Sentiment Analysis from User Reviews Using a Hybrid Generative-Discriminative HMM-SVM Approach

Rim Nasfi(✉) and Nizar Bouguila

Concordia Institute for Information Systems Engineering, 1455 De Maisonneuve Blvd. West, Montreal, QC, Canada
r_nasfi@encs.concordia.ca, nizar.bouguila@concordia.com

Abstract. Sentiment analysis aims to empower automated methods with the capacity to recognize sentiments, opinions and emotions in text. This recognition capacity is now highly demanded to process and extract proper knowledge from the exponentially-growing volume of user-generated data. Applications such as analyzing online products reviews on e-commerce marketplaces, opinion mining from social networks and support chat-bots optimization are putting into practice various methods to perform this complex natural language processing task. In this paper, we apply a hybrid generative-discriminative approach using Fisher kernels with generalized inverted Dirichlet-based hidden Markov models to improve the recognition performance in the context of textual analysis. We propose a method that combines HMMs as a generative approach, with the discriminative approach of Support Vector Machine. This strategy allows us to deal with sequential information of the text, and at the same time use the special focus on the classification task that SVM could provide us. Experiments on two challenging user reviews datasets i.e. Amazon for products reviews and IMDb movies reviews, demonstrate an effective improvement of the recognition performance compared to the standard generative and Gaussian-based HMM approaches.

Keywords: Sentiment analysis · Opinion mining · Hybrid Generative-Discriminative · Hidden Markov models · Support Vector Machines

1 Introduction

Opinion mining received massive interest in recent years, particularly with the important role that reviews and shared experiences over e-commerce and marketplaces platforms play in shaping purchase intentions. User-generated data is increasing drastically, especially when it comes to reviews and feedback shared over the internet [1]. This huge volume of data calls for automated methods to process and extract proper knowledge from it [2]. Analyzing users' opinions from different perspectives can considerably help not only the customer to buy or adopt the best product available in the market, but also the merchant, to better understand what are the good or bad features related to their products and determine their effect on the buyers' opinion and feeling regarding the

A. Krzyzak et al. (Eds.): S+SSPR 2022, LNCS 13813, pp. 74–83, 2022.
https://doi.org/10.1007/978-3-031-23028-8_8

product [3]. These reviews are for the most part available in a text format in an unstructured way and naturally need to be modeled appropriately in order to provide useful insights to both customer and seller. Therefore, recognizing sentiments and attitudes in textual data can provide a better understanding of trends and tendencies related to products [4].

Sentiment analysis, also known as opinion mining analyzes people's opinions as well as their emotions towards a product, an event or an organization [5]. It has been widely investigated in different research works and approached through different methodologies such as lexicon-based approaches as well as hybrid approaches [6]. Nevertheless, there has been rarely a solid explainability or knowledge behind decisions resulting from these methods, and the latter were oftentimes handled as black-box methods. Challenges in sentiment analysis as a natural language processing application are numerous. In fact, analyzing text reviews implies dealing with text sequences that are usually limited in length, have many misspellings and shortened forms of words [7]. As a result, we have an immense vocabulary size and vectors representing each review are highly sparse.Two main approaches are used in machine learning to perform recognition tasks: generative techniques that model the underlying distributions of classes, and discriminative techniques that give a sole focus on learning the class boundaries [8]. Both techniques have been widely used in sentiment analysis to effectively recognize divergent users' attitudes [9].

Hidden Markov Models (HMMs) represent a powerful tool to properly model sequential information within textual data. Their generative aspects constitute a quite potent way to handle sentiment recognition and they tend to require less training data than discriminative models. In the case where the task to be performed is classification, Support Vector Machines (SVM) can clearly distinguish the differences between categories and can thus outperform generative models especially if a large number of training examples are available. SVM is extensively used due to its great capacity to generalize, often resulting in better performance than traditional classification techniques [10]. As a discriminative approach, the main functioning of SVM is to find surfaces that better separate the different data classes using a kernel that allows efficient discrimination in non-linearly separable input feature spaces. Hence the importance of adopting the convenient kernel function which needs to be suitable for the classified data and the objective task. Standard kernels include linear polynomial and radial basis function kernels [11]. Adopting these kernels is not always possible, especially when it comes to classifying objects represented by sequences of different lengths [12]. Consequently, the mentioned kernels may not be a good fit to model our text data. Therefore, a hybrid generative-discriminative approach is adopted to permit the conversion of data into fixed-length and hence provide additional performance to the model.

In this work, we introduce a novel implementation of a hybrid generative-discriminative model and examine its performance on real-life benchmark datasets. We propose the use of Fisher Kernels (FK) generated with Generalized inverted Dirichlet-based HMMs (GIDHMM) to model textual data. Moreover, the use of GID (Generalized Inverted Dirichlet) to model emission probabilities is backed by the several interesting mathematical properties that this distribution has to offer. These properties allow for a representation of GID samples in a transformed space where features are independent and follow inverted Beta distributions. Adopting this distribution allows us to take advantage of conditional independence among features. This interesting strength is used in this paper to develop a statistical model that essentially handles positive vectors.

In light of the existing methods in sentiment analysis, the main contributions of this paper are the following: First, we apply for the first time a non-Gaussian HMM, i.e. generalized inverted Dirichlet-based HMM on the challenging product reviews benchmark by Amazon and the IMDb movie reviews dataset. Second, we derive a hybrid generative-discriminative approach of our HMM-based framework with FK for SVM-based modeling of positive vectors. This novel approach is also tested on the aforementioned datasets, as an unprecedented attempt of using Fisher Vectors-based hybrid generative-discriminative models to handle textual data analysis. The remainder of the paper is organized as follows, Sect. 2 presents background topics on sentiment analysis and examines related works. Section 3 discusses the proposed model. Section 4 presents the performed experiments and obtained results. We finally conclude the paper in Sect. 5.

2 Related Work

Sentiment analysis has been the focus of numerous research works, where it has been approached in different levels namely document, sentence and aspect level [13]. While the document level focuses on classifying the whole opinion document into either a positive or negative sentiment, sentence-level looks at determining whether the sentence expresses the nature of opinion (negative, positive, neutral). On the other hand, the aspect-level analysis provides a detail-oriented approach to handle the broad aspect. Thus, it focuses on determining whether or not a part of the text is opinion-oriented towards a certain aspect. It can present a positive polarity towards one aspect and a negative polarity towards another. Classifying the text as positive or negative depends on the chosen aspect and applied knowledge [14]. It is noteworthy to mention that expressions associated with sentiment are mainly the words or features that express the sentiment of the text, such as adjectives or adverbs. Furthermore, when tackling sentiment analysis, there are mainly three types of machine learning approaches, i.e. supervised, unsupervised and semi-supervised learning and they are respectively used in cases where data is labelled, unlabeled and partially labelled [15].

In HMM-based sentiment analysis, models analyze the input textual data and formulate clusters. After that HMMs are utilized to perform the categorization by considering the clusters as hidden states. Every model analyzes a given instance in order to specify its sentimental polarity. In comparison to related works in the literature, HMM-based methods possess higher interpretability and can model the changing aspects of sentiment information. Multiple sentiment analysis applications adopt HMMs as the main model. In [16] Rabiner proposes a method of predicting sentiments from voice. Also, in [17] authors use HMMs to detect sentiments by considering the label information as positive, negative or neutral. Knowledge about the words position and hidden states is available and injected into the model. While this approach has shown effective results, it clearly assumes knowledge of the labels and thus requires a significant human effort.

In our work, we do not require knowledge about the states labels and we propose another alternative where we estimate the similarity between the pattern of input text and that of sentences expressing either a positive or negative sentiment, plus we make use of SVM to increase the model's performance when it comes to the classification accuracy. We detail our method in the next section.

3 Hybrid Generative-Discriminative Approach with Fisher Kernels

When it comes to our adopted approach, a single HMM is trained for every class in the data depending on the context aspect. The resulting likelihoods will be further classified by the SVM classifier to identify the sentiment. In this section, we present the proposed approach. To illustrate our model, we are first listing various HMM notations and enumerating the upcoming used work script. We then recall the main process behind the forward-backward algorithm. Lastly, we perform a complete derivation of the FK-based model.

3.1 Hidden Markov Models

A Hidden Markov Model is a statistical model that can be used to describe real-world processes with observable output signals. HMMs are defined as an underlying stochastic process formed by a Markov chain that is not observable (hidden). For each hidden state, a stochastic model creates observable output signals or observations, based on which hidden states can be estimated [18]. We consider a HMM with continuous emissions and K hidden states. We put a set of hidden states $H = \{h_1, ..., h_T\}; h_j \in [1, K]$. The transition probabilities matrix: $B = \{b_{ij} = P(h_t = j|h_{t-1} = j')\}$ and the emission probabilities matrix: $C = \{c_{ij} = P(m_t = i|h_t = j)\}; i \in [1, M]$ where M is the number of mixture components associated with state j. We define the initial probability: π_j which is the probability to start the observation sequence from the state j. We denote an HMM as: $\Delta = \{B, C, \varphi, \pi\}$ where φ is the set of mixture parameters depending on the chosen type of mixture. In this work, we focus on the generalized inverted Dirichlet distribution. Let $\overrightarrow{X}$ a D-dimensional positive vector following a GID distribution. The joint density function is given by Lingappaiah [19] as:

$$p(\overrightarrow{X}|\overrightarrow{\alpha}, \overrightarrow{\beta}) = \prod_{d=1}^{D} \frac{\Gamma(\alpha_d + \beta_d)}{\Gamma(\alpha_d)\Gamma(\beta_d)} \frac{X_d^{\alpha_d - 1}}{\left(1 + \sum_{l=1}^{d} X_d\right)^{\eta_d}} \tag{1}$$

where $\overrightarrow{\alpha} = [\alpha_1, ..., \alpha_D]$, $\overrightarrow{\beta} = [\beta_1, ..., \beta_D]$. η is defined such that $\eta_d = \alpha_d + \beta_d - \beta_{d+1}$ for $d = 0, ..., D$ with $\beta_{D+1} = 0$.

3.2 Inference on Hidden States: Forward-Backward Algorithm

The forward algorithm computes the probability of being in state h_j up to time t for the partial observation sequence produced by the model Δ. We consider a forward variable $\gamma_t(i) = P(X_1, X_2, \ldots, X_t, , i_t = h_i|\Delta)$. There is a recursive relationship that is used to compute the former probability. We can resolve for $\gamma_t(i)$ recursively as follows:

1. Initialization:

$$\gamma_t(i) = \pi_i \varphi_i(X_1) \quad 1 \leq i \leq K \tag{2}$$

2. Recursion:

$$\gamma_{t+1}(j) = \left[\sum_{i=1}^{K} \gamma_t(i) b_{ij} \right] \varphi_j(X_{t+1}) \quad for \quad 1 \leq t \leq T-1, \; 1 \leq j \leq K \tag{3}$$

3. Termination:

$$P(X|\Delta) = \sum_{i=1}^{K} \gamma_T(i) \tag{4}$$

The backward variable, which is the probability of the partial observation sequence $X_{t+1}, X_{t+2}, \ldots, X_T$ given the current state is denoted by $\delta_t(i)$ and can similarly be determined as follows:

1. Initialization:

$$\delta_t(i) = 1, \; 1 \leq i \leq K \tag{5}$$

2. Recursion:

$$\delta_t(i) = \sum_{j=1}^{K} b_{ij} \varphi_j(X_{t+1} \delta_{t+1}(j) \quad for \quad t = T-1, T-2, \ldots, 1 \quad 1 \leq i \leq K \tag{6}$$

3. Termination:

$$P(X|\Delta) = \sum_{i=1}^{K} \gamma_i(T) = \sum_{i=1}^{K} \pi_i b_i(X_1) \varphi_i(1) = \sum_{i=1}^{K} \sum_{j=1}^{K} \gamma_t(i) b_{ij} \varphi_j(X_{t+1}) \delta_{t+1}(j) \tag{7}$$

3.3 Fisher Kernels

Non-linear SVM serves our discrimination needs in the context of realistic recognition tasks. The strategy is to use a Kernel method to avoid calculation cost and memory consumption problems that might arise from performing inner product calculation of high-dimensional feature vectors. It will allow us to implicitly project objects to high-dimensional space by using a kernel function $\kappa(x_\zeta, x_\eta) = \langle \phi(x_\zeta), \phi(x_\eta) \rangle$ and solving the problem with observations x_ζ and x_η represented as Bag of Features (BoF) or Bag of Words (BoW) [20] in general with ϕ being a projection function and $\langle \cdot, \cdot \rangle$ meaning the inner product. Here we choose Fisher Kernel as the kernel function. This choice is motivated by FK being a general way of fusing generative and discriminative approaches for classification. FK is formulated as

$$FK(X_\zeta, X_\eta) = \langle FS(X_\zeta, \Delta), FS(X_\eta, \Delta) \rangle \tag{8}$$

where X_ζ and X_η are two observations, Δ is the parameters set of a generative model defined by $P(X|\Delta)$ and $FS(X_\zeta,\Delta)$ is the Fisher score.

$$FS(X,\Delta) = \nabla_\Delta \log P(X|\Delta) \tag{9}$$

Given a particular HMM:

$$L(X|\Delta) = \log P(X|\Delta) = \log \sum_{i=1}^{K} \gamma_T(i) = \log \sum_{i=1}^{K} \pi_i \varphi_i(X_1)\delta_1(i) \tag{10}$$

The derivatives for GID-based HMM can be defined as follows

$$\nabla_\Delta L(X|\Delta) = \left[\frac{\partial L(X|\Delta)}{\partial \pi_i}, \frac{\partial L(X|\Delta)}{\partial b_{ij}}, \frac{\partial L(X|\Delta)}{\partial \alpha_{id}}, \frac{\partial L(X|\Delta)}{\partial \beta_{id}}\right] \tag{11}$$

$$\frac{\partial L(X|\Delta)}{\partial \pi_i} = \frac{\varphi_i(X_1)\delta_1(i)}{\sum_{i=1}^{K} \pi_i \varphi_i(X_1)\delta_1(i)} \tag{12}$$

$$\begin{aligned}\frac{\partial L(X \mid \Delta)}{\partial b_{ij}} &= \frac{1}{P(X \mid \Delta)} \sum_{k=1}^{K} \frac{\partial \gamma_T(k)}{\partial b_{ij}} \\ &= \frac{1}{P(X \mid \Delta)} \sum_{k=1}^{K} \sum_{l=1}^{K} \frac{\partial \gamma_{T-1}(l)}{\partial b_{ij}} b_{lk} \varphi_k(X_T) + \partial \gamma_{T-1}(i) \varphi_{ij}(X_T)\end{aligned} \tag{13}$$

$$\frac{\partial L(X \mid \Delta)}{\partial \alpha_{id}} = \frac{1}{P(X \mid \Delta)} \left(\sum_{j=1}^{K} \sum_{k=1}^{K} \frac{\partial \gamma_{T-1}(k)}{\partial \alpha_{id}} b_{kj} \varphi_j(X_T) + \sum_{k=1}^{K} \partial \gamma_{T-1}(k) b_{ki} \frac{\partial \varphi_i(X_T)}{\partial \alpha_{id}} \right) \tag{14}$$

$$\frac{\partial L(X \mid \Delta)}{\partial \beta_{id}} = \frac{1}{P(X \mid \Delta)} \left(\sum_{j=1}^{K} \sum_{k=1}^{K} \frac{\partial \gamma_{T-1}(k)}{\partial \beta_{id}} b_{kj} \varphi_j(X_T) + \sum_{k=1}^{K} \partial \gamma_{T-1}(k) b_{ki} \frac{\partial \varphi_i(X_T)}{\partial \beta_{id}} \right) \tag{15}$$

$$\frac{\partial \varphi_i(X_t)}{\partial \alpha_{id}} = \Psi(\alpha_{id} + \beta_{id}) - \Psi(\alpha_{id}) + \log\left(\frac{X_d}{1+X_d}\right) \tag{16}$$

$$\frac{\partial \varphi_i(X_t)}{\partial \beta_{id}} = \Psi(\alpha_{id} + \beta_{id}) - \Psi(\beta_{id}) + \log\left(\frac{1}{1+X_d}\right) \tag{17}$$

4 Experiments

4.1 Problem Modeling

The main motive behind the use of HMMs in sentiment analysis is the strong analogy behind the process of understanding a sentiment from a text in real-life and the predictive aspect of HMMs. In fact, an opinion consists of a number of words that together represent what the person is trying to express. To understand it, a person would first proceed by reading the words sequentially from left to right, knowing that each word would normally be related to the previous one in a certain way to create a meaningful sentence. Accordingly, words forming an emotion are modeled as observations in a HMM, while the emotion is the hidden state, which needs to be unveiled. In this work, we propose to tackle this problem of sentiment analysis in a hybrid way; where the HMM states can be modeled approximately by considering a hidden variable given by patterns that are independent of the class of text. An abstraction of this process is illustrated in Fig. 1. We first perform word clustering to indicate a certain word pattern, in a way that all negative and positive connotations are clustered separately. After that, we make use of our SVM to further classify the output into negative and positive classes. This treatment requires a dictionary constructed for use by sentiment analysis models.

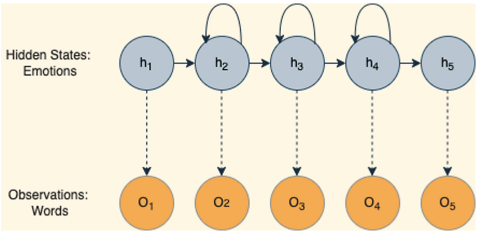

Fig. 1. Problem modeling through hidden state-observation HMM

4.2 Datasets

We choose to experiment on the Amazon[1] reviews dataset from the Stanford Network Analysis Project (SNAP), which spans 18 years period of product reviews [21]. Amazon dataset is a popular corpus of product reviews collected from the Amazon marketplace, which includes ratings and plain text reviews for a multitude of product niches. In this work, we choose to work on the Electronics niche and randomly pick a mix of 30,000 sample reviews from training and testing sets with a vocabulary size of 72.208 unique words. As a word segmentation approach, we use the Part-of-Speech tagger designed by the Stanford NLP group applying default settings. We have also removed numerals, auxiliary words and verbs, punctuation and stop words as a part of the pre-processing. An output vector is then generated corresponding to the input word. Each review is modeled by a text vector that is obtained after adding up all the word vectors and dividing by the number of words. We also test our work on the IMDb dataset [22], which is mainly developed for the task of binary sentiment classification of movie reviews. It consists of an equal number of positive and negative reviews. The dataset is evenly divided between training and test sets with 25,000 reviews each. We choose to work on both subsets uniformly and we hence deal with 50,000 samples from each group with 76,340 unique words in total. A similar pre-processing to the one adopted with the Amazon dataset is then applied.

[1] Publicly available at: https://snap.stanford.edu/data/web-Amazon.html.

Experiments are carried out in a total of 10 independent runs, each time starting with a different initial observation. The resulting log-likelihoods and accuracies are averaged on these 10 independent runs. It is worth mentioning that through the training process our aim is to maximize the log-likelihood of the data and we target by performing these experiments a higher recognition accuracy each time. Therefore we will not focus on assessing the time complexity in this work. We also use the totality of the dataset for training and testing and do not opt for splitting the data into train and test segments for both datasets.

4.3 Results

A HMM is trained for each class, in this case, positive and negative. A test text is sent to each model and the probabilities of occurrence are computed. Each HMM returns a probability and the model with the highest probability of occurrence will indicate the class of this text. We choose the number of hidden states to be $K = 2$ for both experiments. In our discussion, we focus on the effect of using a hybrid model as opposed to a HMM-based model. We also shed some light on the usefulness of using the GID distribution as emission probabilities. Results on both datasets are presented in Fig. 2. We notice that, on both datasets, Hybrid HMM-SVM models perform remarkably better than Generative-only models, be it GID-based or Gaussian-based. The hybrid GID-based HMM/SVM model achieved average accuracies of 86.40% and 88.94% on the Amazon and IMDb datasets respectively, while we only yielded 79.72% and 82.09% using the GID-HMM Generative-only model. Most importantly, we notice that GID-based models achieved the highest accuracy on both hybrid and generative approaches compared to the Gaussian-based models. This increasing recognition capacity was expected and is once more validated when it comes to positive vector modeling. Using GID as an emission probability distribution, clearly improved the modeling accuracy. Conclusively, results achieved by applying the hybrid generative-discriminative approach demonstrate the striking increase in terms of the modeling accuracy and further validate the improved performance that SVM provided to the generative technique.

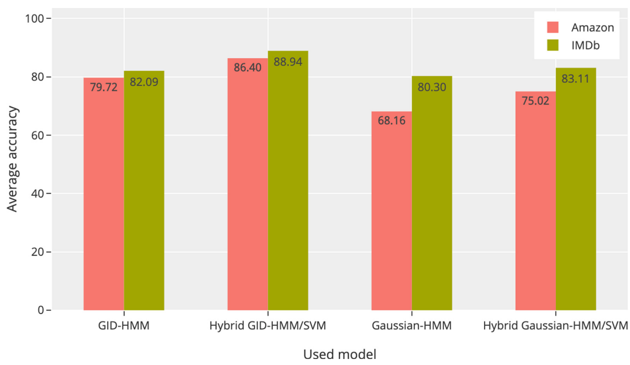

Fig. 2. Average accuracies for sentiment recognition on the Amazon and IMDb datasets with each of the tested models

5 Conclusion

In this work, we presented a hybrid generative-discriminative approach to automatically identify sentiments expressed in user reviews online, using a combination of HMMs as a generative approach, along with the discriminative SVM. The main motivation behind this choice is to be able to enhance the model's capacity by taking advantage of the powerful classification role that SVM plays without neglecting the sequential aspect of text data. We also gave a special focus on modeling positive vectors by using non-Gaussian Generalized Inverted Dirichlet distributions as emission probabilities for our HMM. The interest in adopting the GID for modeling our data arose from the limitations encountered when inverted Dirichlet was adopted, in particular its restraining strictly positive covariance. We carried out what we believe to be the first attempt of applying GIDHMM both in generative and hybrid modes on the challenging Amazon product reviews and IMDb reviews. A comprehensive solution for sentiment detection was introduced, where we allowed the automatic recognition of positive and negative emotions, in textual data. According to the results obtained from the conducted experiments, we proved that the proposed approach obtained highly accurate recognition rates compared to both generative GIDHMM and Gaussian-based HMM. Future works are intended to be done in the near future extending this work to different Natural Language Process applications, such as product recommendation and understanding user intent.

References

1. Chakraborty, K., Bhattacharyya, S., Bag, R.: A survey of sentiment analysis from social media data. IEEE Trans. Comput. Soc. Syst. **7**(2), 450–464 (2020)
2. Liu, B.: Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge University Press, (2020)
3. Wang, Y., Pal, A.: Detecting emotions in social media: a constrained optimization approach, In: Proceedings of the 24th International Conference on Artificial Intelligence, ser. IJCAI'15. AAAI Press, pp. 996–1002 (2015)
4. Liang, R., Wang, J.-Q.: A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in e-commerce. Int. J. Fuzzy Syst. **21**(3), 963–977 (2019)
5. Liu, B.: Sentiment analysis and opinion mining. synth. lect. hum. lang. technol. **5**(1), 1–167 (2012)
6. Cortis, K., Davis, B.: Over a decade of social opinion mining, arXiv e-prints, pp. arXiv-2012, (2020)
7. Zamzami, N., Bouguila, N.: High-dimensional count data clustering based on an exponential approximation to the multinomial beta-liouville distribution. Inf. Sci. **524**, 116–135 (2020)
8. Rubinstein, Y. D., Hastie, T., et al.: Discriminative vs informative learning. in KDD, vol. 5, pp. 49–53 (1997)
9. Perikos, I., Kardakis, S., Paraskevas, M., Hatzilygeroudis, I.: Hidden markov models for sentiment analysis in social media, In: 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science Engineering (BCD), pp. 130–135 (2019)
10. Nasfi, R.,Bouguila , N.: Online learning of inverted beta-Liouville HMMs for anomaly detection in crowd scenes, In: Hidden Markov Models and Applications. Springer, 2022, pp. 177–198 https://doi.org/10.1007/978-3-030-99142-5_7
11. Bdiri, T., Bouguila, N.: Bayesian learning of inverted dirichlet mixtures for svm kernels generation. Neural Comput. Appl. **23**(5), 1443–1458 (2013)

12. Wang, C., Zhao, X., Wu, Z., Liu, Y.: Motion pattern analysis in crowded scenes based on hybrid generative-discriminative feature maps, In 2013 IEEE International Conference on Image Processing. IEEE, pp. 2837–2841 (2013)
13. Feldman, R.: Techniques and applications for sentiment analysis. Commun. ACM **56**(4), 82–89 (2013)
14. Liu, N., Shen, B.: Rememnn: a novel memory neural network for powerful interaction in aspect-based sentiment analysis. Neurocomputing **395**, 66–77 (2020)
15. G. Gautam, G., Yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis, In: 2014 Seventh International Conference on Contemporary Computing (IC3), pp. 437–442 (2014)
16. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
17. Jin, W., Ho, H. H., Srihari, R. K.: Opinionminer: a novel machine learning system for web opinion mining and extraction, In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '09. New York, NY, USA: Association for Computing Machinery, pp. 1195–1204. (2009) https://doi.org/10.1145/1557019.1557148
18. Rabiner, L., Juang, B.: An introduction to hidden markov models, IEEE assp magazine, 3(1), pp. 4–16 (1986)
19. Lingappaiah, G.: On the generalised inverted dirichlet distribution. Demostratio Math. **9**(3), 423–433 (1976)
20. Harris, Z.S.: Distributional structure. Word **10**(2–3), 146–162 (1954)
21. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: understanding rating dimensions with review text, In: Proceedings of the 7th ACM Conference on Recommender Systems, ser. RecSys '13. Association for Computing Machinery, pp. 165–172 (2013)
22. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis, In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies Portland, Oregon, USA: Association for Computational Linguistics, pp. 142–150 June 2011 http://www.aclweb.org/anthology/P11-1015