



# Opinion mining in management research: the state of the art and the way forward

Soumya Mukhopadhyay<sup>1</sup> 

Accepted: 27 December 2017 / Published online: 15 January 2018  
© Operational Research Society of India 2018

**Abstract** In the past decade, the explosive growth of social media has led to the emergence of a wide variety of information sources that can significantly impact individual level decision making processes. This has resulted in an increasing availability of unstructured textual data and automated evaluation of opinions, attitudes, and emotions has been accepted as an indispensable analytical tool in diverse domains. Consequently, there is a strong need to understand the underlying technical aspects of this emerging new field of analysis. In the current paper we address this need by reviewing the state of the art in sentiment analysis, summarize some of the important recent applications of sentiment analysis and offer future directions for further research. This paper differs from earlier reviews in a number of ways: first, it offers preliminary technical exposition of various techniques following a simple classification scheme so as to help potential future users to develop overall understanding of this rapidly developing field; second, it discusses in greater detail some of the more recently proposed techniques to solve a set of problems in specific management domains; third, it also presents some examples to elucidate how combining sentiment analysis techniques with conventional econometric approaches can help us solve business specific problems. The main goal of this paper is to generate more interest about this interesting new domain among management researchers.

**Keywords** Pre-processing · Sentiment analysis · Financial sentiment · Customer sentiment

---

✉ Soumya Mukhopadhyay  
soumyam@iima.ac.in

<sup>1</sup> Indian Institute of Management, Ahmedabad 380015, India

## 1 Introduction

With the advent of online technologies and social media, individuals are increasingly sharing their views and opinions through the Internet. Consequently, a wide variety of information sources are influencing individual level decision making process, and this in turn is affecting their behavior in business, socio-political and personal contexts. The increasing impact of public opinion is being felt by policy makers and business managers alike. As a result, monitoring and understanding the opinion dynamics and taking appropriate corrective actions are emerging as dominant activities that can make or mar the future of any entity having a public interface. The past one and half decades have witnessed the emergence of a number of techniques to detect, extract and classify opinions, sentiments and attitudes concerning different topics, from large structured and unstructured textual content. These techniques, often called sentiment analysis or opinion mining, essentially focus on offering automated means to study of people's opinions, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes using computational and statistical approaches [1]. The sheer amount of data available from public sources makes it a formidable task to monitor and analyze the available information through purely manual efforts. Moreover, human intervention also introduces various types of cognitive biases into the analysis process. Objective automated opinion mining tools (e.g. natural language processing, polarity analysis, textual analysis etc.) can help us overcome these inherent cognitive limitations of manual approaches and make effective decisions.

Therefore, it is not surprising that automated sentiment detection systems have emerged as an indispensable tool in diverse domains in order to achieve wide variety of goals like assessment of political mood, development of market intelligence, determination of customer satisfaction, sales and business predictions, determination of financial market sentiments etc. [2–4]. For example, with the growth of e-commerce and review sharing platforms like Amazon, IMDB, epinions.com, TripAdvisor etc., business entities across the globe have started taking active interest in using sentiment analysis and text mining techniques to understand consumers' requirements and design their offerings accordingly. Opinions contributed by consumers in blogs, forums, and product related websites are providing managers an avenue to learn consumers' preferences, market trends and competitors' reactions. Social networks (e.g. Facebook, Twitter etc.) are generating massive amount individual level information about billions of users, which can be used to develop algorithms that can predict behavioral patterns with extreme accuracy. A similar trend is sweeping the socio-political scene where online media is slowly displacing conventional communication channels [5, 6]. Common people are increasingly getting actively involved in issue based discussions and sharing their views in a way that can be accessed by millions of others. Emerging technologies are slowly creating an environment where individual level opinions are increasingly initiating changes at a global scale. Much of the information generated in the online environment contain highly unstructured textual data that often requires context specific treatments. Therefore, there is a need to

understand various approaches that can be used to analyze and utilize such information. In the current paper we attempt to address this need by reviewing the state of the art in sentiment analysis, summarizing the recent trends and offering future directions for further research.

This paper differs from existing reviews [1, 7, 8] in a number of ways. First, it offers preliminary technical exposition of various sentiment analysis techniques following a simple classification scheme so as to help potential future users of these techniques. Second, rather than giving an abstract overview of a large number of approaches this paper discusses in greater detail some of the more recently proposed techniques to solve a set of problems in specific management domains. Third, it also presents some examples to show how combining sentiment analysis techniques with conventional econometric approaches can help us solve business specific problems.

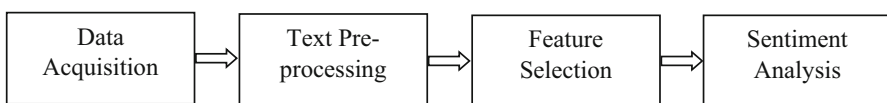
## 2 Sentiment analysis

### 2.1 Definition

The main goal of opinion mining is to extract opinions from unstructured text using algorithmic, statistical or a mixture of both techniques. Sentiment analysis is essentially concerned with the following fundamental elements: the entity or the target that is being evaluated (e.g. a Hotel), the attribute of the target at which the opinion is directed (e.g. service quality or food), the polarity of the opinion (e.g. positive, negative or neutral), the opinion holder (e.g. the individual consumer) and the date when the opinion was emitted. Formally, an opinion is defined as a tuple  $(e_i, a_{ij}, h_k, s_{ijkl}, t_l)$  where  $e_i$  is the  $i$ th entity,  $a_{ij}$  denotes the  $j$ th aspect of entity  $e_i$  at which the opinion is directed,  $h_k$  is the  $k$ th opinion holder,  $s_{ijkl}$  is  $h_k$ 'S opinion polarity (or sentiment) towards aspect  $a_{ij}$  of entity  $e_i$  at time  $t_l$  [9]. An opinionated document  $d \in D$  contains opinions of a set of opinion holder's opinions about a number of entities. Therefore, the main objective of sentiment analysis is to find all the opinion tuples  $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$  in a given document, across a set of documents or across many sets of documents. As stated earlier, the opinion polarity  $s_{ijkl}$  can be generally defined in terms of three levels: positive, negative or neutral.

### 2.2 Process flow

The basic process of sentiment analysis consists of a series of preliminary steps that includes data acquisition, text preprocessing and feature selection. These initial



**Fig. 1** Basic process flow of sentiment mining

steps are followed by actual sentiment classification process. The overall process flow of a typical sentiment or opinion mining is described in Fig. 1.

### 2.2.1 Data acquisition

In the data acquisition phase, the corpus or the text documents are acquired using either an Application Programming Interface (API) or by employing Web Crawlers [10]. In the API based approach, data is collected using a broad range of clients (e.g. browsers and mobile devices) through appropriate HTTP protocols. These data acquisition strategies are easy to implement; however, they may have accessibility limitations (e.g. Twitter REST API imposes a rate limit of 180 requests/15-min). Moreover, many websites do not provide the API interface for public consumption. In contrast, crawler-based approaches offer virtually unrestricted access to topically-relevant content. However, the text collected through crawlers are often noisy and its data structure is prone to change with the change in the design of the target website. Moreover, the design of web crawlers must obey the etiquette protocols by minimizing the frequency and extent of overlap between consecutive requests. These requirements impose a constraint on how much data can be collected within a given span of time. Apart from these two most popular approaches, many firms also generate a large amount of opinionated unstructured text through their customer interfaces. For example, many independent e-Sellers (e.g. Myntra in India), online marketplaces (Flipkart, Amazon etc.) and intermediate aggregators (e.g. TripAdvisor) provide in-built interfaces through which customers can contribute their opinions and reviews about different products. These interfaces can generate a lot of opinionated textual corpus that can be analyzed through sentiment classification in order to generate business relevant knowledge.

### 2.2.2 Text pre-processing

Data acquisition is followed by a text sanitization step that is intended to clean up the text and format it in a manner that can be used for the subsequent steps of feature extraction and sentiment/opinion classification. The usual steps involved in this stage are as follows:

- *Language detection* In this stage the language used in the text is detected and only the documents written in relevant language is identified and extracted from the collection of text documents or corpora [11]. This is important because every language has its own structural characteristics. Consequently, natural language models are generally specific to a particular language and these models require some extent of structural homogeneity for any meaningful analysis to be carried out.
- *Tokenization and lemmatization* Most sentiment classification algorithms operate on a *bag-of-words* assumption whereby the document is represented as sparse vector of occurrence frequencies of its vocabulary (words). This assumption disregards the grammatical structure and word-orders within the document [12]. Accordingly, in the tokenization step, the document is broken

down in terms of all the constituent words. Subsequently, each of these words are converted to its invariant canonical form or “stem” in the lemmatization step. This conversion is usually achieved through a morphological analysis but it can also be achieved through a less rigorous stemming heuristic (e.g. Porter’s stemmer [13]) to remove the word affixes.

- *Stopword removal* Many words (e.g. *the, that, of* etc.) in any language (e.g. determiners, coordinating conjunction and prepositions in English) are used to impose a structure rather than to contribute meaningfully to the underlying expression and emotion. These words are called *stopwords* and they can be safely removed from a document without any significant detrimental impact on final opinion analysis [14, 15]. The process of removing these words is called *stopword removal*. This step can reduce the computational resources needed to analyze the corpora.
- *Part-of-speech (POS) tagging* In this step each word is labeled with its associated part of speech (i.e. noun, verb, adjective, adverb and preposition). POS tagging is often useful for further processing like dependency parsing or machine learning. The identification of adjectives and adverbs are also sometimes useful for determining opinion polarity and subjectivity.

As the process flow suggests, the text pre-processing step is followed by feature or aspect extraction and selection step. This step is presented below with a specific focus on the approaches that have been most frequently used in past research works.

### 2.2.3 Feature extraction/selection

The definition of sentiment classification (presented in Sect. 2.1) suggests that the main task of this process is to identify the polarity of an opinion held by opinion holder ( $h_K$ ) that is targeted towards a specific ( $j$ th) aspect  $a_{ij} \in \{a_{i1}, a_{i2}, \dots, a_{iJ}\}$  of target entity  $e_i$ . Therefore, one of the primary tasks is to identify these aspects. The feature selection or aspect identification process is mainly concerned with this goal. Most of these approaches aim to identify a smaller subset of variables that can efficiently describe the underlying characteristics of the input data. The main approaches of feature selection are as follows.

- Criterion based approaches

In these approaches, words are scored based on a suitable ranking criterion based on their relevance. Subsequently, all words with a score below a predefined threshold are removed. Therefore, in this approach, a feature selection criterion must be determined which can measure the relevance of each word with reference to output classes/labels. One way to define the relevance of a word is in terms of its conditional dependence, i.e. a word that is conditionally independent of the output class labels, can be considered irrelevant. Therefore, an appropriate ranking criterion can be expressed in terms of the interdependence between the class labels and the word under consideration. Accordingly, the most frequently used selection criteria can be based on *correlation, pairwise mutual information* and *Chi square*

estimate [16]. As expected, each of these criteria essentially measures the extent of dependency between the class labels and the target variable or word. The correlation based criterion is defined as follows:

$$R(i) = \frac{\text{cov}(x_i, Y)}{\sqrt{\text{var}(x_i) \cdot \text{var}(Y)}}, \quad (1)$$

where,  $x_i$  is the  $i$ th variable (word),  $Y$  is the output (class labels),  $\text{cov}()$  indicates the covariance and  $\text{var}()$  denotes the variance. The next important criterion for feature selection approach uses is Chi square statistic as a measure of dependency between the class labels and the target word. The Chi square statistic of the word between word  $x$  and class  $k$  is defined as:

$$\chi^2 = \frac{n \cdot F(x)^2 \cdot (p_k(x) - P_k)^2}{F(x) \cdot (1 - F(x)) \cdot P_k \cdot (1 - P_k)}, \quad (2)$$

where  $n$ , Total number of documents;  $p_k(x)$ , The conditional probability of class  $k$  for documents containing word  $x$ ;  $P_k$ , Overall fraction of documents containing class  $k$ ;  $F(x)$ , Overall fraction documents containing word  $x$ .

The pairwise mutual criteria also measure mutual dependency between two variables in an information theoretic manner. Considering the Shannon entropy or the uncertainty associated with the output  $Y$  is defined as  $H(Y) = -\sum p(y) \log(p(y))$ , the corresponding conditional entropy can be expressed as  $H(Y|X) = -\sum_x \sum_y p(x, y) \log(p(y|x))$ . Consequently, we can measure the extent of uncertainty reduction ( $I(Y, X)$ ) in output  $Y$  by observing variable  $X$  as  $I(Y, X) = H(Y) - H(Y|X)$ . This provides a way to estimate the point-wise mutual information ( $M_k(x)$ ) as the ratio of true co-occurrences ( $F(x) \cdot p_k(x)$ ) of class  $k$  and word  $x$  and their expected mutual co-occurrences ( $P_k(x) \cdot F(x)$ ). Hence, we can write:

$$M_k(x) = \log\left(\frac{p_k(x)}{P_k(x)}\right). \quad (3)$$

The point-wise mutual information varies between  $-1$  to  $1$  depending on the extent of dependency between  $Y$  and  $x$ . A distance based measure of mutual information can also be derived using the Kullback–Leibler divergence between two densities defined by the probability density functions  $f(\cdot)$  and  $g(\cdot)$ :

$$K(f, g) = f(y) \log\left(\frac{f(y)}{g(y)}\right). \quad (4)$$

The ranking criterion based approach of feature selection is computationally less demanding and it does not suffer from the problem of overfitting. However, it suffers from the problem of redundancy, in that it often identifies subsets with more than optimum number of variables that can adequately describe the underlying characteristics of the input data.

- Latent semantic analysis

Latent semantic analysis is an unsupervised learning technique that aims to uncover underlying similarity among structures by first creating a rectangular term-document matrix ( $X_{t \times d}$ ) from a large collection of text where the rows represent individual words, columns represent the document and individual cells show the frequency with which a specific term occurs in a document. Subsequently, these frequencies are transformed into an inverse document frequency or entropy-based score and a reduced-rank or truncated singular value decomposition (i.e.  $X \sim T_k \times S_k \times D_k^T$ ) is applied on this matrix. The  $k$  largest singular values and their associated vectors generated by the Singular Value Decomposition (SVD) process are retained so as to represent each document-term as a  $k$ -dimensional vector in the derived space. Specifically, the rows in  $T_k$  represent the term vectors and the rows in  $D_k$  represent the document vectors in a reduced latent semantic space. Finally, similarities among entities (e.g. document–document, term–term and term–document) are computed in this reduced-dimensional space. In effect, this method transforms the text space to a new axis system that can explain the variations in the underlying attributes values in terms of a linear combination of the original word features. The main disadvantage of latent semantic analysis is that it may not necessarily discover those features that would lead to the best separation of the underlying document class-distributions.

#### 2.2.4 Sentiment analysis

Once the data has been collected, pre-processed and a set of appropriate features or aspects have been identified, sentiment analysis can be applied to find out the opinion polarity  $s_{ijk}$ . Sentiment analysis is essentially a process of classifying a given text into two or more (e.g. “positive/negative” or “thumbs up/thumbs down”) opinion categories. It can also take the form of ordinal outputs such as number “stars” for a product etc. Moreover, the opinion determination can happen at various levels of granularity such as words, sentences and documents. Sentiment classification can be performed using either *lexicon-based* based or *machine learning* (ML) approaches. The *Lexicon-based* approach requires human interventions to create a set of annotated seed words that can be used in a subsequent bootstrapping method that relies on synonym detection algorithms to create a larger lexicon. This collection of known and precompiled sentiment terms called a *sentiment lexicon*. Subsequently, either the strength or the probability of occurrence of a sentiment word can be used for sentiment classification. The first approach is called a *dictionary-based approach* while the second type of methods are called *corpus-based approach*. However, it must be noted that, the much needed manual intervention makes lexical based approaches costlier and less appropriate for most of the large-scale sentiment mining exercises. The Machine Learning (ML) approach, on the other hand, uses linguistic features of the text and it can further be categorised based on whether it uses annotated data for training of the classifier (*Supervised ML*) or not (*Unsupervised ML*). Following Medhat et al. [17], a broad description of available classification schemes is presented in Fig. 2. We describe each of these approaches in greater detail in subsequent subsections.

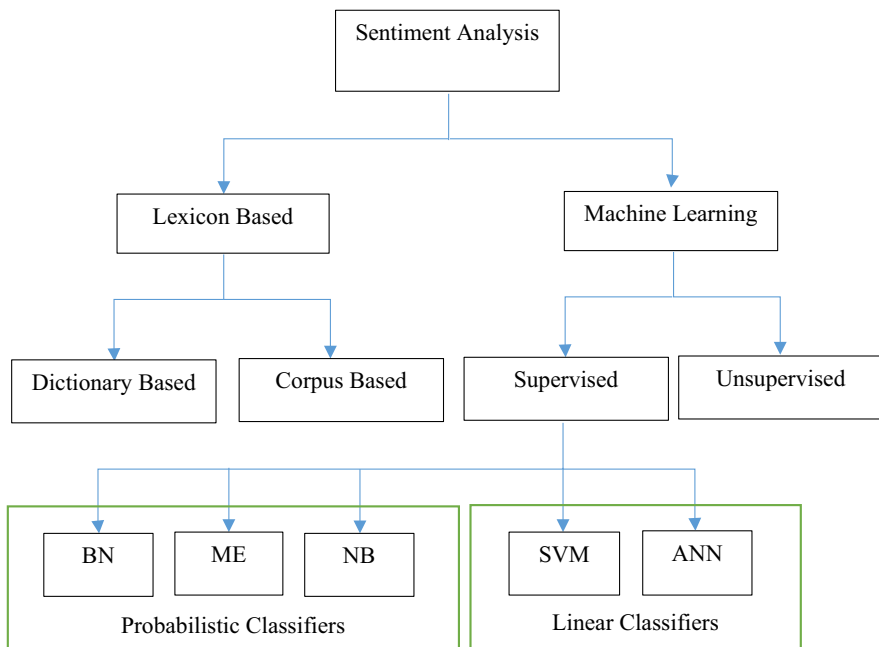
- Machine learning based sentiment classification

*Supervised ML approaches* The supervised machine learning sentiment classifiers can be categorized into two broad classes: linear classifiers and probabilistic classifiers. Under the linear classifier category, we have *support vector machine* and *neural network* based approaches. On the other hand, under *probabilistic classifier* category, there are three major approaches available: *Naïve Bayes classifiers*, *Bayesian networks* and *maximum entropy based classifiers*.

*Support vector machines (SVM)* SVM's were developed from Statistical Learning Theory [18]. This is essentially a class of linear algorithms that tries to find out a hyperplane that can optimally separate out the data into two more classes. For a given  $n$ -dimensional input vector  $\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , a weight vector  $\vec{w} = (w_1, w_2, \dots, w_n)$  and an output value  $y_i$ , the derived hyperplane can be defined as:

$$\vec{w} \cdot \vec{x}_i - b = 0. \quad (5)$$

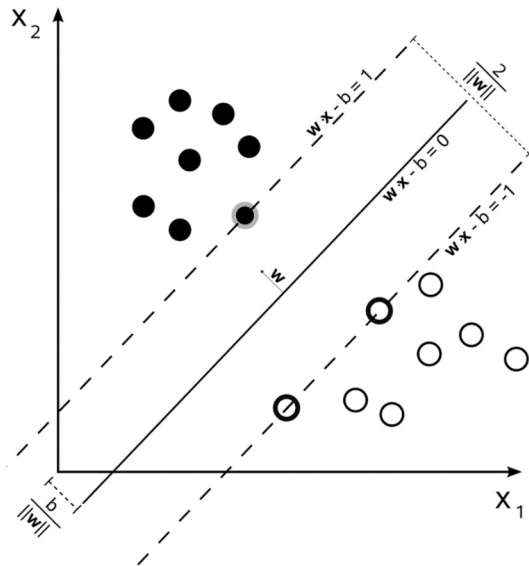
The weight vector  $\vec{w} = (w_1, w_2, \dots, w_n)$  is determined using an appropriate training process, and given these weights, the classification of a new input vector  $\vec{x}_i$  can be carried out based on whether  $\vec{w} \cdot \vec{x}_i - b \geq 0$  or not. We want to maximize the distance between the parallel hyperplanes (or margins) by choosing appropriate values of  $w$  and  $b$  (Fig. 3). These hyperplanes can be described by the equations:



**Fig. 2** Approaches to sentiment analysis. (Reproduced with permission from Medhat et al. [17])



**Fig. 3** Support vector machine. (adapted from wikipedia)



$$\vec{w} \cdot \vec{x}_i - b = 1, \text{ and} \tag{6}$$

$$\vec{w} \cdot \vec{x}_i - b = -1$$

In order to have sufficient separation between the identified classes, we additionally impose the following constraint:

$$\vec{w} \cdot \vec{x}_i - b \begin{cases} \geq 1 & \text{if } y_i = +1 \\ \leq -1 & \text{if } y_i = -1 \end{cases} \tag{7}$$

Learning the SVM can now be formulated as the quadratic optimization problem subject to linear constraints where there is a unique minimum:

$$\min_w \frac{1}{2} \|w\|^2 \text{ subject to } y_i(\vec{w} \cdot \vec{x}_i) - b \geq 1 \text{ for } i = 1, 2, \dots, N. \tag{8}$$

Considering a corresponding Lagrangian multiplier  $\alpha_i$ , the necessary conditions for this optimization process are  $\frac{\partial L(\alpha, w, b)}{\partial w} = w - \sum \alpha_i y_i x_i = 0$  and  $\frac{\partial L(\alpha, w, b)}{\partial b} = \sum \alpha_i y_i = 0$ . Therefore, we get:

$$\min_{w, b} \{L(\alpha, w, b)\} = \sum \alpha_i - \frac{1}{2} \sum_{i, j} \alpha_i \alpha_j y_i y_j x_i^T x_j = \tilde{L}(\alpha). \tag{9}$$

Therefore, we need to find  $\max_{\alpha} \tilde{L}(\alpha)$  subject to following conditions:  $\alpha_i \geq 0$  and  $\sum \alpha_i y_i = 0$ . The corresponding values of  $w$  is given by  $\sum \alpha_i y_i x_i$  and  $b = w \cdot x_i - y_i$  when  $\alpha_i \geq 0$ . The classical SVM system assumes that a single hyperplane that the dataset is linearly separable. For non-linear datasets, a Kernel function can be used

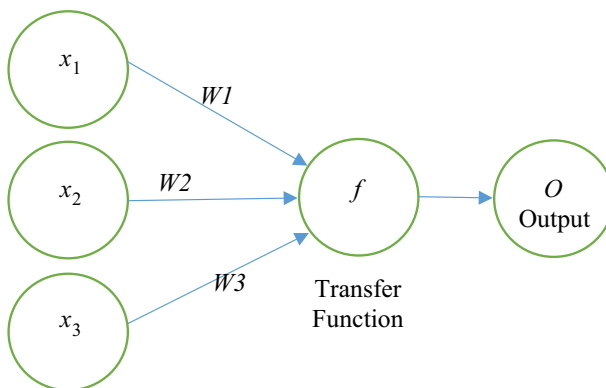
to map the data to a higher dimensional space in which it is linearly separable. Subsequently, the classical SVM system can be used to construct a hyperplane in this higher dimensional feature space. In other words, the product  $x_i^T x_j$  in Eq. (9) can be replaced by an appropriate kernel function  $K(x_i, x_j)$  such as a polynomial kernel ( $K(x_i, x_j) = (x_i^T x_j + 1)^p$  for  $p > 0$ ) or a Gaussian RBF kernel ( $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$  for  $\gamma > 0$ ). Classical binary SVM systems can be also extended to handle multi-class classification problems as well. For further understanding the reader is encouraged to read some of classic texts in this area [18, 19].

**Artificial neural networks (ANN)** The fundamental concept of artificial neural networks (ANN) was developed by McCulloch and Pitts [20]. At a very basic level, an artificial neural network is represented by a number of interconnected nodes (*neurons*) that can take a set of inputs, process these inputs and generate a set of outputs (Fig. 4). The basic processing unit of these networks is a *neuron* that can generate neural impulses taking weighted sums of the input signals and transforming them by applying a transfer function ( $f$ ). The processing of the input signals (stimuli) depends on the respective connection weights. The transfer function accommodates for any possible nonlinearities. The learning ability in this setup is achieved by modulating the weights through some predefined algorithms called learning algorithms.

Referring to Fig. 4, the unidirectional the signal flow from inputs  $x_1, x_2, \dots, x_n$  leads to a neuron's output signal flow (O) that is generated as follows:

$$O = f\left(\sum_{i=1}^n w_i x_i\right), \quad (10)$$

where  $w_i$  is a weight vector and  $f(\cdot)$  is the transfer function. The output  $O$  is classified according to a threshold ( $\theta$ ) based rule. Accordingly, we can define:



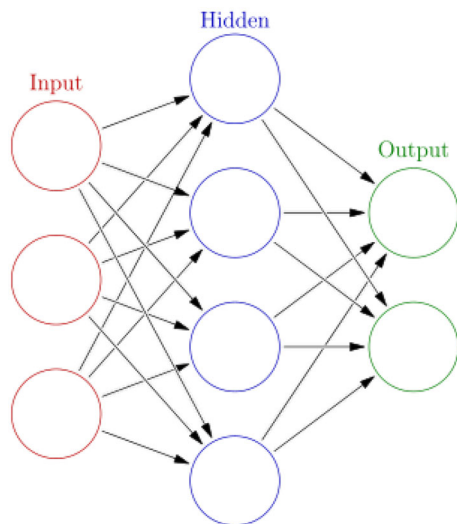
**Fig. 4** Mechanism of artificial neuron

$$O = f\left(\sum_{i=1}^n w_i x_i\right) \begin{cases} = 0, & \text{if } w^T x \geq \theta \\ = 1, & \text{if } w^T x < \theta \end{cases} \tag{11}$$

Extending this idea of the basic processing unit (or *neuron*), the overall architecture of a neural network can be defined in terms of three interconnected layers: *input layer*, *hidden layer* and *output layer* (Fig. 5). In case of feed-forward networks, the signal flows strictly in the forward direction (i.e. from input to output units) only. Recurrent networks on the other hand allows feedback connections where signals can flow backwards. In most of the neural networks the connection weights are updated through different learning approaches: supervised, unsupervised and reinforcement learning. In supervised learning the network is fed with an input vector and a set of desired responses. Subsequently, the errors between the desired and actual response (produced by the neural network) for each node in the output layer are used to modify the connection weights. The best known examples of ANN are the *perceptron algorithm*, the *delta rule*, and the *backpropagation algorithm*. In case of unsupervised learning approach, instead of using a priori set of desired outputs, the ANN automatically discovers statistically salient features of the input vector. Finally, in case of reinforcement learning, the system learns through trials and identifies those actions that would maximize a reward signal. Once trained, the learned rules defined in terms of these estimated weights can be used to predict document class memberships.

*Naïve Bayes classifier (NB)* The Naïve Bayes classifier is the simplest and most commonly used classifier. Bayesian inference was first applied in text classification by Mosteller and Wallace [21]. Naïve Bayes classification model works with the Bag-of-Words (BOW) assumption that ignores the position of the word in the document. Given a document  $d$  and a set of classes  $c \in C$ , this model tries to find out the class that has maximum posterior probability. Following the Bayes Rule, the

**Fig. 5** Artificial neural network



fundamental problem of a NB classifier is to find out the most likely class that a document belongs to:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(c|d) = \operatorname{argmax}_{c \in C} \frac{P(d|c)P(c)}{P(d)} = \operatorname{argmax}_{c \in C} P(d|c)P(c). \quad (12)$$

A document  $d$  can be assumed to be defined by a set of key words  $w_1, w_2, \dots, w_n$ . Further assuming that the probabilities  $P(w_i|c)$  are conditionally independent given the class  $c$ , the probability a document belongs to a class  $c$  is given by the class probability multiplied by the product of the conditional probabilities of each word for that class:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(w_1, w_2, \dots, w_n|c)P(c) = \operatorname{argmax}_{c \in C} P(c) \prod_w P(w|c). \quad (13)$$

In other words, we can write:

$$P(c) \prod_i P(w_i|c) = P(c) \prod_i^n \frac{\operatorname{count}(w_i, c)}{\sum_i \operatorname{count}(w_i, c)} = P(c) \prod_i \frac{\operatorname{count}(w_i, c)}{V_c}. \quad (14)$$

Here  $\operatorname{count}(w_i, c)$  is the number of occurrences of word  $w_i$  in class  $c$ ,  $V_c$  is the total number of words in class  $c$  and  $n$  is the number of words in the target document. Now,  $V_c$  being a constant for given training set, it can be taken outside to get  $P(c) \prod_i \frac{\operatorname{count}(w_i, c)}{V_c} = \frac{P(c)}{V_c^n} \prod_i^n \operatorname{count}(w_i, c)$ . For any word absent from the training set, the conditional probability of the word can be replaced by 1. Based on Eq. (13), the maximum a posteriori decision rule or can be developed to assign a class label to each document based on their respective word frequencies.

*Maximum entropy (ME) classifier* The maximum entropy based approach is closely related to the Naïve Bayesian classification. In case of ME classifier, an indicator function (or *joint feature*) is defined for each word  $w$  and class  $c$ :

$$f_i(d, c) = \begin{cases} 1, & \text{if } d \text{ belongs to class } c \text{ and contains word } w \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

The expected value of feature  $f_i$  with respect to the model  $p(d|c)$  is given by:

$$P(f_i) = \sum_{c,d} \tilde{p}(d)p(d|c)f_i(d, c) = \sum_{c,d} \tilde{p}(d, c)f_i(d, c). \quad (16)$$

where  $\tilde{p}(d, c)$  is the empirical distribution of the training data and it is equal to  $\tilde{p}(d, c) = \frac{1}{N}(\eta)$ ,  $\eta$  being the number of times  $(d, c)$  occurs in training dataset. Subsequently, a weight ( $\lambda_i$ ) is assigned to each of these joint features so as to maximize the log-likelihood of the training data. This weight assignment is carried out by using an iterative optimization algorithm. Given the feature weight vector ( $\lambda$ ), the probability that a given document  $d$  belongs to class  $c$  is given by:

$$p(c|d, \lambda) = \frac{\exp(\sum_i \lambda_i f_i(d, c))}{\sum_c \exp(\sum_i \lambda_i f_i(d, c))}. \quad (17)$$

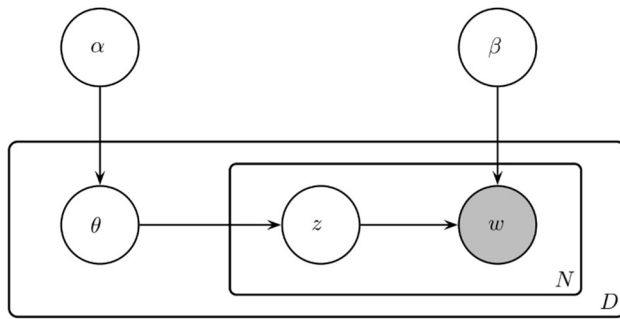
Once again, just like a Naïve Bayes classifier, this probability rule can help classify a given document by indicating its most probable class membership.

*Bayesian networks (BN)* In contrary to Naïve Bayes or Maximum Entropy classifiers, the Bayesian network model aims to capture the complete relationship structure among a set of variables in terms of their conditional dependencies and specify a complete joint probability distribution over all the variables. In general, given a Bayesian network with  $n$  nodes  $x_1, x_2, \dots, x_n$ , the corresponding joint probability distribution is given by:  $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | P_a(x_i))$ , where  $P_a(x_i)$  is the set of the probability distributions corresponding to the parents of node  $x_i$ . Bayesian Networks can often be represented adequately using directed acyclic graphs. However, complete probabilistic treatments can prove to be computationally expensive for most the real-life sentiment mining problems. Although, directional separation or d-separation and Markov property assumptions can alleviate this issue to an extent, BN is still used very infrequently in opinion mining.

*Unsupervised sentiment analysis (Bayesian topic sentiment models)* It is evident from our discussion till now that sentiment analysis aims to classify documents into a set of predefined categories. Supervised techniques archive this by using a large number of pre-annotated training texts. However, creating pre-annotated training documents is an expensive and time consuming task. Moreover, documents created through human intervention inherently introduce biases into the training process. Moreover, many domain specific sentiment models often fail to produce satisfactory results in a different domain. Unsupervised learning tries to address these issues by minimizing dependency on annotated training data. However, using unsupervised approaches for sentiment analysis is a challenging task because the predictive accuracy of these models barely beat the chance baseline. Recently Bayesian approaches to unsupervised sentiment analysis have received a lot of attention, because these models allow the inclusion of prior information into the model estimation and can improve the prediction accuracy to a significant extent. Most of the Bayesian unsupervised sentiment analysis models rely on Latent Dirichlet Allocation [22] to jointly detect the underlying topics in a document and identify whether the semantic orientation of the given text is positive, negative or neutral.

Latent Dirichlet Analysis (LDA) is a generative probabilistic framework that automatically discovers hidden topics from the text by considering each document from a corpus to be a mixture of topics. Each topic, in turn, is assumed to follow a distribution over a fixed vocabulary terms. Accordingly, the LDA process can be expressed by a simple plate notation presented in Fig. 6. In this graphical representation  $\alpha$  is a  $k$ -dimensional vector of symmetric Dirichlet priors, where  $k$  is the number of underlying topics,  $\beta$  denotes the conditional probabilities of words to topics,  $\theta$  is a document specific vector of topic probabilities,  $w$  is the observed document specific vector of words and  $z$  is the document specific choice of topics for each word.

Consequently, the generative process can be described as follows:



**Fig. 6** Latent Dirichlet allocation. The generative process. (Reproduced with permission from Blei et al. [22])

- For each document  $d \in D$  in collection  $D$ :
  - Draw  $\theta \sim \text{Dirichlet}(\alpha)$ .
  
- For each word ( $w_n$ ) in a specific document  $d$  of length  $N$ :
  - Draw a topic  $z_n \sim \text{Multinomial}(\theta)$
  - Draw a word  $w_n$  from a multinomial probability  $p(w_n|z_n, \beta)$  conditioned on the topic  $z_n$ .

This generative process leads to a joint probability of  $\theta^d, \mathbf{z}^d$  and  $\mathbf{w}^d$  given parameters  $\alpha$  and  $\beta$ :

$$p(\theta^d, \mathbf{z}^d, \mathbf{w}^d | \alpha, \beta) = p(\theta^d | \alpha) \prod_{n=1}^N p(z_n^d | \theta^d) p(w_n^d | z_n^d; \beta) \tag{18}$$

Marginalizing out  $z$  and  $\theta$  leads to the probability of words in a document:

$$p(\mathbf{w}^d | \alpha, \beta) = \int p(\theta^d | \alpha) \prod_{n=1}^N \sum_{z_n^d} p(z_n^d | \theta^d) p(w_n^d | z_n^d; \beta) d\theta^d. \tag{19}$$

Consequently, the probability of all documents in the corpus is given by the product of marginal distributions:

$$p(D | \alpha, \beta) = \prod_{d=1}^D \int p(\theta^d | \alpha) \prod_{n=1}^N \sum_{z_n^d} p(z_n^d | \theta^d) p(w_n^d | z_n^d; \beta) d\theta^d. \tag{20}$$

Now, the posterior of the hidden variables  $\theta^d$  and  $\mathbf{z}^d$  for a given document  $d$  can be expressed in terms of Eqs. (18) and (19)

$$p(\boldsymbol{\theta}^d, \mathbf{z}^d | \mathbf{w}^d, \alpha, \beta) = \frac{p(\boldsymbol{\theta}^d, \mathbf{z}^d, \mathbf{w}^d | \alpha, \beta)}{p(\mathbf{w}^d | \alpha, \beta)}. \tag{21}$$

However, the denominator of the posterior is not tractable as both variables  $\boldsymbol{\theta}^d$  and  $\mathbf{z}^d$  are latent in nature and consequently the posterior expectation cannot be calculated in a straight forward manner. Researchers have proposed a number of ways to solve this problem beginning with variational Bayes inference, where the generative model is expressed in terms of a simpler distribution without many dependencies by removing the edges between  $\theta$ ,  $z$  and  $w$  (Fig. 7).

Based on this simplified notation, the approximate posterior distribution takes the following form:

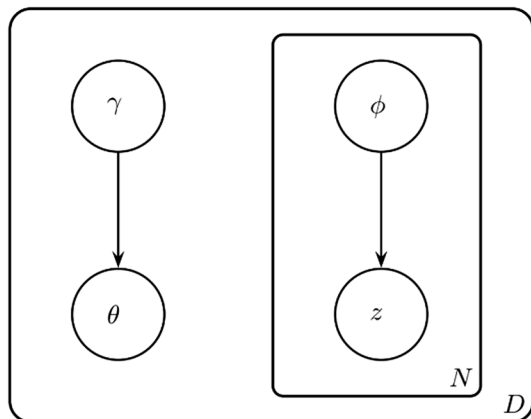
$$q^d(\theta^d, \mathbf{z}^d | \gamma^d, \phi^d) = q(\theta^d | \gamma^d) \prod_{n=1}^N q(z^n_d | \phi^n_d), \tag{22}$$

where  $q(\cdot)$  denotes an approximate posterior function and  $\phi$  denotes the variational parameters. Using an EM (Expectation Maximization) algorithm the posterior can now be calculated by iterating through the following alternating steps:

1. E(Expectation)-step: Find out the best approximate posterior function  $q^d(\theta^d, \mathbf{z}^d | \gamma^d, \phi^d)$  for each document
2. M(Maximization)-step: Maximize the bounds of  $q(\cdot)$  function with respect to  $\alpha$  and  $\beta$

Although variational EM is the most frequently used algorithm for parameter estimation in any model related to LDA, still in reality it is an approximation method. Consequently, Gibbs based samplers have been proposed recently for estimating LDA [23]. Gibbs approach is a Markov chain Monte Carlo (MCMC) algorithm for sampling a sequence of approximate observations from a specified joint probability distribution of two or more random variables  $p(\mathbf{z}) = p(z_1, z_2, \dots, z_n)$ . The process begins by initializing state values for

**Fig. 7** Latent Dirichlet allocation. The variational inference. (Reproduced with permission from Blei et al. [22])



$\{z_i : i = 1, \dots, N\}$  and then iterating through a sampling process where each variable  $z_i$  is sampled from its conditional distribution with reference to the remaining variables i.e.  $p(z_i|z_{-i})$ . This procedure is repeated a number of times until the samples begin to converge to the true target distribution.

The Gibbs sampling approach to LDA aims to find the latent document specific topic portions ( $\theta_d$ ), the topic specific word distributions ( $\varphi$ ), and the probability of topic  $z$  being assigned to a word  $w_i$  denoted by  $z_i$ . However, these topic index assignments ( $z_i$ ) are sufficient to find out both  $\theta_d$  and  $\varphi$ . Therefore, in principle these parameters ( $\theta_d$  and  $\varphi$ ) can be integrated out so as to focus on just computing the topic indexes ( $z_i$ ) given all other topic assignments to all other words. This kind of Gibbs sampling scheme is called a *collapsed Gibbs sampler* (with both  $\theta_d$  and  $\varphi$  being “*collapsed out*”). Therefore, denoting all topic allocations other than for  $z_i$  by  $z_{-i}$ , the following posterior probability can be defined up to a normalizing constant:

$$p(z_i|z_{-i}, \alpha, \beta, w) = \frac{p(z_i, \mathbf{z}_{-i}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{z}_{-i}, \mathbf{w}|\alpha, \beta)} \propto p(z_i, \mathbf{z}_{-i}, w|\alpha, \beta) = p(\mathbf{w}, \mathbf{z}|\alpha, \beta). \tag{23}$$

Moreover,

$$\begin{aligned} & p(\mathbf{w}, \mathbf{z}|\alpha, \beta) \\ &= \iint p(\mathbf{z}, \mathbf{w}, \theta, \varphi|\alpha, \beta) d\theta d\varphi = \iint p(\varphi|\beta) p(\theta|\alpha) p(\mathbf{z}|\theta) p(\mathbf{w}|\varphi, \mathbf{z}) d\theta d\varphi \\ &= \int p(\mathbf{z}|\theta) p(\theta|\alpha) d\theta \int p(\mathbf{w}|\varphi, \mathbf{z}) p(\varphi|\beta) d\varphi. \end{aligned} \tag{24}$$

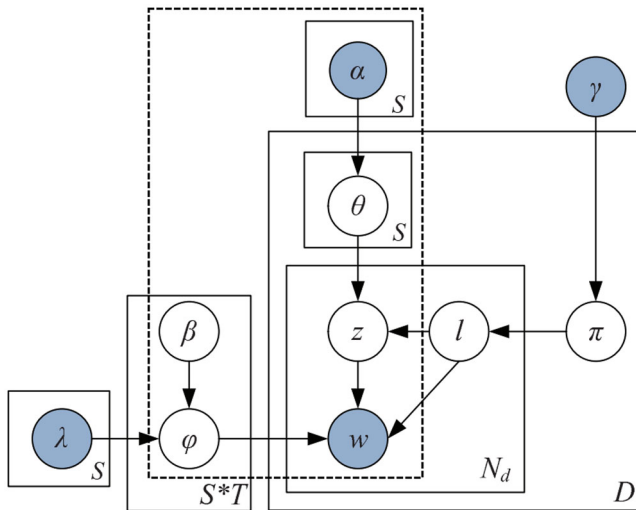
Both the terms in Eq. (24) being multinomial with Dirichlet prior, using conjugacy property it can be shown that

$$p(\mathbf{w}, \mathbf{z}|\alpha, \beta) = \prod_d \frac{B(n_{d\cdot} + \alpha)}{B(\alpha)} \prod_k \frac{B(n_{k\cdot} + \beta)}{B(\beta)}, \text{ and} \tag{25}$$

$$\begin{aligned} p(z_i|\mathbf{z}^{(-i)}, \mathbf{w}) &= \frac{p(\mathbf{w}, \mathbf{z})}{p(\mathbf{w}, \mathbf{z}^{(-i)})} = \frac{p(\mathbf{z})}{p(\mathbf{z}^{(-i)})} \cdot \frac{p(\mathbf{w}|\mathbf{z})}{p(\mathbf{w}^{(-i)}|\mathbf{z}^{(-i)})p(w_i)} \\ &\propto \left( n_{d,k}^{(-i)} + a_k \right) \frac{n_{k,w}^{(-i)} + \beta_w}{\sum_{w'} n_{k,w'}^{(-i)} + \beta_{w'}}, \end{aligned} \tag{26}$$

where  $k$  represents the topic,  $n_{d,k}$  indicates the number of words assigned to topic  $k$  in document  $d$  and  $n_{k,w}$  denotes the number of times word  $w$  is assigned to topic  $k$ . Once the Gibbs sample is finished, the counts can be used to compute the latent distributions  $\theta_d$  and  $\varphi_k$ . This basic framework has been extended in many ways to incorporate sentiment analysis. Joint Sentiment Topic model (JST) is one such framework that tries to model the word generation for positive or negative sentiment conditioned on topics [24]. The plate diagram for JST model is presented in Fig. 8.





**Fig. 8** Joint sentiment/topic model plate diagram. (Reproduced with permission from Lin and He [24])

Compared to some of the existing semi-supervised methods JST shows significant performance gain of between 10% to 20%.

Denoting a collection of  $D$  documents  $(d_1, d_2, \dots, d_D)$  that are each associated with a sequence of words  $(w_1, w_2, \dots, w_{N_d})$  where each word is considered to be from a vocabulary of distinct words  $1, 2, \dots, V$ , JST tries to model  $S$  distinct sentiments and  $T$  topic labels simultaneously. The corresponding generative mechanism can be described as follows [24]:

- For each sentiment label  $l \in \{1, \dots, S\}$ , for each topic  $j \in \{1, \dots, T\}$ , draw  $\varphi_{lj} \sim \text{Dirichlet}(\lambda_l \times \beta_{lj}^T)$ .
- For each document  $d$ , choose a distribution  $\pi_d \sim \text{Dirichlet}(\gamma)$
- For each sentiment label  $l$  under document  $d$ , choose a distribution  $\theta_{d,l} \sim \text{Dirichlet}(\alpha)$ .
- For each word  $w_i$  in document  $d$ 
  - Choose a sentiment label  $l_i \sim \text{Multinomial}(\pi_d)$
  - Choose a topic  $z_i \sim \text{Multinomial}(\theta_{d,l_i})$
  - Choose a word  $w_i$  from a Multinomial distribution over words conditioned on topic  $z_i$  and sentiment label  $l_i$  denoted by  $\varphi_{l_i z_i}$

A very similar approach is also taken by Aspect Sentiment Unification Model (ASUM) that, in contrast to JST, focuses on regional co-occurrence of the words in a document by imposing a constraint that all the words in a given sentence must originate from the same language model [25]. In some sense both ASUM and JST are semi supervised in nature in that they use a small set of sentiment seed words.

There are several other models (e.g. Topic Sentiment Mixture (TSM) model, Multi-Aspect Sentiment (MAS) model etc.) that extend the basic LDA model to model topic and sentiment together [14, 26]. Very recently, researchers have proposed more advanced methods based on text-based hidden Markov models, that can use a sequence of words in training texts instead of a predefined sentiment lexicon to classify implicit opinions [27].

Apart from machine learning based and lexical sentiment classifications, the sentiment classification can be carried out using *dictionary based* and *corpus based* classification methods also. Therefore, for the sake of completeness, we briefly touch upon these two methods and cite some relevant literature.

- Dictionary-based approach

The dictionary based approach begins with small set of seed words collected manually and recursively extends this initial word list by collecting related synonyms and antonyms from appropriate dictionaries e.g. WordNet [28]. This process continues till no new words can be found to be added to the seed word list. However, the main weakness of these family of methods is in their inherent inability to find domain and context specific opinion words. The Corpus-based approach tries to solve the problem of finding context specific opinion words by relying on the syntactic patterns of co-occurrences. In addition to WordNet, there are several other dictionaries which have been developed to examine specific aspects of human sentiment. Most prominent among them are: (1) Harvard General Inquirer [29], the oldest manually constructed word list organized into 17 distinct semantic categories; (2) SenticNet [30], an extension of WordNet consisting of words related to four emotional dimensions (sensitivity, aptitude, attention, and pleasantness) and their polarity; (3) the Valence Aware Dictionary for Sentiment Reasoning (VADER) [31], developed specifically for shorter texts found in social media contexts; (4) EmoLex [32] database that consists of word list related to particular emotions (e.g., anger, anticipation, disgust, fear, joy, sadness, surprise, and trust); (5) the Affective Norms for English Words (ANEW) database [33], that includes affective norms for valence, pleasure, arousal, and dominance; and (6) the SentiWordNet dictionary that enriches the WordNet word list with three sentiment scores: positivity, negativity, objectivity [34].

- Corpus-based approach

Augmenting the word search process with linguistic constraints, corpus-based approach aims to find contextual opinion words and their opinion directions. The extraction of opinion words and their sentiment polarity is often facilitated by sequential learning algorithms such as Conditional Random Field (CRF) [35]. This approach is suitable for creation and visualization of comparative relation maps that are often used as an important tool in the area of enterprise risk management and decision making. The two broad categories of corpus based sentiment analysis are: statistical and semantic approaches. Statistical methods use a simple underlying principle that the frequency of occurrence of a word in an annotated text corpus can be a robust indicator of its polarity, i.e. words occurring more frequently in positive

text are positive while words with higher frequency of occurrence in negative texts are negative. Those words which have equal frequency across positive and negative texts are neutral. On the other hand, semantic approach assigns similar sentiment polarities to words that are semantically similar. Corpus based approaches often adopt a mixture of tools such as pairwise mutual information (PMI) or its extensions (semantic oriented PMI), HowNet-based similarity measures, Latent Semantic Analysis (LSA), higher dimensional semantic spaces derived from lexical co-occurrence patterns (e.g. Hyperspace Analogue to Language or Sentiment Hyperspace Analogue to Language) etc. [36]. The corpus-based approach has been used in a wide variety of contexts due to its ability to identify the domain specific sentiments and their orientations, however, their dependency on corpus is an inherent weakness of such methods. Given this overview of the overall opinion mining process, we now move onto the next section that present a brief overview of some significant research works that have applied text/sentiment mining in a wide variety of managerial contexts.

### 3 Sentiment analysis in management research

#### 3.1 Understanding market structure and customer perceptions

Understanding consumer sentiments regarding products and brands is an important area of research as such knowledge can be used by firms to design their product and service offerings, communication strategies and branding decisions. There is a large body of research that handles these tasks using econometric and psychometric tools. In most cases these tools rely on consumer responses collected through statistically robust survey designs, field studies, or experimental setups. However, increasingly user generated content is emerging as a dominant source of information that can be effectively used by firms to handle these tasks.

Given its strategic implications, a number of studies have examined various aspects of online opinion. These include how online evaluations affect demand [37], how such content gets created [38], and how firms should strategically respond to online consumer reviews [39]. Due to its informational richness, review text has been successfully used to discover key product attributes [40], to make product recommendations [41], and to determine market structure [42]. It has been demonstrated that by analyzing the linguistic cues associated with deception it is possible to weed out those product reviews which are written by reviewers who have not actually tried the concerned product [43]. Studies have also shown that even the affective content and linguistic style of the online text reviews can have a significant effect on conversion rates [44]. More recently, researchers have utilized online reviews to predict consumer's purchase intention of durable goods [45] and to infer ratings of different product attributes [46]. User generated content from social networks such as Twitter has also been used for consumer insight mining [47].

Research in this domain has also examined consumers' behaviors [2], role of social networks [48], and social influence [49]. Online textual content has been

found to have significant impact on firm's performance in terms of brand images, purchase intentions, sales, return on investment [50], and stock prices [51, 52]. More recently, researchers have focused on how brand specific sentiments differ across various types on social platforms [3, 52]. From a methodological perspective, modified LDA based topic models have been recently proposed that can effectively analyze unstructured consumer reviews and identify consumer opinions at a sentence level. These models have also been extended to accommodate opinion stickiness where the reviewer talks about the same topic over a number of consecutive sentences [53]. Such inertia across sentences needs to be handled properly as it violates the independent and identical topic distribution assumption of classical topic models such as LDA. Examination of interaction dynamics in virtual communities and their underlying emotional antecedents have also attracted some attention. Muniz and O'Guinn [6] find that consumers often choose to participate in communities whose members share their interests and opinions. In a virtual community the inclination to post a comment individual level goal (e.g. solving problems or helping others by offering technical advice) and motivations that can either be intrinsic satisfaction or social benefits such as reputation [54]. Such works are important because many of the social media has emerged as an important component of firm's communication strategy.

### 3.2 Analysing financial sentiments

The concept of financial market sentiment has its root in the fundamental assumption that investors decisions are driven by their sentiments [55]. The market dynamics arise out of two factors: the transient sentiments of irrational traders (who are subject to exogenous sentiment) and the limited ability of the rational arbitrageurs to arbitrage. These limits arise because of several reasons: short time horizons, the costs associated with trading or short selling, and the fact that betting against irrational traders is inherently risky [56]. The vulnerability of the sentimental or irrational investors towards exogenous sentiment makes their behavior prone to outside information available from various sources such as financial news, press releases etc. Past research has shown that financial news can potentially affect the market [57] by impacting market returns [58, 59], intra-market volatility [60], and the profitability of different types of portfolios [61].

Analyzing financial sentiments using linguistic and opinion mining techniques has started drawing more research attention due to increasing acknowledgment of its prominent role in influencing market dynamics. In an early paper in this direction Knowles showed that the state of the financial market is often described in terms of health metaphors [62]. More recently, there has been work applying various sentiment analysis techniques to financial news analysis. Sentiments expressed in stock discussion boards have been found to affect price level in a technology stock index [63]. Specialized computational linguistics systems have been developed to predict stock prices and market volatility [64]. It has been found that the extent of pessimism in financial news column can significantly affect a company's cash flow [65, 66]. Tetlock and colleagues [67] used a dictionary based approach for sentiment analysis to examine the relation between the Dow Jones Industrial Index and a

pessimism index, specifically they utilized the General Inquirer Dictionary proposed by Stone et al. [68] in their studies. However, despite arguing that opinion sentiments can play an important role in determining the financial market dynamics the results from this research domain is not always unanimous or conclusive. Contrary to the findings of Tetlock [65], who suggest that textual sentiments can effectively supplants information related to a firms' fundamentals, Tetlock et al. [67] finds that sentiment analysis does not provide any significant additional information. Li [69] also reports that stock market fail to reflect the textual information regarding firms' future profitability available from the annual reports. Sinha [70] finds evidence that the stock market generally underreacts to the news sentiments. These contradictive results, however, are just an indication that our understanding of opinion dynamics in financial market is still incomplete and there is a strong need for further research in this field. In conclusion, it seems that textual information can informationally enrich and augment the conventional measures indicators of financial performance and play a significant role in determining market movements. Consequently, in agreement with the strongest form of efficient market hypotheses, a good financial market models might be justified to incorporate textual sentiment as an additional factor along with other firm-level characteristics.

### 3.3 Examining accounting practices

Understanding the textual information in corporate disclosures is important for financial accounting research. With significant advancement in the fields of computational linguistics, text mining, and machine learning in the past 2 decades accounting researchers have now access to powerful tools to understand financial disclosures and corporate communications better. These communications often indicate important managerial characteristics of the firm and thus have significant implications for understanding managers' behavioral biases and predicting corporate decisions. This argument is supported by past research in this field that suggests that communication patterns during critical decision making processes can reveal critical organizational characteristics and indicate firm's future performances. Recent findings also suggests that the level of optimism in earnings releases are positively associated with the market's short-term response [71]. In general, past evidence indicates that the extent of pessimistic sentiment cues in the earnings disclosures is correlated with lower future return on assets [72], while, the optimism and certainty embedded in such announcements are positively associated with future earnings and expected earnings uncertainty [73]. However, negative sentiment prior to earning announcements has found to be inversely associated with earnings surprise [74]. Based on these findings, it seems fair to conclude that firms' fundamentals often determine the textual sentiment in corporate disclosures and it does have the potential to offer additional information about firms' future performance that cannot be captured completely by conventional quantitative measures. Consequently, incorporating textual sentiment as an additional covariate along with usual firm-level fundamentals can prove to be a fruitful avenue for future research.

## 4 Future directions

Till now, we have seen how sentiment analysis can be defined, what are the basic steps involved in the implementation of a sentiment analysis (SA) system, different approaches to extract sentiment from a corpus of opinionated texts, and a brief overview of various applications of sentiment analysis or opinion mining in business domain. Now, we discuss a few possible future extensions of the existing sentiment analysis techniques solve some very specific problems. In this section we touch upon managerial issues related to customer and competition along with an important methodological issue that need further attention.

### 4.1 Understanding customers

Conventionally the notion of customer driven marketing strategy development depends on various tools to understand the perceptions and need of customers. The success of these tools largely depends on input data that is collected through survey instruments. However, such data often suffer from various biases introduced by the survey tools. Therefore, an alternative approach would be to utilize data that is voluntarily contributed by consumers. Online reviews offer such data but in the form of an unstructured text. In specific, it can be argued that given a product specific textual review, we can ideally extract various product specific aspects and their respective sentiments from the text. Subsequently, the review specific overall sentiment  $r_{dl}$  associated with a review  $d$  contributed by reviewer  $l$  can be expressed as weighted sum of latent sentiments regarding various product specific aspects. In principle this is similar to a generative latent rating regression (LRR) model proposed by Wang et al. [15]. The aspect identification part can be executed using a Bootstrapping step that relies on Chi square based measure of dependencies between aspects and words. At the end of the aspect segmentation process each review  $d$  is associated with a word frequency matrix ( $w_d$ ) that gives the normalized frequency of words in each aspect. The LRR model treats  $w_d$  as independent variables and the overall rating  $r$  of the review as the dependent variable. Formally, an aspect sentiment rating  $X_{il}$  is determined as follows:

$$X_{il} = \sum_{j=1}^n \omega_{ijl} w_{dijl} \quad (27)$$

where  $w_{dijl}$  represents the frequency of  $j$ th term belonging to  $i$ th aspect in the  $d$ th review contributed by  $l$ th reviewer and  $\omega_{ijl}$  represents the corresponding individual specific term weights. Similarly, the overall rating  $r_l$  associated with  $l$ th reviewer can be assumed to follow a Gaussian distribution. Thus we have

$$r_l \sim N\left(\sum_{i=1}^p \beta_{il} X_{il}, \sigma^2\right) \quad (28)$$

where  $X_{il}$  denotes the aspect ( $i$ ) and subject ( $l$ ) specific opinion rating and  $\sigma^2$  indicates the uncertainty of the overall rating predictions. This basic modelling

framework can be easily extended to incorporate a model based segmentation approach using finite mixture models. Specifically, instead of assuming that all the overall ratings are generated from a population level multivariate normal distribution with mean  $\mu_\beta$  and covariance matrix  $\Sigma_\beta$  i.e.  $\beta_l \sim N_p(\mu_\beta, \Sigma_\beta)$ , we can assume that each customer  $l$  belongs to one of  $K$  segments. The distribution of parameter heterogeneity in segment  $k \in \{1, 2, \dots, K\}$  assumed to follow a Gaussian distribution with mean  $\theta_k$  and variance- covariance matrix  $A_k$ :

$$P(\beta_l|k, \theta_k, A_k) = N(\theta_k, A_k). \tag{29}$$

Segment membership is assumed to be unknown, while the prior probability of belonging to segment  $k$  is denoted by  $\psi_k$ . In order to identify the model, the probabilities are ordered: the first segment is the smallest, and the last segment is the largest. This conceptualization induces a mixture model for the marginal distribution of  $r_l$ . Denoting an identity matrix by  $I$  and integrating out  $\beta_l$ , it can be shown that for a subject belonging to segment  $k$ :

$$r_l = X_l\theta_k + \varepsilon_l(k), \text{ and} \tag{30}$$

$$V(\varepsilon_l(k)) = \sigma^2 I + X_l A_k X_l'. \tag{31}$$

Following standard Bayesian estimation approach we further assume following priors for various parameters

$$P(\sigma^2|r_0, s_0) = IG\left(\sigma^2 \middle| \frac{r_0}{2}, \frac{s_0}{2}\right) \tag{32}$$

$$P(\theta_k|u_0, V_0) = N_p(\theta_k|u_0, V_0) \tag{33}$$

$$P(A_k|f_0, G_0) = IW_p(A_k|f_0, G_0^{-1}), \text{ and} \tag{34}$$

$$P(\psi|W_0) = ODir_K(\psi|W_0). \tag{35}$$

The joint distribution of the proposed hierarchical Bayesian mixture model becomes:

$$\prod_l N(r_l|X_l\beta_l, \sigma^2) \times \prod_l N_p(\beta_l|\theta_k, A_k) \times \prod_k N_p(\theta_k|u_0, V_0) IW_p(A_k|f_0, G_0^{-1}) \tag{36}$$

$$\times MN_K(N|\Psi) ODir_K(\Psi|W_0) \times IG(\sigma^2 \middle| \frac{r_0}{2}, \frac{s_0}{2}).$$

In above equations  $IG$ ,  $IW$ ,  $ODir$ ,  $MN$  represent inverse Gamma, inverse Wishart, Ordered Dirichlet and multinomial distributions respectively. Setting proper values initial values for prior parameters  $r_0, s_0, u_0, V_0, f_0, G_0$  and  $W_0$ , the estimation algorithm can now follow a standard Markov Chain Monte Carlo method such as a Gibbs sampler or a Metropolis–Hastings algorithm. The main objective of this extension is to show that sentiment analysis or a variant thereof, (i.e. in this case we don't treat sentiments in the conventional sense of positive or negative orientation

but as a much more granular rating expression) can be applied solve real life problems in customer management and marketing.

### 4.2 Developing competitive insight

Traditional corporate finance theories posit that firms faced with financial constraints—broadly defined as frictions that prevent firms from funding all desired investments have higher costs of external financing [75]. Financially constrained firms preserve internal finance to generate funds for future investment opportunities. Consequently, it can be argued that the way firms allocate funds across long term (forward looking) decisions (e.g. R&D, new product development, branding) and short term decisions (e.g. promotions) would also be dictated to a large extent by their ability to generate funds. In other words, the expected relative intensity of such decisions in these areas can be predicted beforehand if we can assess the extent of financial constraint a priori. At a very basic level, this idea can be implemented using the following linear model:

$$z_t = \beta_0 + \sum_{j=1}^J \beta_{1j} y_{t-j} + \sum_{j=0}^J \sum_{f=1}^F \beta_{2j}^f X_{t-j}^f + \sum_{j=0}^J \sum_{a=1}^A \beta_{3j}^a S_{t-j}^a + \varepsilon_t. \tag{37}$$

The dependent variable  $z_t$  denotes a latent performance indicator that the firm management can assess but unobserved by outside world (e.g. the competitor),  $X^f$  is a vector of control variables, and  $S$  is a vector of sentiment measures. This is a dynamic model that assumes that firm’s performance at time  $t$  depends on its prior performances for  $j$  time periods captured by  $y_{t-j}$ , control variables  $X_{t-j}^f$  and sentiment terms  $S_{t-j}^a$ . Where, both  $X_{(\cdot)}^f$  and  $S_{(\cdot)}^a$  can be contemporaneous or time lagged. The control variables can be various characteristics of the firm and market specific variables (e.g. cash flow from operations, the book-to-market ratio, the market value of equity, accruals and leverage, current earnings surprises, analyst earnings forecast revisions and dispersions, volatilities, stock market index returns, and trading volumes [76]). Now, defining  $\beta = \{ \beta_0, \vec{\beta}_1, \vec{\beta}_2, \vec{\beta}_3 \}$  and  $X_t = \{ 1, y_{t-j; j=1,..,j}, X_{t-j; j=1,..,j}^f, S_{t-j; j=1,..,j}^a \}$  we can simply write:

$$z_t = X_t \beta + \varepsilon_t. \tag{38}$$

Following Albert, Chib [77], the observed decisions (whether the firm engages in long-term investments or not) can therefore be expressed as:

$$y_t = I(z_t > 0). \tag{39}$$

The corresponding Gibbs sampler involves iterative sampling of  $p(\beta|z)$  and  $p(z|\beta)$ . Dropping indices for notational simplicity and assuming a normal prior on the parameter  $\beta$  i.e.  $\beta \sim N(\mu_\beta, \Sigma_\beta)$ , the conditional distributions are:



$$\begin{aligned}
 p(\beta|z) &\propto p(z|\beta)p(\beta) \\
 &\propto \exp\left[-\frac{1}{2}\left\{(z - x_t\beta)^2 + (\beta - \mu_\beta)' \Sigma_\beta (\beta - \mu_\beta)\right\}\right] \\
 &\propto \exp\left[-\frac{1}{2}\left\{(\beta - \tilde{\mu}_\beta)' \tilde{\Sigma}_\beta (\beta - \tilde{\mu}_\beta) + z' P z\right\}\right] \propto N(\beta; \tilde{\mu}_\beta, \tilde{\Sigma}_\beta) N(z; 0, P^{-1})
 \end{aligned}$$

where,  $\tilde{\Sigma}_\beta = (\Sigma_\beta^{-1} + X'X)^{-1}$ ,  $\tilde{\mu}_\beta = \tilde{\Sigma}_\beta (\Sigma_\beta^{-1} \mu_\beta + Xz)$  and  $P = I_T - X\tilde{\Sigma}_\beta X'$ .

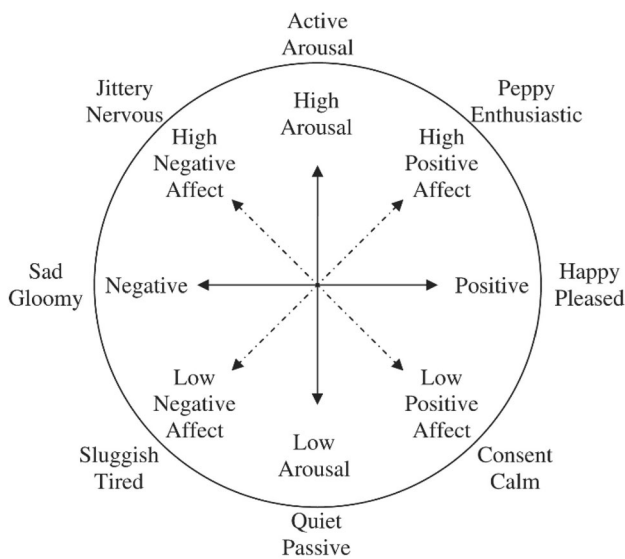
(40)

$$\text{So, } p(z|\beta, y, X) \propto p(y|z)p(z|\beta, X) = \prod_{t=1}^T p(y_t|z_t)p(z_t|\beta, x_t).$$
(41)

Similar approaches have earlier been used to examine whether textual sentiment can be used to predict occurrence or non-occurrence of specific events [78]. On the contrary we used a binary probit model to understand the impact of sentiment on firm’s decisions. Moreover, in contrast to conventional maximum likelihood based approach we have shown how such models can be solved using Bayesian estimation approach.

### 4.3 Going beyond sentiment: role of emotions

Contrary to the most frequent conceptualization of sentiment just in terms of positive and negative orientations, human emotions can be multidimensional in nature. Psychology literature has established the multifaceted nature of emotions that can assume various forms (Fig. 9) based not only on polarity but also on the



**Fig. 9** Human emotions (Reproduced with permission from Watson and Tellegen [79])

level of arousal [79]. Recognising this limitation, many researchers have tried to broaden the scope of sentiment analysis by incorporating various emotions suggested by psychological literature.

Recently, Kim et al. [80] proposed an interesting approach where large set of emotions are assumed to be embedded in a low dimensional Euclidean space. Instead of the conventional binary (positive or negative) conceptualization of emotion in a document, they introduce a multivariate response variable that corresponds to a complex emotional state. Consequently, the discrete emotion label  $Y \in \{1, 2, \dots, C\}$  for given document  $X$  depends on the position  $Z$  on a continuous manifold  $Z \in R^l$  so that the distribution of  $Z$  given a specific emotion label is assumed to be Gaussian. In other words,

$$p(Z|Y = y) \sim N(\mu_y, \Sigma_y). \quad (42)$$

Moreover, the distribution of  $Z$  given the document  $X$  (represented typically in a bag-of-words form) is assumed to be given by a linear regression model

$$p(Z|X = x) \sim N(\theta^T x, \Sigma_x). \quad (43)$$

They further assume that the distances between the vectors in  $E(Z|Y = y)$  are similar to the respective distances in  $E(X|Y = y)$ . Consequently, the parameter  $\mu_y = E(Z|Y = y); y \in C$  can be estimated by using either a multidimensional scaling or Kernel Principal Component Analysis (Kernel PCA) on  $\{E(X|Y = y) : y \in C\} = \frac{1}{n_k} \sum_{y(i)=k} x(i)$ , where  $n_k$  is the number of documents belonging to category  $k$ . The estimate ( $\hat{\theta}$ ) for the parameter  $\theta$  can be found by solving the regression model presented in Eq. (44) by using a maximum likelihood approach that gives

$$\hat{\theta} \sim \operatorname{argmax}_{\theta} \sum_i \log \int_Z p(z|y^{(i)}) \frac{p(y^{(i)})p_{\theta}(z|x^{(i)})}{\sum_y p(z|y)p(y)} dz. \quad (44)$$

Finally, the covariance matrices  $\Sigma_y$  can be estimated by computing the variance of  $Z$  values simulated from  $p_{\hat{\theta}}(z|x^{(i)})$  for all documents having the right labels  $Y^{(i)} = y$ . Once estimated, the parameters  $\hat{\theta}$ ,  $\mu_y$  and  $\Sigma_y$  can now be used to predict the most likely emotion class membership of a new document using the following relationship:

$$\hat{y} = \operatorname{argmax}_{\theta} \int p(y|z)p_{\hat{\theta}}(z|x)dx. \quad (45)$$

These models can be combined very successfully inside a temporal or dynamic econometric framework to track how the emotional states of a target group (e.g. consumers, voters etc.) are changing in response to various actions (e.g. campaign) and events (e.g. product failures). For example, there have been a number of recent cases where a product has suffered loss of reputation due to safety issues (e.g. Maggi in India, Samsung Note 7). The corresponding companies would be

interested in tracking whether subsequent measures taken by them and the related post event campaigns have been able to successfully address the consumers' concern and whether the corresponding brands/products have recovered from the loss of reputation they suffered or not. A multidimensional representation of the consumer's mood can potentially be more effective in determining the exact communication strategy (in terms of its content and message) to be adopted.

## 5 Conclusion

The main objective of this paper has been to present how opinion or sentiment analysis has been used in management research. In contrast to some of the existing reviews that offer a wider theoretical coverage of this process, we adopted a more applied orientation and provided a structured description of the technical details associated with various steps of this process. Moreover, with specific examples drawn from three important domains of management (marketing, financial market and accounting), we demonstrated how these techniques have been used in past. Finally, we presented a number of examples pointing out where opportunities for further research exist. Given that this research area is expanding rapidly, it is virtually impossible to keep track of all the developments that is happening unless we are actively associated with this line of research. Hence, we feel there is a strong need for such works not only to keep researchers in other fields (e.g. management and social sciences) informed about the underlying concepts of these field but also to provide a more applied outlook of the possibilities that exist in more concrete manner. The main intention of this paper is thus to give the readers a consolidated understanding of the state of the arts for this rapidly growing field and to encourage them to think of possible applications in their field of choice.

## References

1. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Aggarwal, CC., Zhai, CX. (eds.) *Mining Text Data*, pp. 415–463. Springer, Berlin (2012)
2. Berger, J., Milkman, K.L.: What makes online content viral? *J. Mark. Res.* **49**(2), 192–205 (2012)
3. Schweidel, D.A., Moe, W.W.: Listening in on social media: a joint model of sentiment and venue format choice. *J. Mark. Res.* **51**(4), 387–402 (2014)
4. Toubia, O., Stephen, A.T.: Intrinsic vs. image-related utility in social media: why do people contribute content to twitter? *Mark. Sci.* **32**(3), 368–392 (2013)
5. Chmiel, A., Sienkiewicz, J., Thelwall, M., Paltoglou, G., Buckley, K., Kappas, A., Holyst, J.A.: Collective emotions online and their influence on community life. *PLoS ONE* **6**(7), e22207 (2011)
6. Muniz, A.M., O'guinn, T.C.: Brand community. *J. Consum. Res.* **27**(4), 412–432 (2001)
7. Sun, S., Luo, C., Chen, J.: A review of natural language processing techniques for opinion mining systems. *Inf. Fusion* **36**(Supplement C), 10–25 (2017)
8. Piryani, R., Madhavi, D., Singh, V.K.: Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Inf. Process. Manag.* **53**(1), 122–150 (2017)
9. Liu, B.: Sentiment analysis and subjectivity. *Handb. Nat. Lang. Process.* **2**, 627–666 (2010)
10. Kumar, S., Morstatter, F., Liu, H.: *Twitter Data Analytics*. Springer, Berlin (2014)

11. Swanson, B., Charniak, E.: Native language detection with tree substitution grammars. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 2012, pp. 193–197. Association for Computational Linguistics
12. Porter, W.: The Porter stemming algorithm: then and now. *Program* **40**(3), 219–223 (2006)
13. Porter, M.: Lovins revisited. In: Tait, J. (ed.) *Charting a New Course: Natural Language Processing and Information Retrieval*, pp. 39–68. Springer, University of Sunderland, Sunderland, UK (2005)
14. Titov, I., McDonald, R.T.: A joint model of text and aspect ratings for sentiment summarization. In: ACL 2008, pp. 308–316
15. Wang, H., Lu, Y., Zhai, C.: Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2010, pp. 783–792. ACM
16. Hagenau, M., Liebmann, M., Neumann, D.: Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decis. Support Syst.* **55**(3), 685–697 (2013)
17. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* **5**(4), 1093–1113 (2014)
18. Vapnik, V.: *Statistical Learning Theory*. Wiley, New York (1998)
19. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
20. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**(4), 115–133 (1943)
21. Mosteller, F., Wallace, D.L.: *Inference and disputed authorship: The Federalist*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., (1964)
22. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
23. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence 2004, pp. 487–494. AUAI Press
24. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management 2009, pp. 375–384. ACM
25. Jo, Y., Oh, A.H.: Aspect and sentiment unification model for online review analysis. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining 2011, pp. 815–824. ACM
26. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th International Conference on World Wide Web 2007, pp. 171–180. ACM
27. Kang, M., Ahn, J., Lee, K.: Opinion mining using ensemble text hidden Markov models for text classification. *Expert Syst. Appl.* **94**(Supplement C), 218–227 (2018)
28. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: an on-line lexical database. *Int. J. Lexicogr.* **3**(4), 235–244 (1990)
29. Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M.: The general inquirer: a computer approach to content analysis. *Am. J. Sociol.* **73**(5), 634–635 (1968)
30. Cambria, E., Speer, R., Havasi, C., Hussain, A.: SenticNet: a publicly available semantic resource for opinion mining. In: Commonsense knowledge: papers from the AAAI fall symposium, pp. 14–18. Menlo Park, CA, USA: AAAI Press. 2010 AAAI Fall Symposium, Arlington, VA, USA (2010)
31. Gilbert, C.H.E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International AAAI Conference on Weblogs and Social Media (2014)
32. Mohammad, S.M., Turney, P.D.: Emotions evoked by common words and phrases: using mechanical Turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text 2010, pp. 26–34. Association for Computational Linguistics
33. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): instruction manual and affective ratings. In: Technical Report C-1, the Center for Research in Psychophysiology, University of Florida (1999)
34. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC, vol. 10, pp. 2200–2204 (2007)
35. Lafferty, J., McCallum, A., Pereira, F.C.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: proceeding of the International Conference on Machine Learning (ICML), pp. 282–289 (2001)

36. Lund, K., Burgess, C.: Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* **28**(2), 203–208 (1996)
37. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: online book reviews. *J. Mark. Res.* **43**(3), 345–354 (2006)
38. Moe, W.W., Schweidel, D.A.: Online product opinions: incidence, evaluation, and evolution. *Mark. Sci.* **31**(3), 372–386 (2012)
39. Chen, Y., Xie, J.: Third-party product review and firm marketing strategy. *Mark. Sci.* **24**(2), 218–240 (2005)
40. Lee, T.Y., Bradlow, E.T.: Automated marketing research using online customer reviews. *J. Mark. Res.* **48**(5), 881–894 (2011)
41. Ghose, A., Ipeirotis, P.G., Li, B.: Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Mark. Sci.* **31**(3), 493–520 (2012)
42. Netzer, O., Feldman, R., Goldenberg, J., Fresko, M.: Mine your own business: market-structure surveillance through text mining. *Mark. Sci.* **31**(3), 521–543 (2012)
43. Anderson, E.T., Simester, D.I.: Reviews without a purchase: low ratings, loyal customers, and deception. *J. Mark. Res.* **51**(3), 249–269 (2014)
44. Ludwig, S., de Ruyter, K., Friedman, M., Brügger, E.C., Wetzels, M., Pfann, G.: More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. *J. Mark.* **77**(1), 87–103 (2013)
45. Bag, S., Tiwari, M.K., Chan, F.T.S.: Predicting the consumer's purchase intention of durable goods: an attribute-level analysis. *J. Bus. Res.* (2017). <https://doi.org/10.1016/j.jbusres.2017.11.031>
46. Kangale, A., Kumar, S.K., Naeem, M.A., Williams, M., Tiwari, M.K.: Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary. *Int. J. Syst. Sci.* **47**(13), 3272–3286 (2016)
47. Rathan, M., Hulipalad, V.R., Venugopal, K.R., Patnaik, L.M.: Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews. *Appl. Soft Comput.* (2017). <https://doi.org/10.1016/j.asoc.2017.07.056>
48. Goldenberg, J., Oestreicher-Singer, G., Reichman, S.: The quest for content: how user-generated links can facilitate online exploration. *J. Mark. Res.* **49**(4), 452–468 (2012)
49. Trusov, M., Bodapati, A.V., Bucklin, R.E.: Determining influential users in internet social networks. *J. Mark. Res.* **47**(4), 643–658 (2010)
50. Hoffman, D.L., Fodor, M.: Can you measure the ROI of your social media marketing? *MIT Sloan Manag. Rev.* **52**(1), 41 (2010)
51. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. *J. Comput. Sci.* **2**(1), 1–8 (2011)
52. Tirunillai, S., Tellis, G.J.: Does chatter really matter? Dynamics of user-generated content and stock performance. *Mark. Sci.* **31**(2), 198–215 (2012)
53. Büschken, J., Allenby, G.M.: Sentence-based text analysis for customer reviews. *Mark. Sci.* **35**(6), 953–975 (2016)
54. Chen, Y.-J., Kirmani, A.: Posting strategically: the consumer as an online media planner. *J. Consum. Psychol.* **25**(4), 609–621 (2015)
55. De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J.: Positive feedback investment strategies and destabilizing rational speculation. *J. Finance* **45**(2), 379–395 (1990)
56. Shleifer, A., Vishny, R.W.: The limits of arbitrage. *J. Finance* **52**(1), 35–55 (1997)
57. Klein, F.C., Prestbo, J.A.: *News and the Market*. H. Regnery Co., Washington (1974)
58. Engle, R.F., Ng, V.K.: Measuring and testing the impact of news on volatility. *J. Finance* **48**(5), 1749–1778 (1993)
59. Melvin, M., Yin, X.: Public information arrival, exchange rate volatility, and quote frequency. *Econ. J.* **110**(465), 644–661 (2000)
60. Ederington, L.H., Lee, J.H.: How markets process information: news releases and volatility. *J. Finance* **48**(4), 1161–1191 (1993)
61. Chan, W.S.: Stock price reaction to news and no-news: drift and reversal after headlines. *J. Financ. Econ.* **70**(2), 223–260 (2003)
62. Knowles, F.: Lexicographical aspects of health metaphors in financial texts. In: *Euralex96 Proceedings (Part II)*. Göteborg University, Göteborg, Sweden, pp. 789–796 (1996)
63. Das, S.R., Chen, M.Y.: Yahoo! for Amazon: sentiment extraction from small talk on the web. *Manag. Sci.* **53**(9), 1375–1388 (2007)

64. Antweiler, W., Frank, M.Z.: Is all that talk just noise? The information content of internet stock message boards. *J. Finance* **59**(3), 1259–1294 (2004)
65. Tetlock, P.C.: Giving content to investor sentiment: the role of media in the stock market. *J. Finance* **62**(3), 1139–1168 (2007)
66. Loughran, T.I.M., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J. Finance* **66**(1), 35–65 (2011)
67. Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S.: More than words: Quantifying language to measure firms' fundamentals. *J. Finance* **63**(3), 1437–1467 (2008)
68. Stone, P., Dunphy, D.C., Smith, M.S., Ogilvie, D.M.: The general inquirer: a computer approach to content analysis. *J. Reg. Sci.* **8**(1), 113–116 (1968)
69. Li, F.: Textual analysis of corporate disclosures: a survey of the literature. *J. Account. Lit.* **29**, 143 (2010)
70. Sinha, N.R.: Underreaction to news in the US stock market. *Q. J. Finan.* **6**(02), 1650005 (2016)
71. Feldman, R., Govindaraj, S., Livnat, J., Segal, B.: Management's tone change, post earnings announcement drift and accruals. *Rev. Account. Stud.* **15**(4), 915–953 (2010)
72. Davis, A.K., Tama-Sweet, I.: Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&A. *Contemp. Account. Res.* **29**(3), 804–837 (2012)
73. Demers, E., Vega, C.: Soft information in earnings announcements: news or noise? (2008) (**Working Paper**)
74. Chen, X., Cheng, Q., Lo, A.K.: Is the decline in the information content of earnings following restatements short-lived? *Account. Rev.* **89**(1), 177–207 (2013)
75. Lamont, O., Christopher, P., Jesus, S.-R.: Financial Constraints and Stock returns. *Rev. Finan. Stud.* **14**(2), 529–554 (2001)
76. Kearney, C., Liu, S.: Textual sentiment in finance: a survey of methods and models. *Int. Rev. Financ. Anal.* **33**, 171–185 (2014)
77. Albert, J.H., Chib, S.: Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* **88**(422), 669–679 (1993)
78. Rogers, J.L., Van Buskirk, A., Zechman, S.L.C.: Disclosure tone and shareholder litigation. *Account. Rev.* **86**(6), 2155–2183 (2011)
79. Watson, D., Tellegen, A.: Toward a consensual structure of mood. *Psychol. Bull.* **98**(2), 219 (1985)
80. Kim, S., Li, F., Lebanon, G., Essa, I.: Beyond sentiment: the manifold of human emotions. In: *Artificial Intelligence and Statistics 2013*, pp. 360–369