasets.

- *The first dataset*: Software-Product-Reviews[3] consists of software product reviews collected by the thesis author from CNET's software download website. The dataset includes five software product categories: Browser, Antivirus, Video, Action Games and Utilities. Most software reviews are written by customers (average users), but there are some which are written by professionals (CNET editors).
- *The second dataset*: Movie Reviews[4] consists of movie reviews collected by Pang and Lee [53] from the IMDB website. It is a well-known standard benchmark dataset for sentiment analysis.

The datasets have been pre-processed to remove duplicates, spam and inconsistencies. They are also balanced (i.e. each class has a similar number of reviews). The detailed characteristics of these datasets are shown in Table 3.1.

3.4 Model

Our concept-level sentiment-analysis system, *pSenti*, is developed by combining lexicon-based and learning-based approaches. As shown in Figure 3.3, the supervised machine-

[3] http://www.dcs.bbk.ac.uk/~andrius/psenti/
[4] http://www.cs.cornell.edu/People/pabo/movie-review-data/

Dataset		Labels of Reviews	Number of Reviews	Avg Size of Reviews
Software Reviews	Miscellaneous (Editor)	Pos/Neg	1660	1056.82
	Browser (Editor)	Pos/Neg	360	1091.61
	Browser (Customer)	Pos/Neg	2000	158.07
	Antivirus (Customer)	Pos/Neg	2000	165.06
	Video (Customer)	Pos/Neg	2000	152.43
	Action Games (Customer)	Pos/Neg	2000	136.21
	Utilities 1 (Customer)	Pos/Neg	2000	155.80
	Utilities 2 (Customer)	1-5 Stars	1850	295.19
Movie Reviews	Movies 1	Pos/Neg	2000	3892.96
	Movies 2	1-5 Stars	5000	2257.44

Table 3.1. The experimental datasets.

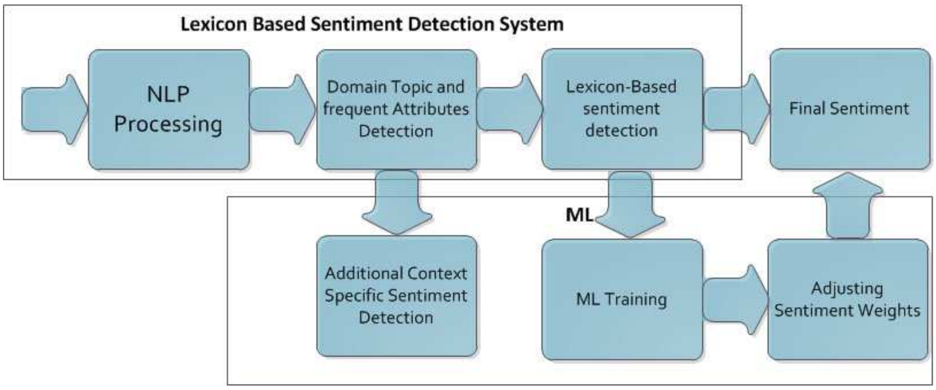


Fig. 3.3: The system architecture of *pSenti*.

learning component is responsible for multiple tasks, such as adjusting sentiment values and new sentiment word discovery. To derive the final output, it performs adjustment of all lexicon components, including semantic rules. The *pSenti* system measures and reports the overall sentiment of a given opinionated text, such as a customer review, as a real-valued score between -1 and $+1$, which can then be easily transformed into a positive/negative classification or a range of 1-5 stars (see Equation (3.3)). It can also output sentiment as an eight-dimensional mood vector.

The system has a rich UI, making it easy to use to analyse sentiment dynamics and understand how sentiment changes over time. Using its interface, we can inspect sentiment changes on both sentence (see Figure 3.4) and word (see Figure 3.5) levels.

Fig. 3.4: Sentence-level *pSenti's* analysis interface

3.4.1 Preprocessing

The core of *pSenti* is its lexicon-based system, so it shares many common NLP processing techniques with other similar approaches. It supports two different NLP frameworks: Stanford CoreNlp [176] and OpenNlp [104]. During the first step of text processing, we carry out tokenisation, POS and entity tagging. Before feeding a piece of a document into the parser, we perform some text clean-up, simplification and transformations.

As part of the transformation, we replace known idioms and emoticons with text masks. *pSenti* can read both text emoticons and those encoded as Unicode images. For example, the emoticon ":-)" or its Unicode representation, will be replaced by the token `EMOTICON_SMILE`. `EMOTICON_SMILE` is listed in the default lexicon as a sentiment word with $+2$ sentiment value. Similarly, ":|", which has a negative sentiment strength -1, will be replaced by the token `EMOTICON_CONFUSED`. The assumption here is that various emoticons express different sentiment strength, which has already been measured, differentiated and added into the standard lexicon. Emoticon tokenisation simplifies their processing and understanding by machine-learning algorithms, and allows them to be further adjusted, depending on the underlying domain.

Fig. 3.5: Word-level *pSenti's* analysis interface

Emoticons are commonly used as a universal method to express sentiment and have similar sentiment strength across many domains [204]. As they are ubiquitous in social media domains, *pSenti* has an option to use the emoticon-only sentiment lexicon. Such an approach is useful in the procedure of bootstrapping the training sample, which we extensively employ in the following chapters.

Idioms follow slightly different heuristic rules, and they are replaced using system-defined tokens. Thus "crocodile tears", known to have sentiment strength -3, should be replaced by `_Bad_Three_` token. The range of sentiment values for emoticons and idioms is from -3 to $+3$. Currently, *pSenti* knows about 116 emoticons and forty idioms.

3.4.2 Aspect and view extraction

Aspect and view extraction play multiple roles in sentiment analysis. People very often express multiple views in a single review (sometimes even of opposite polarity) about distinct aspects of the same item as a software product or a movie. Therefore, it is essential for a practical sentiment-analysis system to extract the discussed aspects and the corresponding views from each document with a sentiment. A view is subjective information expressed in relation to an aspect; thus, we include them in the domain-specific sentiment lexiconinduction step. Aspect detection can also help to create a domain-specific signature and so contribute to recognising to which domain messages belong.

The current implementation of *pSenti* uses a simple aspect and view extraction algorithm as follows:

- **Find candidate aspects**. We generate a list of candidate aspects by including frequent nouns and noun phrases identified by the POS tagger. In this step, we also use Named Entity Recognition (NER) information. If a word has a named entity category assigned, we include only *location*, *organisation* and *person* categories. We excluded all words which have an initial sentiment value, as well as stopwords.
- **Find expressed views**. We generate a list of candidates by including adjectives and known sentiment words which occur near an aspect (in the same sentence) but excluding all stopwords and all types of named-entities.
- **Clean-up**. We further remove all candidate aspects or views that occur less than five times and ensure that the same word can be either an aspect or view.
- **Group similar aspects**. If multiple aspects share the same stem, they are assigned to the same aspect group; we also include the phrases in which they occurred.
- **Generate final aspects and views**. The final list includes only the top 100 of the aspect group, the top 100 views, plus the top ten views for each selected aspect.

Another motivation for *pSenti* to emphasise aspect/view extraction is that the domain-specific aspect words will be excluded from the machine-learning step to reduce the dependence on the current topic domain, writing style or time period. For example, in many of the browser category customer reviews, we can observe very negative sentiments towards "Internet Explorer" and "Microsoft", so if we include these words in the machine-learning step, they would be given high negative values. In the bag-of-word learning-based approach (e.g. using SVM as the learning algorithm), "Microsoft" would be in the top list with a strong negative weight of -1.36, and "Firefox" would have a positive weight of $+1.07$. However, these words do not carry any stable or robust sentiment value, and it is purely a coincidence that, at the time of sentiment analysis, Microsoft IE6 had such negative publicity.

After a couple of years, we might find that the sentiment polarity and strength for these aspect words have become entirely different from their current values. That helps *pSenti* not only to be less sensitive to topic-domain boundaries but also less sensitive to crossing a time-domain boundary.

Besides, aspect/view extraction allows us to find frequently occurring adjectives (views), which can be used to expand the sentiment lexicon and enables us to perform context-aware sentiment-value estimation for such adjectives within the given aspect. For example, the same word, “large”, could have very different sentiment implications in different contexts: the sentiment for a “large monitor” is usually positive, while the sentiment for a “large phone” is probably negative.

3.4.3 Lexicon-based sentiment-detection evaluation

For the first pass of sentiment detection, our system uses the sentiment lexicon constructed using public resources. It is a mixture of various publicly available lexicons, including the Opinion Lexicon compiled by Liu [145], the General Inquirer compiled by Stone et al. [247] and *SentiStrength* by Thelwall, Buckley, and Paltoglou [149]. Currently, the sentiment lexicon consists of 7048 sentiment words including words with wildcards. The wildcard character “*” in such words represents a number of any characters or an empty string (e.g. “graceful*” will match words “graceful”, “gracefully” and “gracefulness”).

Their sentiment values are marked in the range from -3 to $+3$. Based on this sentiment lexicon, we apply the following heuristic linguistic rules to detect sentiments from a text:

- **Negation**. We included both traditional negation words such as “not” and “don’t”, as well as pattern-based negations such as “stop” + “*vb*-ing”, “quit” + “*vb*-ing”. Our system also employs an algorithm in which negation could be applied to more distant sentiments. If a negation word could not be attached to sentiment or another known adjective, it is treated as a negative sentiment word with a weight -1.5 and will generate the feature $w_{not-word}$ for the machine-learning algorithm. As part of the processing, we perform various sentence repairs using heuristic rules for more reliable negation detection. For example, the system detects negation words in phrases such as “not just ... ” and “not only ... but also”, and excludes them as sentiment negations. Besides, it splits words with the “non-” prefix (e.g. the word “non-violent” will be separated into two words, “not violent”, in advance).

- **Modifier**. Since words such as “more” and “less” can boost or reduce the sentiment value of their associated sentiment word, they are considered by our sentiment-detection algorithm. Intensifiers increase the sentiment value by several times, whereas diminishers decrease it several times. Currently, we have forty such handcrafted modifiers with their impact value in the range from 0.4x to 2.5x.

3.4.4 Learning-feature extraction

As was already mentioned above, in our proposed model, the learning phase is responsible for the lexicon part of domain adaptation by adjusting sentiment word values and participating in a domain-specific lexicon expansion. The supervised machine-learning algorithm used in our system is the linear SVM implementation from LibSVM[5], with an L2 objective function for optimisation and grid search for parameter tuning. We chose linear SVM since in previous studies [39] it has been shown that it outperforms other popular learning algorithms for sentiment analysis. Another reason for this selection was the observation identified by Mladenić et al. [52] that weights extracted from a linear SVM can be a good indicator of feature importance. A feature weight obtained from a linear model represents a vector coordinate which is orthogonal to the hyperplane, and its direction indicates the predicted class. The magnitude also tells us how informative a feature is for classification and its importance in a data-separation task.

A classic bag-of-words supervised learning approach takes all words as features; however, not all words carry sentiment information. Thus, if we limit features only to well-known and potential sentiment words, we would be able to use weights extracted from a linear SVM to learn their importance and impact on a sentiment classification task. In other words, we would discover their domain-specific sentiment values. Due to normalisation and standardisation, feature weights discovered in the training phase are on the same scale, and their interpretation can be mapped back into the lexicon-based part and represent their domain-specific sentiment strength. Other lexicon features, such as mood dimensions and overall lexicon-based sentiment strength, are represented by so-called *lexicon bias*. In Section 3.4.7, we will further validate our model and confirm that linear SVM weights can indeed generate a high-quality domain-specific sentiment lexicon.

To perform domain adaptation the following elements are included:

- **Sentiment words**. The weight of this feature is its frequency in a given document multiplied by its *absolute* sentiment strength. For example, if we have a document with the word "good" (sentiment value $+2$), appearing twice, we would generate the feature w_{good} with a weight of $2 \times 2 = 4$. A similar calculation would be performed for the word "bad" (sentiment value -2). It would generate the feature w_{bad} with a weight of $2 \times 2 = 4$. That makes the learning part responsible for the polarity identification. Such a model is agnostic to an initial lexicon polarity, and thus transferable across other lexicons and easier to interpret by human users.

[5] http://www.csie.ntu.edu.tw/cjlin/libsvm/

In the case of sentiment-value modification, the feature and value generation are slightly more complicated. If the sentiment source has been inverted, we generate a new feature to reflect inversion. In the case of the word "good", the feature $w_{not-good}$ would have a sentiment value of -2 with a feature value of $+2$ (absolute value). A similar situation arises with intensifiers or diminishers (e.g. for the bigram "extremely good", where we have the sentiment word "good" appearing next to one of the strongest intensifiers with $x2.5$ impact value, we would generate the $w_{more-good}$ feature with $+2 \times 2.5 = +5$ as its value). In the case of "sometimes good", we would generate $w_{less-good}$ with a value of $+2 \div 1.5 = +1.33$

- **Other adjectives**. Turney [40] was one of the first to identify that adjectives are among the most likely sentiment candidates. Thus, by including adjectives as learning-phase features, we are performing domain-specific sentiment word discovery. For adjective-based features, we use the term frequency weighting. For example, if the word "large" appears twice, we would have the feature w_{large} with a value of 2.0. In this case, a negation, intensifier or diminisher does not modify a feature's weight but only triggers the generation of a new feature. An instance such as the "not clean" bigram, would generate the feature $w_{not-clean}$.

- **Inverted words**. Inverters are typically good indicators of negative sentiment (e.g. the phrase "not working" does not have any well-known sentiment words in it, yet it expresses a strong negative sentiment). Negative sentiments in our lexicon have their strength in the range from -1 to -3, with most words having a value of -1 or -2. Based on empirical evidence, for inverted words in lexicon-based sentiment analysis, we use a value of -1.5, which is in the middle of this range. To mitigate the empirical bias, in the learning phase we use the term frequency weighting.

- **Lexicon-based sentiment score**. We call this feature *lexicon bias*, which is essential in processing documents which contain sentiment words unseen in the training examples. The fallback sentiment lexicon-based classifier plays an important role in a situation where we might have insufficient labelled training data. In such a case, the sentiment strength of these documents can still be derived using a standard fallback sentiment lexicon and lexicon-based heuristic rules. Using this feature, we are measuring how biased the lexicon-based classifier is in a particular domain.

- **Mood dimensions**. We have an option to include eight mood dimensions (see Listing 3.1), extracted using the NRC sentiment lexicon [160]. For each dimension, as its feature weight, we use the probability of occurrence in a given document.

3.4.5 Sentiment scoring

Most of our results are reported in terms of classification into positive and negative classes. However, the actual output is a real-valued sentiment score in the range of $[-1,+1]$. F_{senti} is calculated using the *log-odds*, also known as the *logit* equation (see Equation (3.1)). As discussed in Section 2.2.1, this equation has the smoothest symmetrical distribution in the range of $[-1,+1]$, is symmetric around zero and is most suitable for representing the proportion of two different sentiment poles. In Equation (3.1), w_p is the sum of positive, and w_n of negative (absolute) sentiment values, and β is a fixed coefficient to prevent ill-defined $log_2(0)$. F_{senti} has upper-bound $+1$ and lower-bound -1 (see Equation (3.2)). If the value is greater than +1.0, it will be reset to $+1.0$, and if it is lower than -1.0, it will be reset to -1.0.

$$F_{senti} = \frac{1}{2}(\log_2(\sum w_p + \beta) - \log_2(\sum w_n + \beta)) \tag{3.1}$$

$$F_{senti} = \begin{cases} 1, & \text{if } F_{senti} \geq 1 \\ -1, & \text{if } F_{senti} \leq -1 \\ F_{senti}, & \text{otherwise} \end{cases} \tag{3.2}$$

If neither positive nor negative sentiment is detected, our algorithm treats such text as a neutral text and assigns it the sentiment value 0. The sentiment value can be easily transformed into a five-star scale using the simple Equation (3.3).

$$F_{stars} = 2 * F_{senti} + 3 \tag{3.3}$$

It is important to note that the final sentiment calculation also includes fallback sentiment words adjusted by a penalised *lexicon bias* coefficient.

3.4.6 Sentiment measurement example

To illustrate the calculation process, consider the following review as an example:

"After reading <u>very good</u> reviews online, I bought this one for Evolution class. It is a <u>horrible</u> excuse for a <u>new</u> textbook. <u>Do not buy</u> this <u>horrible</u> book unless it is for a middle school student. If the authors think this book has been written for an <u>advanced</u> audience, then I would suggest that anyone interested in learning evolution <u>not attend</u> University of Washington."

To simplify all calculations, we omit normalisation, mood dimensions and fallback sentiment words. All calculation steps are presented below in Table 3.3 and the sentiment lexicon in Table 3.2.

Word	Sentiment Value
good	$+2$
horrible	-3
excuse	-1
advance	$+1$
interest	$+2$

Table 3.2. The sentiment lexicon snapshot

- **Lexicon-based sentiment-strength calculation**. The review contains five sentiment words: *"good"* (with boosted sentiment value $+2.0 \times 1.5 = +3$), two occurrences of *"horrible"* (with a sentiment value $-3.0 \times 2 = -6$), *"excuse"* (with a sentiment value -1.0), *"advance"* (with a sentiment value $+1.0$) and *"interest"* (with a sentiment value $+2.0$). The review also contains two inverted verbs: *"do not buy"* and *"not attend"*, for which we generate features $w_{not-buy}$ and $w_{not-attend}$. As described in the previous section, all negated words are treated as negative sentiment words and given a weight of -1.5. The sum of all positive sentiment values is $w_p = 3+1+2 = 6$, and the sum of all negative is $w_n = 6+1+1.5+1.5 = 10$. The lexicon-based sentiment value, calculated using Equation (3.1), is -0.387.

- **Learning features extraction**. As it was outlined in the section above, for each sentiment word, we generate a separate feature and use its aggregated sentiment value. For each non-sentiment feature, we will use the term *frequency*. Following the outlined procedure, we would generate the document vector as: $[+3.0, +6.0, +1.0, +1.0, +1.0, +1.0, +2.0, +1.0, -0.387]$ (see Table 3.3). The last value in the vector (-0.387), is the lexicon-based document sentiment value.

- **Learning-based weight discovery**. All documents in a training dataset have to be labelled with a positive or negative label. In this phase, we train linear SVM using a training dataset and extract estimated SVM coefficients from a model. Sample weights are provided in Table 3.3.

- **Learning-based weight adjustment**. To demonstrate how we calculate domain-specific sentiment, we will use the same review. To determine a domain-specific sentiment value for each feature, we will multiply the previously calculated weights by their SVM coefficient (e.g. two occurrences of *"horrible"*, with an original sentiment value of -2 (absolute $+2$), has an overall -1.8 domain-specific sentiment value ($-0.3 \times 2 \times 3 = -1.8$). On the other hand, some sentiment words such as *"advance"* lost their sentiment value ($+1 \times 0 = 0$). For the last feature, which we also call the "*pSenti bias*", we include the originally calculated lexicon sentiment value (-0.387) adjusted by its SVM weight ($+0.05$) plus the SVM hyperplane bias (-0.12), with a final value $-0.387 \times 0.05 + (-0.12) = -0.138$.

To calculate the domain-specific sentiment, we follow the same procedure as above. First, we calculate $w_p = 0.3 + 0.1 = 0.4$ and $w_n = 1.8 + 0.1 + 0.3 + 0.05 + 0.138 = 2.388$. Finally using *log-odds* (see Equation (3.1)) we calculate the review sentiment value. In Table 3.4 the results of the sentiment calculation are shown, with the final sentiment after adjustment (-1), significantly lower than in the original calculations (-0.368). The final result, transformed into five-star grades using Equation (3.3), is just one star out of a possible five.

	Features								
	good	horrible	new	excuse	not-buy	advance	interest	not-attend	F_{senti}
Lexicon step	$+3$	-6	0	-1	-1.5	$+1$	$+2$	-1.5	-0.387
Learning Vector	$+3$	$+6$	$+1$	$+1$	$+1$	$+1$	$+2$	$+1$	-0.387
SVM Weights	$+0.1$	-0.3	$+0.05$	-0.1	-0.3	0	$+0.025$	-0.05	$+0.05$
Learning adjustment	$+0.3$	-1.8	$+0.05$	-0.1	-0.3	0	$+0.05$	-0.05	-0.138

Table 3.3. Weight adjustment stages

Step	$\sum w_p$	$\sum w_n$	F_{senti}	F_{stars}
Lexicon-based	6	10	-0.368	2.263
Learning-based	0.4	2.388	-1	1

Table 3.4. Sentiment rating calculations

3.4.7 Sentiment lexicon information gain evaluation

In this section, to examine the effectiveness of learning-based lexicon adaptation, we attempt to recreate a sentiment lexicon using the Amazon domain dataset. As a benchmark, we took the sentiment lexicon compiled by Liu [145]. This lexicon is not domain specific, but its primary application was in the Amazon domain, and it is representative enough to allow evaluation of a generated lexicon quality.

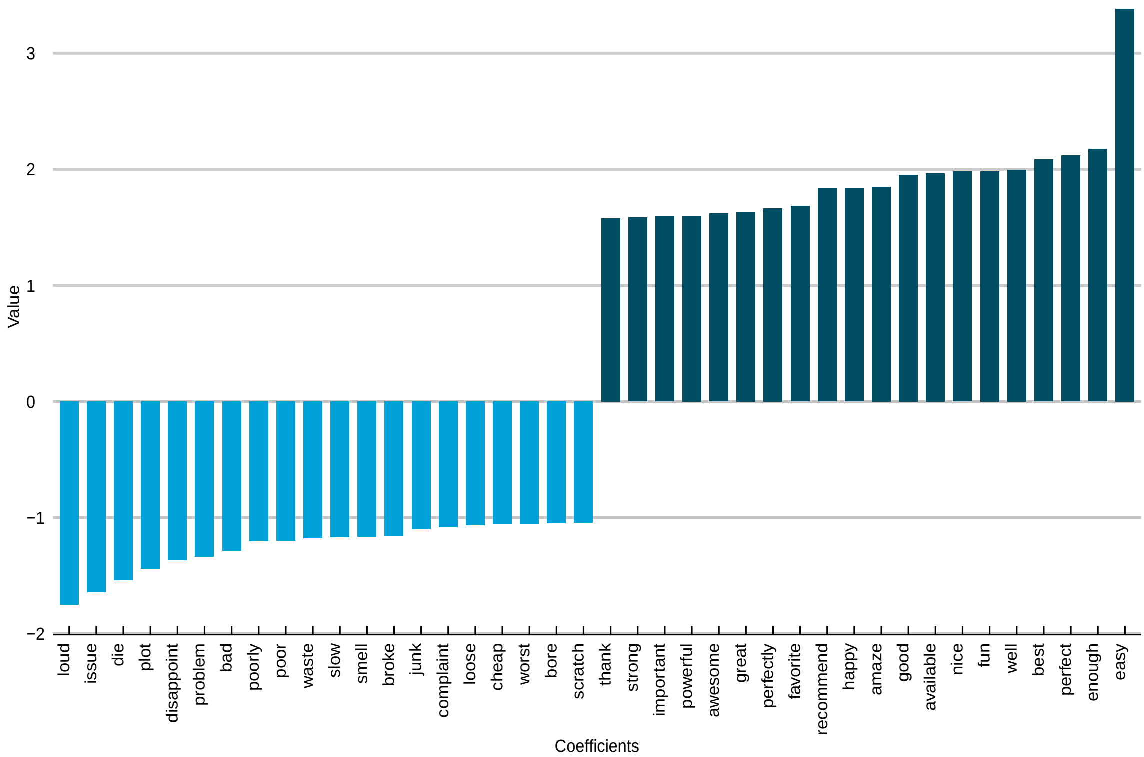


Fig. 3.6: Top 40 SVM feature weights

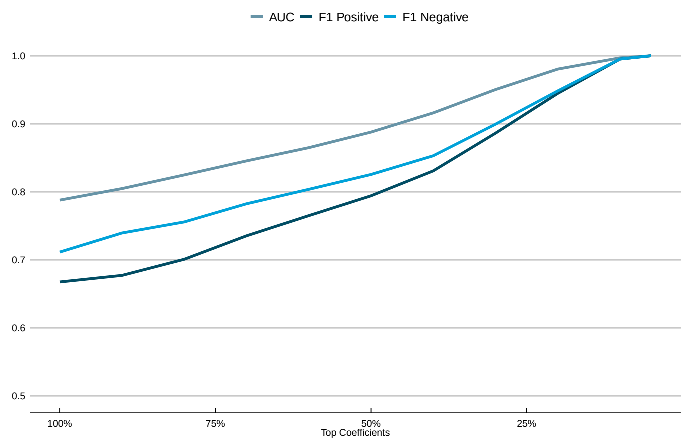


Fig. 3.7: How the performance of lexicon is influenced using feature information gain filtering

To generate a domain-specific lexicon, as described before, we use feature weights extracted from a linear SVM model. Normalisation and standardisation make feature weights extracted from different domains comparable, with the values placed on the same scale.

Top	AUC	F_1^{Positive}	F_1^{Negative}	Size
100%	0.7877	0.6673	0.7113	2064
90%	0.8046	0.6770	0.7393	1872
80%	0.8247	0.7007	0.7556	1654
70%	0.8453	0.7352	0.7823	1444
60%	0.8648	0.7648	0.8035	1238
50%	0.8877	0.7940	0.8254	1032
40%	0.9159	0.8307	0.8529	826
30%	0.9503	0.8862	0.8993	618
20%	0.9803	0.9447	0.9483	412
10%	0.9968	0.9951	0.9951	206
5%	1.0000	1.0000	1.0000	102

Table 3.5. How the performance of lexicon is influenced using feature information gain filtering

Figure 3.6 shows the top 40 features with the most substantial information gain. Different features have different information gain, with most of them having values just fractionally above zero. Thus, to evaluate the adapted lexicon quality, we filter it using information gain and present the results in Figure 3.6 and Table 3.5.

The results indicate that even if we include all generated features, we will get a reasonable-quality lexicon with 0.7877 ROC. Imposing a higher cut-off threshold increases lexicon quality and demonstrates that features with the highest information gain are indeed the most accurate. Taking the top 20% of features generates an impressive ROC value of 0.9803 (see Table 3.5). Results also confirm that the generated lexicon has not only good ROC but also well-balanced F1 scores. Moreover, the actual accuracy is likely to be much higher, as the benchmark lexicon is not domain specific. Manual inspection of the top features (see Figure 3.6) confirms that the strongest sentiment values are indeed identified in words typically associated with customer reviews, such as “loud”, “issue”, “recommend” and “easy”.

3.5 Experimental Results

The *pSenti* system, based on the proposed hybrid approach, is compared with the following baselines:

- *LexiconOnly*: The pure lexicon-based approach using the same sentiment lexicon as *pSenti*;
- *LearningOnly*: The pure learning-based approach using the same learning algorithm (linear SVM) as *pSenti*, with bag-of-words features; and

Dataset		*pSenti*	*LexiconOnly*	*LearningOnly*	*SentiStrength*
Software Reviews	Miscellaneous (Editor)	89.64%	79.40%	**90.78%**	64.93%
	Browser (Editor)	86.94%	76.94%	**91.39%**	62.77%
	Browser (Customer)	79.60%	74.50%	**80.54%**	52.25%
	Antivirus (Customer)	78.55%	70.60%	**82.91%**	47.85%
	Video (Customer)	83.55%	75.95%	**85.83%**	52.80%
	Action Games (Customer)	78.75%	71.55%	**82.92%**	58.25%
	Utilities 1 (Customer)	78.80%	73.70%	**82.03%**	50.50%
Movie Reviews	Movies 1	82.30%	66.00%	**86.85%**	60.70%

Table 3.6. The sentiment-polarity classification performance (accuracy) in the standard (single-style) setting.

- *SentiStrength*[6]: a state-of-the-art sentiment-analysis system free for academic research [149, 111].

All experimental results are reported using 10-fold cross-validation.

3.5.1 Same-domain sentiment analysis

Sentiment-polarity classification

In this experiment, we used two datasets: (i) a set of browser reviews, and (ii) a set of professionally edited reviews of various software products, with both datasets containing balanced data. As explored by Straube and Krell [181], F-measure in Information Retrieval (IR) is a suggested metric for imbalanced classes. However, in the case of a balanced binary classification task, accuracy is the most straightforward measure and is sufficient to explain results. Thus, in this section, we report only accuracy, while the other measures are omitted.

As we can see from the results shown in Table 3.6, our algorithm achieved consistent results across all domains. Performance on customer reviews is lower in comparison to professionally prepared editor reviews, but that could be easily explained by the text quality in customer reviews, rating inconsistencies and different writing styles. In our experiment we tried to mimic real-life situations and made the assumption that an author who wrote a review and assigned a rating is objective in his or her valuation; for that reason, we used all the original review ratings extracted from http://www.donwload.com. However, authors are not always consistent in their ratings — they may write a positive review and assign just a 1-star rating. Also, it is not uncommon to find reviews in which customers express opposite sentiment towards competing products. For example, in our dataset, we have a

[6]http://sentistrength.wlv.ac.uk/

5-star review with the sentence, “Glad to dump Explorer forever!”. In this review, an author expresses negative sentiment towards “Explorer”, yet, the review has a 5-star rating because it refers to the Firefox browser. Nevertheless, as we can see from Table 3.6, the maximum accuracy achieved by our algorithm was 83.55% (for customer video products), and the lowest accuracy was 78.55% (for antivirus products).

In all our experiments using *LexiconOnly* and *SentiStrength* we used default configurations without any training or sentiment-value adjustments, and in both cases final sentiment was calculated using Equation (3.1). *SentiStrength* achieved the lowest scores in all categories; such low accuracy can be explained by the fact that in many reviews *SentiStrength* was not able to detect any sentiment or assigned neutral sentiment value.

To compare our algorithm’s performance with other well-known methods or to a pure machine-learning implementation, we made use of the Pang and Lee [53] movie reviews dataset. Another reason for using this dataset is that movie reviews are usually more difficult to process, as is clearly illustrated in Figure 3.8c, where the pure lexicon-based approach to sentiment analysis would struggle with customer reviews in the movie domain. One reason for such a poor performance is that many of the movie reviews in the given dataset make extensive use of quotes and plot description. For example, in the sentence “when you get out of jail, you can kill him” the author uses several negative words. However, he is not expressing an opinion but just quoting one of the character’s utterances. Such blocks of objective information could be a significant source of sentiment-value distortion, which can be addressed only by processing subjective information blocks. As Pang and Lee [53] have demonstrated in their work, by using such processing it is possible to significantly improve the accuracy of sentiment detection. We have tried to apply a similar (minimum-cut) subjectivity detection algorithm to the datasets we have processed; however, so far, this has not had any positive effect on overall system performance. In this context, we note that subjectivity detection is domain specific and therefore requires a domain-specific training dataset. Nevertheless, as we can see from the results shown in Table 3.6, even without subjectivity detection, our algorithm achieved 82.3% accuracy and was only slightly less accurate than the SVM unigram implementation.

Sentiment-strength detection

All the results have so far been reported as classification into positive/negative classes, but, as previously highlighted, the actual output is sentiment strength. The last experiment results are shown in Table 3.7. This experiment was conducted to illustrate performance on 5-star classifications. As the results show, for utility product customer reviews, our algorithm achieved, on average, an RMSE of 1.56. The one-versus-one (OVO) strategy is regarded as

Dataset		*pSenti*	*LexiconOnly*	*LearningOnly*	*SentiStrength*
Software Reviews	Utilities 2 (Customer)	1.56	1.50	**1.45**	1.77
Movie Reviews	Movies 2	0.87	0.98	**0.60**	1.13

Table 3.7. The sentiment-strength detection performance (RMSE) in the standard (single-style) setting.

Training	Testing	*pSenti*	*LexiconOnly*	*LearningOnly*	*SentiStrength*
Browser (Customer)	Miscellaneous (Editor)	77.47%	**79.40%**	71.92%	64.93%
Browser (Customer)	Browser (Editor)	**77.78%**	76.94%	75.28%	62.77%
Miscellaneous (Editor)	Browser (Customer)	**77.10%**	74.50%	68.55%	52.25%
Browser (Editor)	Browser (Customer)	**75.90%**	74.50%	65.80%	52.25%

Table 3.8. The sentiment-polarity classification performance (accuracy) in the cross-style setting.

one of the most effective SVM strategies available [187] for multi-class sentiment analysis. Thus, in the *LearningOnly* case, we used a five-class one-versus-one classification.

3.5.2 Cross-style sentiment analysis

In this experiment we created a *near-cross-domain* environment, or, in other words, *cross-style*, as both datasets are from the same *topic domain*, yet they use a different writing style. There are many writing styles on the Web, with very distinct features and characteristics. In this section, we define two types of writing style: formal and informal expressions. In a formal journalistic style, writers use well-structured sentences and have a certain composition and length requirement. On the other hand, an informal text is typically short, has irregular grammar and spelling problems, and shows creativity in sentiment expressions. It is not uncommon for a sentiment-analysis system to perform well with one style and significantly worse with another [146]. The experimental results in this section illustrate one of the principal advantages of our algorithm (i.e. lower topic and style dependency compared to a pure SVM implementation). To test this, we made use of two datasets, the professional and informal browser reviews within the same domain, and trained with both SVM and *pSenti* on one type of review to evaluate their performance on another.

As expected, compared to *pSenti*, the SVM-based model excelled in its performance, in all cases, on the same dataset. However, when tested on reviews from another type its performance dropped significantly. In particular, an SVM trained on editor reviews achieved only 68.55% accuracy on customer reviews, as shown in Table 3.8. Such a drop in accuracy

Training	Testing	*pSenti*	*LexiconOnly*	*LearningOnly*	*SentiStrength*
Movies 1	Browser (Customer)	**76.00%**	74.50%	66.95%	52.25%
Movies 1	Utilities 1 (Customer)	**75.30%**	73.70%	65.02%	50.50%
Browser (Customer)	Movies 1	67.70%	66.00%	**67.90%**	60.70%
Utilities 1 (Customer)	Movies 1	67.75%	66.00%	**68.50%**	60.70%

Table 3.9. The sentiment-polarity classification performance (accuracy) in the cross-domain setting.

Training	Testing	*pSenti*	*LexiconOnly*	*LearningOnly*	*SentiStrength*
Utilities 2 (Customer)	Movies 2	1.08	**0.98**	1.24	1.13
Movies 2	Utilities 2 (Customer)	1.65	1.50	**1.40**	1.77

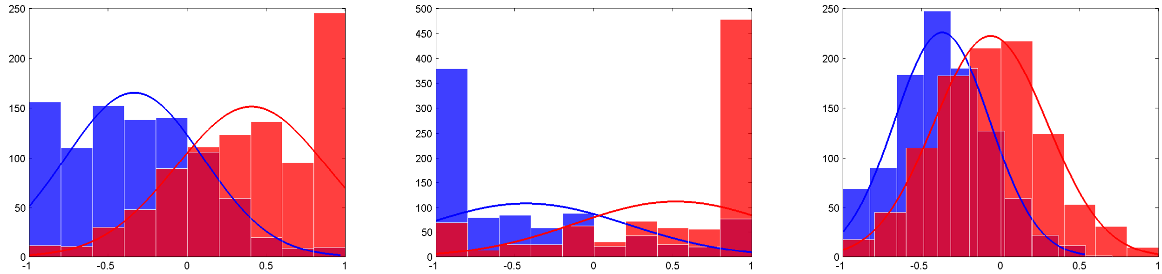
Table 3.10. The sentiment-strength detection performance (RMSE) in the cross-domain setting.

illustrates the weakness of a pure machine-learning method (i.e. overfitting on the training dataset). In contrast, *pSenti* produced consistent results. For customer reviews, it achieved 77.10%, which is just a small 2.5% drop in accuracy, as shown in Table 3.8. This suggests that our mixed algorithm can be trained on one type of reviews and detect sentiments in another type without incurring a significant performance penalty. From the practical point of view, such a system simplifies sentiment processing in less structured social media sources such as Twitter, which usually does not have reliable training data. Moreover, sentiment-strength labelling using professionally prepared text is more reliable, which has more content and is less likely to contain sentiment anomalies. In conjunction with the algorithm's ability to detect the discussed aspects, we can train *pSenti* on different domains and, based on the discussed topic, switch between domain-specific weighting models.

3.5.3 *Distant cross-domain* sentiment analysis

In the final part of our experiments, we analysed various aspects of the system's performance across *distant domain* boundaries. As expected, in all scenarios, *pSenti* was among the best-performing models. On short, informal text processing (see Table 3.9), it was the model with the best overall performance, with only the lexicon-based model producing comparable results.

In the movie reviews dataset, processing performance was significantly lower. We have already highlighted that this domain uses many sentiment words to describe objective information, and, without domain adaptation, all methods failed to take that into account;



(a) Software Reviews (Editor) (b) Software Reviews (Customer) (c) Movie Reviews)

Fig. 3.8: The lexicon-based sentiment-analysis results.

more specifically, the extensive use of quotes and plot description, as well as the domination of negative sentiment words. Looking at the lexicon-based sentiment-analysis results (see Figure 3.8c), we can see a shift towards the negative sentiment scale, which contributes to lower scores. It would be possible to address these issues by looking into the distribution graph and using output calibration.

3.6 Summary and Conclusions

According to the experimental results, the pure lexicon-based approach achieved its best performance on customer software reviews. As Figure 3.8b shows, in this case, the sentiment values have an evident polarisation into positive and negative clusters (i.e. most reviews have a value of either -1 or $+1$). In editor software reviews, as shown in Figure 3.8a, the values are more scattered, but, overall, the classes are still clearly separable. On the other hand, in movie reviews, as shown in Figure 3.8c, the situation is more complicated. In this case, both classes tightly overlap, and that is reflected in the poor performance of the lexicon-based algorithm. Thus, the machine-learning contribution is especially noticeable, and SVM easily detects and offsets such domain anomalies.

Another important topic is how the inclusion of lexicon-based calculated sentiment into the machine-learning vector influenced the system's performance. As expected, its influence can be directly correlated to the original performance in the given domain; in this case, the greatest effect is in customer reviews, and the smallest is in movie reviews. By removing this feature from the movie classification dataset, accuracy drops by only 0.1%, which shows that this feature is entirely ignored by SVM (this can also be seen in the low SVM weight). In editor reviews, on the other hand, it has significantly higher influence. Removing the lexicon-based sentiment feature would result in a loss of 2% in algorithm accuracy.

We have shown that the sentiment-analysis results produced by our *hybrid* approach are favourable compared to the lexicon-only and learning-only baselines. For both sentiment-polarity classification and sentiment-strength detection, the *pSenti* system based on the proposed hybrid approach achieved high accuracy that is very close to the pure learning-based system and much higher than the pure lexicon-based system. Furthermore, *pSenti* can provide sentiment-analysis results in a structured and readable way by dividing the overall sentiment into aspects (e.g., product features) and their corresponding views. Moreover, it has much better tolerance to the writing style of text, as demonstrated by our cross-style experiments where the system is trained on editor reviews and then tested on customer reviews or vice versa. Compared with a representative state-of-the-art sentiment-analysis system *SentiStrength*, the *pSenti* system is consistently and significantly better. In summary, our proposed *hybrid* approach combines the best of two worlds: it provides stability, as well as readability, through a carefully designed lexicon and the high accuracy of a powerful supervised learning algorithm. Results also demonstrated that a supervised linear SVM model could be employed to generate a high-quality domain-specific sentiment lexicon.

It would be promising to explore the potential of this approach further, and we will do so in later chapters.

References

[1] Glenn W Brier. “Verification of Forecasts Expressed in Terms of Probability”. In: *Monthey Weather Review* 78.1 (1950), pp. 1–3.

[2] C. W. J. Granger. “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. In: *Econometrica* 37.3 (Aug. 1969), pp. 424–438. ISSN: 00129682. DOI: 10.2307/1912791.

[3] G. Salton. *The SMART Retrieval System—Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1971.

[4] Greta M Ljung and George EP Box. “On a measure of lack of fit in time series models”. In: *Biometrika* 65.2 (1978), pp. 297–303.

[5] J Welles Wilder. *New concepts in technical trading systems*. Trend Research, 1978.

[6] David A Dickey and Wayne A Fuller. “Distribution of the estimators for autoregressive time series with a unit root”. In: *Journal of the American statistical association* 74.366a (1979), pp. 427–431.

[7] R. Plutchik. “A general psychoevolutionary theory of emotion”. In: *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*. New York: Academic press, 1980, pp. 3–33.

[8] Silvan S Tomkins. “Affect theory”. In: *Approaches to emotion* 163.163–195 (1984).

[9] Nico H Frijda. “The laws of emotion.” In: *American psychologist* 43.5 (1988), p. 349.

[10] Gerard Salton and Christopher Buckley. “Term-weighting approaches in automatic text retrieval”. In: *Information processing & management* 24.5 (1988), pp. 513–523.

[11] JP Wu and Shuony Wei. *Time series analysis*. Hunan Science and Technology Press, ChangSha, 1989.

[12] Janyce Marbury Wiebe. “Recognizing Subjective Sentences: A Computational Investigation of Narrative Text”. UMI Order No. GAX90-22203. PhD thesis. Buffalo, NY, USA: State University of New York at Buffalo, 1990.

[13] George A Miller and Walter G Charles. “Contextual Correlates of Semantic Similarity”. In: *Language and Cognitive Processes* 6.1 (1991), pp. 1–28.

[14] Paul Ekman. “An argument for basic emotions”. In: *Cognition and Emotion* (1992), pp. 169–200.

[15] David Matsumoto. “More evidence for the universality of a contempt expression”. In: *Motivation and Emotion* 16.4 (1992), pp. 363–368.

[16] Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. "Building a Large Annotated Corpus of English: The Penn Treebank". In: *Comput. Linguist.* 19.2 (June 1993), pp. 313–330. ISSN: 0891-2017.

[17] Jan Beran. *Statistics for long-memory processes*. Vol. 61. CRC press, 1994.

[18] Rosalind Wright Picard. *Affective Computing*. Tech. rep. 321. MIT Media Lab, 1995.

[19] Jung-Hua Wang and Jia-Yann Leu. "Stock market trend prediction using ARIMA-based neural networks". In: *Neural Networks, 1996., IEEE International Conference on*. Vol. 4. IEEE. 1996, pp. 2160–2165.

[20] Dirk Geeraerts. *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford University Press, 1997.

[21] Vasileios Hatzivassiloglou and Kathleen R. McKeown. "Predicting the semantic orientation of adjectives". In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. ACL '98. Madrid, Spain: Association for Computational Linguistics, 1997, pp. 174–181.

[22] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

[23] Peter Schauble. *Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases*. Norwell, MA, USA: Kluwer Academic Publishers, 1997. ISBN: 0792398998.

[24] Thorsten Joachims. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features". In: *Proceedings of the 10th European Conference on Machine Learning (ECML)*. Chemnitz, Germany, 1998, pp. 137–142.

[25] Andreas Blank and Peter Koch. *Historical semantics and cognition*. Vol. 13. Walter de Gruyter, 1999.

[26] Thorsten Joachims. "Transductive Inference for Text Classification using Support Vector Machines". In: *Proceedings of the 16th International Conference on Machine Learning (ICML)*. Bled, Slovenia, 1999, pp. 200–209.

[27] Mark Thomason. "The practitioner methods and tool". In: *Journal of Computational Intelligence in Finance* 7.3 (1999), pp. 36–45.

[28] Janyce M. Wiebe, Rebecca F. Bruce, and Thomas P. O'Hara. "Development and Use of a Gold-standard Data Set for Subjectivity Classifications". In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*. ACL '99. College Park, Maryland: Association for Computational Linguistics, 1999, pp. 246–253. ISBN: 1-55860-609-3. DOI: 10.3115/1034678.1034721.

[29] Yiming Yang and Xin Liu. "A Re-examination of Text Categorization Methods". In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. Berkeley, CA, USA, 1999, pp. 42–49.

[30] Vasileios Hatzivassiloglou and Janyce M. Wiebe. “Effects of Adjective Orientation and Gradability on Sentence Subjectivity”. In: *Proceedings of the 18th Conference on Computational Linguistics - Volume 1*. COLING ’00. Saarbrücken, Germany: Association for Computational Linguistics, 2000, pp. 299–305. ISBN: 1-55860-717-X. DOI: 10.3115/990820.990864.

[31] John Platt. “Advances in Large Margin Classifiers”. In: ed. by A. Smola et al. Cambridge, MA, USA: MIT Press, 2000. Chap. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods.

[32] Klaus R Scherer. “Psychological Models of Emotion”. In: *The Neuropsychology of Emotion*. Ed. by Joan C. Borod. Oxford University Press (OUP), 2000, pp. 137–162.

[33] Eric W. Weisstein. *Reversion to the Mean*. 2000. URL: http://mathworld.wolfram.com/ReversiontotheMean.html.

[34] Eibe Frank and Mark A. Hall. “A Simple Approach to Ordinal Classification”. In: *Proceedings of the 12th European Conference on Machine Learning (ECML)*. Freiburg, Germany, 2001, pp. 145–156.

[35] Zonglu He and Koichi Maekawa. “On spurious Granger causality”. In: *Economics Letters* 73.3 (2001), pp. 307–313.

[36] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001.

[37] Yannis Hamilakis. “The past as oral history”. In: *Thinking through the Body*. Springer, 2002, pp. 121–136.

[38] Chih-Wei Hsu and Chih-Jen Lin. “A Comparison of Methods for Multiclass Support Vector Machines”. In: *IEEE Transactions on Neural Networks (TNN)* 13.2 (2002), pp. 415–425.

[39] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. “Thumbs up?: sentiment classification using machine learning techniques”. In: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*. EMNLP ’02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[40] Peter D. Turney. “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, PA, USA, 2002, pp. 417–424.

[41] Li-Juan Cao and Francis Eng Hock Tay. “Support vector machine with adaptive parameters in financial time series forecasting”. In: *IEEE Transactions on neural networks* 14.6 (2003), pp. 1506–1518.

[42] Haidt J. “The moral emotions”. In: *Davidson JR, Scherer KR, Goldsmith HH (eds) Handbook of affective sciences. Oxford University Press, Oxford* (2003).

[43] Thorsten Joachims. “Transductive Learning via Spectral Graph Partitioning”. In: *Proceedings of the 20th International Conference on Machine Learning (ICML)*. Washington, DC, USA, 2003, pp. 290–297.

[44] James Pustejovsky et al. “The timebank corpus”. In: *Corpus linguistics*. Vol. 2003. Lancaster, UK. 2003, p. 40.

[45] Ellen Riloff and Janyce Wiebe. "Learning Extraction Patterns for Subjective Expressions". In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Sapporo, Japan, 2003, pp. 105–112.

[46] Hong Yu and Vasileios Hatzivassiloglou. "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences". In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Sapporo, Japan, 2003, pp. 129–136.

[47] Xiaojin Zhu, Zoubin Ghahramani, and John D. Lafferty. "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions". In: *Proceedings of the 20th International Conference on Machine Learning (ICML)*. Washington, DC, USA, 2003, pp. 912–919.

[48] Philip Beineke et al. *An Exploration of Sentiment Summarization*. Tech. rep. In AAAI Spring Symposium on Exploring Attitude, Affect in Text: Theories, and Applications (AAAI, 2004.

[49] Steven Bethard et al. "Automatic extraction of opinion propositions and their holders". In: *2004 AAAI spring symposium on exploring attitude and affect in text*. Vol. 2224. 2004.

[50] Minqing Hu and Bing Liu. "Mining and summarizing customer reviews". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD '04. Seattle, WA, USA: ACM, 2004, pp. 168–177. ISBN: 1-58113-888-1.

[51] Jaap Kamps et al. "Using WordNet to Measure Semantic Orientations of Adjectives." In: *LREC*. Vol. 4. Citeseer. 2004, pp. 1115–1118.

[52] Dunja Mladenić et al. "Feature selection using linear classifier weights: interaction with classification models". In: *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. 2004, pp. 234–241.

[53] Bo Pang and Lillian Lee. "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts". In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. ACL '04. Barcelona, Spain: Association for Computational Linguistics, 2004.

[54] Anthony Aue and Michael Gamon. "Customizing sentiment classifiers to new domains: A case study". In: *Proceedings of recent advances in natural language processing (RANLP)*. Vol. 1. 1-3. 2005, pp. 2–1.

[55] Giuseppe Carenini, Raymond T Ng, and Ed Zwart. "Extracting knowledge from evaluative text". In: *Proceedings of the 3rd international conference on Knowledge capture*. ACM. 2005, pp. 11–18.

[56] Wei Huang, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine". In: *Computers & Operations Research* 32.10 (2005), pp. 2513–2522.

[57] Alexandru Niculescu-Mizil and Rich Caruana. "Predicting Good Probabilities with Supervised Learning". In: *Proceedings of the 22nd International Conference on Machine Learning (ICML)*. Bonn, Germany, 2005, pp. 625–632.

[58] Ping-Feng Pai and Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting". In: *Omega* 33.6 (2005), pp. 497–505.

[59] Bo Pang and Lillian Lee. "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. 2005, pp. 115–124.

[60] Ana-Maria Popescu and Oren Etzioni. "Extracting product features and opinions from reviews". In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. HLT '05. Vancouver, British Columbia, Canada: Association for Computational Linguistics, 2005, pp. 339–346.

[61] Jonathon Read. "Using emoticons to reduce dependency in machine learning techniques for sentiment classification". In: *Proceedings of the ACL Student Research Workshop*. ACLstudent '05. Ann Arbor, Michigan: Association for Computational Linguistics, 2005, pp. 43–48.

[62] Janyce Wiebe and Ellen Riloff. "Creating subjective and objective sentence classifiers from unannotated texts". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer. 2005, pp. 486–497.

[63] Janyce Wiebe, Theresa Wilson, and Claire Cardie. "Annotating expressions of opinions and emotions in language". In: *Language resources and evaluation* 39.2 (2005), pp. 165–210.

[64] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis". In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics. 2005, pp. 347–354.

[65] Steven Bird. "NLTK: the natural language toolkit". In: *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics. 2006, pp. 69–72.

[66] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[67] Giuseppe Carenini, Raymond Ng, and Adam Pauls. "Multi-document summarization of evaluative text". In: *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*. 2006, pp. 305–312.

[68] Yejin Choi, Eric Breck, and Claire Cardie. "Joint extraction of entities and relations for opinion recognition". In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2006, pp. 431–439.

[69] Andrew B. Goldberg and Xiaojin Zhu. "Seeing Stars when There Aren'T Many Stars: Graph-based Semi-supervised Learning for Sentiment Categorization". In: *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*. TextGraphs-1. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 45–52.

[70] Minqing Hu and Bing Liu. "Opinion Feature Extraction Using Class Sequential Rules". In: *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. Stanford, CA, USA, 2006, pp. 61–66.

[71] Xiangji Huang et al. "Applying Data Mining to Pseudo-Relevance Feedback for High Performance Text Retrieval". In: *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM)*. Hong Kong, China, 2006, pp. 295–306.

[72] Soo-Min Kim and Eduard Hovy. "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text". In: *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. SST '06. Sydney, Australia: Association for Computational Linguistics, 2006, pp. 1–8. ISBN: 1-932432-75-2.

[73] Moshe Koppel and Jonathan Schler. "The importance of neutral examples for learning sentiment". In: *Computational Intelligence* 22.2 (2006), pp. 100–109.

[74] Qiaozhu Mei et al. "A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs". In: *Proceedings of the 15th International Conference on World Wide Web*. WWW '06. Edinburgh, Scotland: ACM, 2006, pp. 533–542. ISBN: 1-59593-323-9. DOI: 10.1145/1135777.1135857.

[75] Sara Owsley, Sanjay Sood, and Kristian J. Hammond. "Domain Specific Affective Classification of Documents". In: *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. Stanford, CA, USA, 2006, pp. 181–183.

[76] Shlomo Argamon et al. "Stylistic text classification using functional lexical features: Research Articles". In: *J. Am. Soc. Inf. Sci. Technol.* 58.6 (Apr. 2007), pp. 802–822. ISSN: 1532-2882.

[77] John Blitzer, Mark Dredze, and Fernando Pereira. "Biographies, Bollywood, Boomboxes and Blenders: Domain Adaptation for Sentiment Classification". In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*. Prague, Czech Republic, 2007, pp. 440–447.

[78] Eric Breck, Yejin Choi, and Claire Cardie. "Identifying Expressions of Opinion in Context." In: *IJCAI*. Vol. 7. 2007, pp. 2683–2688.

[79] Chrysanthos Dellarocas, Xiaoquan Michael Zhang, and Neveen F Awad. "Exploring the value of online product reviews in forecasting sales: The case of motion pictures". In: *Journal of Interactive marketing* 21.4 (2007), pp. 23–45.

[80] Andrea Esuli and Fabrizio Sebastiani. "SentiWordNet: a high-coverage lexical resource for opinion mining". In: *Evaluation* (2007), pp. 1–26.

[81] Dirk Geeraerts and Hubert Cuyckens. "Introducing cognitive linguistics". In: (2007).

[82] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. "Large-Scale Sentiment Analysis for News and Blogs." In: *ICWSM* 7.21 (2007), pp. 219–222.

[83] Nobuhiro Kaji and Masaru Kitsuregawa. "Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents." In: *EMNLP-CoNLL*. 2007, pp. 1075–1083.

[84] Jingjing Liu et al. "Low-Quality Product Review Detection in Opinion Summarization." In: *EMNLP-CoNLL*. Vol. 7. 2007, pp. 334–342.

[85] Qiaozhu Mei et al. "Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs". In: *Proceedings of the 16th International Conference on World Wide Web*. WWW '07. Banff, Alberta, Canada: ACM, 2007, pp. 171–180. ISBN: 978-1-59593-654-7. DOI: 10.1145/1242572.1242596.

[86] Marvin Minsky. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster, 2007.

[87] Bo Qian and Khaled Rasheed. “Stock market prediction with multiple classifiers”. In: *Applied Intelligence* 26.1 (2007), pp. 25–33.

[88] Carlo Strapparava and Rada Mihalcea. “SemEval-2007 Task 14: Affective Text”. In: *Proceedings of the 4th International Workshop on Semantic Evaluations*. SemEval ’07. Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 70–74.

[89] Mashek DJ Tangney JP Stuewig J. “Moral emotions and moral behavior”. In: . *Annu Rev Psychol 58:345–372* (2007).

[90] Alina Andreevskaia and Sabine Bergler. “When specialists and generalists work together: overcoming domain dependence in sentiment tagging”. In: *In Proceedings of ACL-08: HLT*. 2008.

[91] Yubo Chen and Jinhong Xie. “Online consumer review: Word-of-mouth as a new element of marketing communication mix”. In: *Management science* 54.3 (2008), pp. 477–491.

[92] Xiaowen Ding, Bing Liu, and Philip S. Yu. “A holistic lexicon-based approach to opinion mining”. In: *Proceedings of the international conference on Web search and web data mining*. WSDM ’08. Palo Alto, California, USA: ACM, 2008, pp. 231–240. ISBN: 978-1-59593-927-2.

[93] Mulligan K. “Moral Emotions”. In: *Sander D, Scherer KR (eds) Oxford companion to the affective sciences. Oxford University Press, Oxford* (2008).

[94] Yue Lu and Chengxiang Zhai. “Opinion integration through semi-supervised topic modeling”. In: *Proceedings of the 17th international conference on World Wide Web*. ACM. 2008, pp. 121–130.

[95] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research (JMLR)* 9.Nov (2008), pp. 2579–2605.

[96] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008. ISBN: 0521865719, 9780521865715.

[97] Bo Pang, Lillian Lee, et al. “Opinion mining and sentiment analysis”. In: *Foundations and Trends® in Information Retrieval* 2.1–2 (2008), pp. 1–135.

[98] Alec Go, Richa Bhayani, and Lei Huang. “Twitter sentiment classification using distant supervision”. In: *CS224N Project Report, Stanford* 1.2009 (2009), p. 12.

[99] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. Springer, 2009.

[100] Bernard J Jansen et al. “Twitter power: Tweets as electronic word of mouth”. In: *Journal of the Association for Information Science and Technology* 60.11 (2009), pp. 2169–2188.

[101] Chenghua Lin and Yulan He. “Joint sentiment/topic model for sentiment analysis”. In: *Proceedings of the 18th ACM conference on Information and knowledge management*. CIKM ’09. Hong Kong, China: ACM, 2009, pp. 375–384. ISBN: 978-1-60558-512-3.

[102] Xiaowei Lin, Zehong Yang, and Yixu Song. “Short-term stock price prediction based on echo state networks”. In: *Expert systems with applications* 36.3 (2009), pp. 7313–7317.

[103] Rein Ahas et al. “Daily rhythms of suburban commuters’ movements in the Tallinn metropolitan area: Case study with mobile positioning data”. In: *Transportation Research Part C: Emerging Technologies* 18.1 (2010), pp. 45–54.

[104] Apache. *OpenNLP*. 2010. URL: http://opennlp.apache.org (visited on 07/19/2013).

[105] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.” In: *LREC*. Vol. 10. 2010, pp. 2200–2204.

[106] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. “Twitter mood predicts the stock market”. In: *CoRR* abs/1010.3003 (2010).

[107] Samuel Brody and Noemie Elhadad. “An Unsupervised Aspect-sentiment Model for Online Reviews”. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. HLT ’10. Los Angeles, California: Association for Computational Linguistics, 2010, pp. 804–812. ISBN: 1-932432-65-5.

[108] Bing Liu. “Sentiment analysis and subjectivity”. In: *Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca*. 2010.

[109] S. J. Pan and Q. Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. ISSN: 1041-4347. DOI: 10.1109/TKDE.2009.191.

[110] Sinno Jialin Pan et al. “Cross-domain Sentiment Classification via Spectral Feature Alignment”. In: *Proceedings of the 19th International Conference on World Wide Web*. WWW ’10. Raleigh, North Carolina, USA: ACM, 2010, pp. 751–760. ISBN: 978-1-60558-799-8. DOI: 10.1145/1772690.1772767.

[111] Mike Thelwall et al. “Sentiment strength detection in short informal text”. In: *Journal of the American Society for Information Science and Technology* 61.12 (2010), pp. 2544–2558. ISSN: 1532-2890.

[112] Andranik Tumasjan et al. “Predicting elections with twitter: What 140 characters reveal about political sentiment.” In: *Icwsm* 10.1 (2010), pp. 178–185.

[113] Yunfang Wu and Peng Jin. “SemEval-2010 Task 18: Disambiguating Sentiment Ambiguous Adjectives”. In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. SemEval ’10. Los Angeles, California: Association for Computational Linguistics, 2010, pp. 81–85.

[114] Feng Zhu and Xiaoquan Zhang. “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics”. In: *Journal of marketing* 74.2 (2010), pp. 133–148.

[115] Jian Zhu, Hanshi Wang, and JinTao Mao. “Sentiment classification using genetic algorithm and conditional random fields”. In: *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*. IEEE. 2010, pp. 193–196.

[116] Adam Bermingham and Alan F. Smeaton. *On Using Twitter to Monitor Political Sentiment and Predict Election Results*. 2011.

[117] Samuel Brody and Nicholas Diakopoulos. “Coooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs”. In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2011, pp. 562–570.

[118] Murphy Choy et al. “A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction”. In: *arXiv preprint arXiv:1108.5520* (2011).

[119] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. “Forecasting time series with complex seasonal patterns using exponential smoothing”. In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1513–1527.

[120] Shangkun Deng et al. “Combining technical analysis with sentiment analysis for stock price prediction”. In: *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*. IEEE. 2011, pp. 800–807.

[121] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. “Domain adaptation for large-scale sentiment classification: A deep learning approach”. In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 513–520.

[122] Joseph M Hilbe. “Logistic regression”. In: *International Encyclopedia of Statistical Science*. Springer, 2011, pp. 755–758.

[123] Yohan Jo and Alice H. Oh. “Aspect and sentiment unification model for online review analysis”. In: *Proceedings of the fourth ACM international conference on Web search and data mining*. WSDM ’11. Hong Kong, China: ACM, 2011, pp. 815–824. ISBN: 978-1-4503-0493-1.

[124] Shou-Shan Li, Chu-Ren Huang, and Cheng-Qing Zong. “Multi-Domain Sentiment Classification with Classifier Combination”. In: *Journal of Computer Science and Technology* 26.1 (2011), pp. 25–33. ISSN: 1860-4749. DOI: 10.1007/s11390-011-9412-y.

[125] Bing Liu. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Second Edition*. Data-Centric Systems and Applications. Springer, 2011. ISBN: 978-3-642-19459-7. DOI: 10.1007/978-3-642-19460-3.

[126] Will Lowe et al. “Scaling policy preferences from coded political texts”. In: *Legislative studies quarterly* 36.1 (2011), pp. 123–155.

[127] Yue Lu et al. “Automatic construction of a context-aware sentiment lexicon: an optimization approach”. In: *Proceedings of the 20th international conference on World wide web*. ACM. 2011, pp. 347–356.

[128] Andrew L. Maas et al. “Learning Word Vectors for Sentiment Analysis”. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*. HLT ’11. Portland, Oregon: Association for Computational Linguistics, 2011, pp. 142–150. ISBN: 978-1-932432-87-9.

[129] Jean-Baptiste Michel et al. “Quantitative analysis of culture using millions of digitized books”. In: *science* 331.6014 (2011), pp. 176–182.

[130] Viswa Mani Kiran Peddinti and Prakriti Chintalapoodi. “Domain Adaptation in Sentiment Analysis of Twitter.” In: *Analyzing Microtext* 11 (2011), p. 05.

[131] Guang Qiu et al. "Opinion Word Expansion and Target Extraction Through Double Propagation". In: *Comput. Linguist.* 37.1 (Mar. 2011), pp. 9–27. ISSN: 0891-2017. DOI: 10.1162/coli_a_00034.

[132] Rajhans Samdani and Wen-Tau Yih. "Domain Adaptation with Ensemble of Feature Groups". In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*. IJCAI'11. Barcelona, Catalonia, Spain: AAAI Press, 2011, pp. 1458–1464. ISBN: 978-1-57735-514-4. DOI: 10.5591/978-1-57735-516-8/IJCAI11-246.

[133] Richard Socher et al. "Semi-supervised recursive autoencoders for predicting sentiment distributions". In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2011, pp. 151–161.

[134] Rui Xia and Chengqing Zong. "A POS-based Ensemble Model for Cross-Domain Sentiment Classification". In: *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*. Chiang Mai, Thailand, 2011, pp. 614–622.

[135] Xue Zhang, Hauke Fuehres, and Peter A Gloor. "Predicting stock market indicators through twitter "I hope it is not as bad as I fear"". In: *Procedia-Social and Behavioral Sciences* 26 (2011), pp. 55–62.

[136] P Thomas Barthelemy, Devin Guillory, and Chip Mandal. *Using Twitter Data to Predict Box Office Revenues*. 2012.

[137] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. "Machine learning strategies for time series forecasting". In: *European Business Intelligence Summer School*. Springer. 2012, pp. 62–77.

[138] Javier Borondo et al. "Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study". In: *Chaos: an interdisciplinary journal of nonlinear science* 22.2 (2012), p. 023138.

[139] Erik Cambria, Andrew Livingstone, and Amir Hussain. "The hourglass of emotions". In: *Cognitive behavioural systems* (2012), pp. 144–157.

[140] Angel X. Chang and Christopher Manning. "SUTime: A library for recognizing and normalizing time expressions". In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. ISBN: 978-2-9517408-7-7.

[141] Shom Prasad Das and Sudarsan Padhy. "Support vector machines for prediction of futures prices in Indian stock market". In: *International Journal of Computer Applications* 41.3 (2012).

[142] Stefano Faralli and Roberto Navigli. "A New Minimally-supervised Framework for Domain Word Sense Disambiguation". In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. EMNLP-CoNLL '12. Jeju Island, Korea: Association for Computational Linguistics, 2012, pp. 1411–1422.

[143] Fabian Gieseke et al. "Sparse Quasi-Newton Optimization for Semi-Supervised Support Vector Machines". In: *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. Vilamoura, Algarve, Portugal, 2012, pp. 45–54.

[144] Yulan He. "Incorporating Sentiment Prior Knowledge for Weakly Supervised Sentiment Analysis". In: *ACM Transactions on Asian Language Information Processing (TALIP)* 11.2 (June 2012), 4:1–4:19.

[145] Bing Liu. "Sentiment analysis and opinion mining". In: *Synthesis lectures on human language technologies* 5.1 (2012), pp. 1–167.

[146] Hiroshi Maeda, Kazutaka Shimada, and Tsutomu Endo. "Twitter sentiment analysis based on writing style". In: *Advances in Natural Language Processing*. Springer, 2012, pp. 278–288.

[147] Yelena Mejova and Padmini Srinivasan. "Crossing Media Streams with Sentiment: Domain Adaptation in Blogs, Reviews and Twitter." In: *ICWSM*. 2012.

[148] Andrius Mudinas, Dell Zhang, and Mark Levene. "Combining Lexicon and Learning Based Approaches for Concept-level Sentiment Analysis". In: *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*. WISDOM '12. Beijing, China: ACM, 2012.

[149] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. "Sentiment Strength Detection for the Social Web". In: *Journal of the American Society for Information Science and Technology (JASIST)* 63.1 (2012), pp. 163–173.

[150] Hao Wang et al. "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle". In: *Proceedings of the ACL 2012 System Demonstrations*. ACL '12. Jeju Island, Korea: Association for Computational Linguistics, 2012, pp. 115–120.

[151] Xiaohui Yu et al. "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain". In: *IEEE Transactions on Knowledge and Data Engineering* 24.4 (2012), pp. 720–734. ISSN: 1041-4347.

[152] Danushka Bollegala, David Weir, and John Carroll. "Using a Sentiment Sensitive Thesaurus". In: *IEEE Trans. on Knowl. and Data Eng.* 25.8 (Aug. 2013), pp. 1719–1731. ISSN: 1041-4347. DOI: 10.1109/TKDE.2012.103.

[153] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for research and an integration of findings*. Elsevier, 2013.

[154] Mohsen Farhadloo and Erik Rolland. "Multi-class sentiment analysis with clustering and score representation". In: *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*. IEEE. 2013, pp. 904–912.

[155] Alexander Hogenboom et al. "Exploiting emoticons in sentiment analysis". In: *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM. 2013, pp. 703–710.

[156] Julian McAuley and Jure Leskovec. "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text". In: *Proceedings of the 7th ACM Conference on Recommender Systems*. RecSys '13. Hong Kong, China: ACM, 2013, pp. 165–172. ISBN: 978-1-4503-2409-0. DOI: 10.1145/2507157.2507163.

[157] Tomas Mikolov et al. “Distributed Representations of Words and Phrases and Their Compositionality”. In: *Advances in Neural Information Processing Systems 26: Annual Conference on Neural Information Processing Systems (NIPS)*. Vol. 26. Lake Tahoe, NV, USA, 2013, pp. 3111–3119.

[158] Tomas Mikolov et al. “Distributed Representations of Words and Phrases and their Compositionality”. In: *CoRR* abs/1310.4546 (2013).

[159] Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. “NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets”. In: *CoRR* abs/1308.6242 (2013).

[160] Saif M. Mohammad and Peter D. Turney. “Crowdsourcing a Word-Emotion Association Lexicon”. In: 29.3 (2013), pp. 436–465.

[161] Preslav Nakov et al. *SemEval-2013 Task 2: Sentiment Analysis in Twitter*. 2013.

[162] Reynier Ortega et al. “Improving subjectivity detection using unsupervised subjectivity word sense disambiguation”. In: *Procesamiento del Lenguaje Natural* 51 (2013).

[163] Elvis Picardo. *How To Trade The News*. Nov. 2013. URL: https://www.investopedia.com/articles/active-trading/111313/how-trade-news.asp.

[164] Matthew A Russell. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. " O’Reilly Media, Inc.", 2013.

[165] Stefan Wager, Sida I. Wang, and Percy Liang. “Dropout Training as Adaptive Regularization”. In: *Advances in Neural Information Processing Systems 26: Annual Conference on Neural Information Processing Systems (NIPS)*. Lake Tahoe, NV, USA, 2013, pp. 351–359.

[166] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. “Norms of valence, arousal, and dominance for 13,915 English lemmas”. In: *Behavior Research Methods* 45.4 (2013), pp. 1191–1207. ISSN: 1554-3528.

[167] Mohd Ridzwan Yaakub, Yuefeng Li, and Jinglan Zhang. “Integration of sentiment analysis into customer relational model: the importance of feature ontology and synonym”. In: *Procedia Technology* 11 (2013), pp. 495–501.

[168] Marta Arias, Argimiro Arratia, and Ramon Xuriguera. “Forecasting with Twitter Data”. In: *ACM Trans. Intell. Syst. Technol.* 5.1 (Jan. 2014), 8:1–8:24. ISSN: 2157-6904. DOI: 10.1145/2542182.2542190.

[169] Darina Benikova et al. “Germeval 2014 named entity recognition shared task: companion paper”. In: *Organization* 7 (2014), p. 281.

[170] Stephan Günnemann, Nikou Günnemann, and Christos Faloutsos. “Detecting anomalies in dynamic rating data: A robust probabilistic model for rating evolution”. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2014, pp. 841–850.

[171] Jasleen Kaur and Jatinderkumar R Saini. “Emotion detection and sentiment analysis in text corpus: a differential study with informal and formal writing styles”. In: *International Journal of Computer Applications* 101.9 (2014).

[172] Yoon Kim. "Convolutional Neural Networks for Sentence Classification". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar, 2014, pp. 1746–1751.

[173] Yoon Kim. "Convolutional Neural Networks for Sentence Classification". In: *CoRR* abs/1408.5882 (2014).

[174] Himabindu Lakkaraju, Richard Socher, and Chris Manning. "Aspect specific sentiment analysis using hierarchical deep learning". In: *NIPS Workshop on Deep Learning and Representation Learning*. 2014.

[175] Heeyoung Lee et al. "On the Importance of Text Analysis for Stock Price Prediction". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.* 2014, pp. 1170–1175.

[176] Christopher D Manning et al. "The stanford corenlp natural language processing toolkit." In: *ACL (System Demonstrations)*. 2014, pp. 55–60.

[177] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey". In: *Ain Shams Engineering Journal* 5.4 (2014), pp. 1093–1113.

[178] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1532–1543.

[179] Sara Rosenthal et al. "SemEval-2014 Task 9: Sentiment Analysis in Twitter." In: *SemEval@COLING*. Ed. by Preslav Nakov and Torsten Zesch. The Association for Computer Linguistics, 2014, pp. 73–80. ISBN: 978-1-941643-24-2.

[180] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research (JMLR)* 15.1 (2014), pp. 1929–1958.

[181] Sirko Straube and Mario M. Krell. "How to evaluate an agent's behavior to infrequent events?—Reliable performance estimation insensitive to class distribution". In: *Frontiers in Computational Neuroscience* 8 (2014), p. 43. ISSN: 1662-5188. DOI: 10.3389/fncom.2014.00043.

[182] Danushka Bollegala, Takanori Maehara, and Ken-ichi Kawarabayashi. "Unsupervised Cross-Domain Word Representation Learning". In: *CoRR* abs/1505.07184 (2015).

[183] Kai Chen, Yi Zhou, and Fangyan Dai. "A LSTM-based method for stock returns prediction: A case study of China stock market". In: *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE. 2015, pp. 2823–2824.

[184] Emil Şt Chifu, Tiberiu Şt Leţia, and Viorica R Chifu. "Unsupervised aspect level sentiment analysis using Ant Clustering and Self-organizing Maps". In: *Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on*. IEEE. 2015, pp. 1–9.

[185] Andrew M. Dai and Quoc V. Le. "Semi-Supervised Sequence Learning". In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems (NIPS)*. Montreal, Quebec, Canada, 2015, pp. 3079–3087.

[186] Xiao Ding et al. "Deep Learning for Event-driven Stock Prediction". In: *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI'15. Buenos Aires, Argentina: AAAI Press, 2015, pp. 2327–2333. ISBN: 978-1-57735-738-4.

[187] Mikel Galar et al. "DRCW-OVO: distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems". In: *Pattern recognition* 48.1 (2015), pp. 28–42.

[188] Klaus Greff et al. "LSTM: A Search Space Odyssey". In: *CoRR* abs/1503.04069 (2015).

[189] James Hong and Michael Fang. *Sentiment Analysis with Deeply Learned Distributed Representations of Variable Length Texts*. Tech. rep. Stanford University, 2015.

[190] James Hong and Michael Fang. "Sentiment Analysis with Deeply Learned Distributed Representations of Variable Length Texts". In: *Stanford* (2015).

[191] Aamera ZH Khan, Mohammad Atique, and VM Thakare. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis". In: *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)* (2015), p. 89.

[192] Sarah Kohail. "Unsupervised Topic-Specific Domain Dependency Graphs for Aspect Identification in Sentiment Analysis." In: *RANLP*. 2015, pp. 16–23.

[193] Vivek Kulkarni et al. "Statistically significant detection of linguistic change". In: *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2015, pp. 625–635.

[194] Richard Lawson. *Web Scraping with Python*. Packt Publishing Ltd, 2015.

[195] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. "Deep Learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[196] Hongmin Li et al. "Twitter Mining for Disaster Response: A Domain Adaptation Approach." In: *ISCRAM*. 2015.

[197] Bing Liu. *Sentiment Analysis — Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2015.

[198] Elliot Marx and Zachary Yellin-Flaherty. "Aspect Specific Sentiment Analysis of Unstructured Online Reviews". In: (2015).

[199] Tristan Miller, Darina Benikova, and Sallam Abualhaija. "GermEval 2015: LexSub–A shared task for German-language lexical substitution". In: *Proceedings of GermEval* (2015), pp. 1–9.

[200] Sara Rosenthal and Veselin Stoyanov. "SemEval-2015 Task 10: Sentiment Analysis in Twitter." In: *SemEval*. 2015.

[201] Amir Sadeghian and Ali Reza Sharafat. "Bag of Words Meets Bags of Popcorn". In: *web* (2015).

[202] Kartik Singhal, Basant Agrawal, and Namita Mittal. "Modeling Indian general elections: sentiment analysis of political Twitter data". In: *Information Systems Design and Intelligent Applications*. Springer, 2015, pp. 469–477.

[203] Gang Wang et al. "Crowds on wall street: Extracting value from collaborative investing platforms". In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM. 2015, pp. 17–30.

[204] Hao Wang and Jorge A Castanon. "Sentiment expression via emoticons on social media". In: *arXiv preprint arXiv:1511.02556* (2015).

[205] Haibing Wu et al. "Aspect-based Opinion Summarization with Convolutional Neural Networks". In: *CoRR* abs/1511.09128 (2015). arXiv: 1511.09128.

[206] Min Xiao and Yuhong Guo. "Feature space independent semi-supervised domain adaptation via kernel matching". In: *IEEE transactions on pattern analysis and machine intelligence* 37.1 (2015), pp. 54–66.

[207] Alex Marino Goncalves de Almeida, Sylvio Barbon, and Emerson Cabrera Paraiso. "Multi-class Emotions Classification by Sentic Levels as Features in Sentiment Analysis". In: *Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on*. IEEE. 2016, pp. 486–491.

[208] Edgar Altszyler et al. "Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database". In: *arXiv preprint arXiv:1610.01520* (2016).

[209] Danushka Bollegala, Tingting Mu, and John Yannis Goulermas. "Cross-domain sentiment classification using sentiment sensitive embeddings". In: *IEEE Transactions on Knowledge and Data Engineering* 28.2 (2016), pp. 398–410.

[210] Chris Chatfield. *The analysis of time series: an introduction*. CRC press, 2016.

[211] *Exit Polls 2016*. https://edition.cnn.com/election/2016/results/exit-polls. Nov. 2016.

[212] Qiyuan Gao. "Stock market forecasting using recurrent neural network". PhD thesis. University of Missouri–Columbia, 2016.

[213] Anastasia Giachanou and Fabio Crestani. "Tracking sentiment by time series analysis". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM. 2016, pp. 1037–1040.

[214] William L. Hamilton et al. "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Austin, TX, USA, 2016, pp. 595–605.

[215] Soufian Jebbara and Philipp Cimiano. "Aspect-Based Sentiment Analysis Using a Two-Step Neural Network Architecture". In: *Semantic Web Evaluation Challenge*. Springer. 2016, pp. 153–167.

[216] Mirko Lai et al. "Friends and enemies of clinton and trump: using context for detecting stance in political tweets". In: *Mexican International Conference on Artificial Intelligence*. Springer. 2016, pp. 155–168.

[217] Zihang Liu et al. "Discovering Opinion Changes in Online Reviews via Learning Fine-Grained Sentiments". In: *Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on*. IEEE. 2016, pp. 1–10.

[218] Marco Loog. "Contrastive Pessimistic Likelihood Estimation for Semi-Supervised Classification". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 38.3 (2016), pp. 462–475.

[219] Mika Viking Mäntylä, Daniel Graziotin, and Miikka Kuutila. "The Evolution of Sentiment Analysis-A Review of Research Topics, Venues, and Top Cited Papers". In: *arXiv preprint arXiv:1612.01556* (2016).

[220] Saif Mohammad et al. "SemEval-2016 Task 6: Detecting Stance in Tweets". In: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 2016, pp. 31–41.

[221] Shannon M Monnat. "Deaths of despair and support for Trump in the 2016 Presidential Election". In: *Pennsylvania State University Department of Agricultural Economics Research Brief* (2016).

[222] Preslav Nakov et al. "SemEval-2016 Task 4: Sentiment Analysis in Twitter." In: *SemEval@ NAACL-HLT*. 2016, pp. 1–18.

[223] Filipe N Ribeiro et al. "Sentibench — A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods". In: *EPJ Data Science* 5.1 (2016), pp. 1–29.

[224] Sascha Rothe, Sebastian Ebert, and Hinrich Schütze. "Ultradense Word Embeddings by Orthogonal Transformation". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL)*. San Diego, CA, USA, 2016, pp. 767–777.

[225] Harald Sack et al., eds. *Semantic Web Challenges - Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*. Vol. 641. Communications in Computer and Information Science. Springer, 2016.

[226] Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. "Detecting stance in tweets and analyzing its interaction with sentiment". In: *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*. 2016, pp. 159–169.

[227] Ignasi Vegas, Tina Tian, and Wei Xiong. "Characterizing the 2016 US Presidential Campaign using Twitter Data". In: *International Journal of Advanced Computer Science and Applications* 7.10 (2016).

[228] Yu Wang, Yuncheng Li, and Jiebo Luo. "Deciphering the 2016 US Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists." In: *ICWSM*. 2016, pp. 723–726.

[229] Wan Wei et al. "pkudblab at semeval-2016 task 6: A specific convolutional neural network system for effective stance detection". In: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 2016, pp. 384–388.

[230] Mathieu Cliche. "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs". In: *arXiv preprint arXiv:1704.06125* (2017).

[231] Mathieu Cliche. "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs". In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval@ACL 2017)*. Vancouver, Canada, 2017, pp. 573–580.

[232] Jeffrey Dastin. *Amazon trounces rivals in battle of the shopping 'bots'*. May 2017. URL: https://www.reuters.com/article/us-amazon-com-bots-insight/amazon-trounces-rivals-in-battle-of-the-shopping-bots-idUSKBN1860FK.

[233] Jacob Eisenstein. "Unsupervised Learning for Lexicon-Based Classification". In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*. San Francisco, CA, USA, 2017, pp. 3188–3194.

[234] Yoav Goldberg. "Neural Network Methods for Natural Language Processing". In: *Synthesis Lectures on Human Language Technologies* 10.1 (2017), pp. 1–309.

[235] Xin-Yao Qian and Shan Gao. “Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods”. In: *arXiv preprint arXiv:1706.00948* (2017).

[236] Alec Radford, Rafal Józefowicz, and Ilya Sutskever. “Learning to Generate Reviews and Discovering Sentiment”. In: *CoRR* abs/1704.01444 (2017).

[237] Salvador Rodriguez. *U.S. judge says LinkedIn cannot block startup from public profile data*. Aug. 2017. URL: https://uk.reuters.com/article/us-microsoft-linkedin-ruling/u-s-judge-says-linkedin-cannot-block-startup-from-public-profile-data-idUKKCN1AU2BV.

[238] Sara Rosenthal, Noura Farra, and Preslav Nakov. “SemEval-2017 Task 4: Sentiment Analysis in Twitter”. In: *Proceedings of the 11th International Workshop on Semantic Evaluation*. SemEval ’17. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017.

[239] Mariona Taulé et al. “Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017”. In: *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017*. Vol. 1881. CEUR-WS. 2017, pp. 157–177.

[240] Shuting Zhi et al. “Aspects Opinion Mining Based on Word Embedding and Dependency Parsing”. In: *Proceedings of the International Conference on Advances in Image Processing*. ICAIP 2017. Bangkok, Thailand: ACM, 2017, pp. 210–215. ISBN: 978-1-4503-5295-6. DOI: 10.1145/3133264.3133305.

[241] Mirko Lai et al. “Stance Evolution and Twitter Interactions in an Italian Political Debate”. In: *International Conference on Applications of Natural Language to Information Systems*. Springer. 2018, pp. 15–27.

[242] Dilin Liu and Lei Lei. “The appeal to political sentiment: An analysis of Donald Trump’s and Hillary Clinton’s speech themes and discourse strategies in the 2016 US presidential election”. In: *Discourse, Context & Media* (2018).

[243] Narges Tabari et al. “Causality Analysis of Twitter Sentiments and Stock Market Returns”. In: *Proceedings of the First Workshop on Economics and Natural Language Processing*. 2018, pp. 11–19.

[244] Frank Z. Xing, Erik Cambria, and Roy E. Welsch. “Intelligent Bayesian Asset Allocation via Market Sentiment Views”. In: *IEEE Computational Intelligence Magazine* (2018).

[245] Frank Z. Xing, Erik Cambria, and Roy E. Welsch. “Natural Language based Financial Forecasting: A Survey”. In: *Artificial Intelligence Review* 50.1 (2018), pp. 49–73.

[246] Hu Xu et al. “Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction”. In: *arXiv preprint arXiv:1805.04601* (2018).

[247] Philip J. Stone et al. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press, 1966.