ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2022.0397

SENTIMENT ANALYSIS USING VOTING BASED UNSUPERVISED ENSEMBLE MACHINE LEARNING IN CANCER DETECTION

M. Uma Maheswari and A. Aloysius

Department of Computer Science, St. Joseph's College, Affiliated to Bharathidasan University, Tiruchirappalli, India

Abstract

Within the field of natural language processing, sentiment analysis is one form of data mining used to make inferences about the emotional tenor of a speakers words. Computational linguistics is employed to examine the text in order to deduce and assess ones mental knowledge of the Web, social media, and associated references. One of the numerous advantages of sentiment analysis is that it can help improve the quality of healthcare by making use of medical data to produce the most positive outcome possible. Natural language processing challenges can change how sentiment analysis looks and works in a variety of contexts. Some of the challenges are specific to the data type, while others are universal to any method of text analysis. The primary objective of this study was to evaluate how challenging it is to analyse sentiment in the healthcare sector. Given the aforementioned complexities, the objective was to look into whether or not the currently available SA tools are adequate for handling any healthcare-related issue. With such motivation, in this paper, we develop an unsupervised ensemble machine learning (ML) algorithm that includes K-means clustering; Principle Component Analysis; Independent Component Analysis and k-nearest neighbors. The unsupervised ensemble ML model is assessed via voting meta-classifier over various cancer datasets. The simulation is conducted to test the efficacy of the model in terms of accuracy, precision, recall and f-measure over various datasets. The results of simulation against the cancer datasets show that the proposed method achieves higher rate of ensemble accuracy than the other existing ensemble models.

Keywords:

Natural Language Processing; Sentiment Analysis; Unsupervised ML

1. INTRODUCTION

To generate innovative medical ideas, it important to set aside time and resources for the study of emotions and how they might be classified. With no domain-specific dictionaries and no domain scholars eager in exploring this area of study, the challenges that can be accepted are substantial [1]. A further challenge is closing the information gap between knowledge-dependent features and the healthcare domain semantic linkages. The root cause of this is that, historically speaking, medical lexicons have lacked vital qualities like category and mood.

Experts have spent years developing data extraction technologies like PennBioIE 2 and GENIAI to address the problems that have arisen over this time. There is an immediate need for the creation of new versions of both organized and unstructured corpora. Linguistics, machine learning (ML), economics, and ontology are only few of the other disciplines whose contributions are used. Health care providers employ them to differentiate between syntactic and semantic features [2].

Recent studies have led to the development of two novel strategies for disentangling semantic links in the healthcare industry. The first method is known as tokenization (allotting the groups). For the second method, we examine how people feel in various medical settings and how those experiences shape their perspectives. A healthcare subject is just a term or word that denotes matters, details, and facts related to healthcare [3].

There are two main categories operating within the current market. The first set is concerned with one physical well-being, while the second set is not. The goal of this drill is to help you pinpoint the section of a sentence that houses both the negative and the period [4] [5].

The patient has regular headache complaints and tics. The success of any industry outside of healthcare depends entirely on healthcare ability to function. In addition to being a symptom of cancer in its early stages, headaches are also taken as an early warning sign of the disease. Therefore, it is a medical setting rich in medical terminology [6]. Additionally, our approach considers the larger corpus context for each sentence and word.

Classification and emotion recognition systems are the means by which distinct domains and settings can be distinguished from one another. Five separate classification schemes are used to organize this particular area of study [7]. The term headache is often used to describe a specific illness category. Researches and scientists from the healthcare industry contributed these five categories to the corpora.

What determined them was how the first instance of separation thinking was named and measured [8]. There is a different set of rules for how things inside a certain category ought to be structured for each distinct category. There are eleven subspecialties in medicine, each of which is divided into subspecialties based on pairings of concepts such disease symptoms and therapies [9].

Improves the model ability to process data based on concepts and contexts, allowing for more accurate sentiment analysis. When applied to a wide range of scenarios, emotion recognition algorithms yield wildly varying and even contradicting results.

The term anatomy of the human body is neutral in that it can be interpreted positively or negatively depending on the nature of the symptoms. When comparing these models to older versions of themselves, the lexicon-viz model is used. WordNet of Medical Events helps to isolate the medical field from other fields of study. The first emotions connected with medical thoughts are separated from the other groups using the negative character when the unigram and the bigram are used to arrange items into groups [10] [11].

In this paper, we develop an unsupervised ensemble machine learning (ML) algorithm that includes K-means clustering; Principle Component Analysis; Independent Component Analysis and k-nearest neighbors. The unsupervised ensemble ML model is assessed via voting meta-classifier over various cancer datasets.

2. RELATED WORKS

In order to devise the most productive treatment plan for patients, Leilei Sun [12] carried out a comprehensive examination of a substantial amount of information regarding various types of treatment. The purpose of this endeavor was to determine, through the use of a method known as quick semantic clustering, the degree to which the several treatment data sets were connected to one another. In addition, the author has developed a plan for evaluating whether or not the proposed treatment would be successful. This system is able to provide patients with the most effective treatment regimens since it takes into account the patient demographic information, medical history, and other significant criteria.

One of the tools that may be used is called GalenOWL, and it is an online framework that is enabled with semantics. It was described in the study [13] to help professionals locate information on the pharmaceuticals, and it is one of the tools that can be utilized. The findings of this study provide a model for deciding which treatments would be most beneficial for a particular patient by taking into account the patient specific infection in addition to any sensitivities they may have and any possible drug interactions. The model was developed by taking into account the patient specific infection in addition to any sensitivities they may have and any possible drug interactions. Because it connects clinical data with ontological concepts that are derived from clinical data and have been translated using international standards such as ICD-10 and UNII, GalenOWL is able to successfully perform its function.

Jiang et al. [14] investigated three distinct kinds of algorithms by applying them to treatment data: the decision tree technique, a support vector machine (SVM), and a backpropagation neural network. Due to its exceptional performance across all three distinct limits — model proficiency, model variety, and accuracy — SVM was ultimately chosen as the best option for the medication suggestion module. In addition to this, a method to check for errors was made available in order to ensure that accurate analysis and effective management could continue.

The investigation [15] is the idea that medical treatment ought to be adapted specifically to the capacities of the individual patient. To give just one example, if the patient immunity is weakened, the only medications that are likely to be of use should be prescribed to them by the doctor. It was recommended that risks be categorized in a certain way in order to arrive at a judgment regarding whether or not a patient have immunity. This conclusion would determine whether or not the patient possessed immunity. In addition, a web-based prototype of a system that uses a decision support system to help in the selection of first-line medications for patients was constructed. This system was constructed.

Li et al. [16] built a hashtag recommender system by training semantic sentence vectors with the usage of the skip-gram model and convolutional neural networks (CNN). These vectors make it feasible to classify hashtags with LSTM RNN. When typical artificial intelligence (AI) approaches are utilized, such as support vector machines (SVM) and collaborative filtering, the semantic properties are erased. This has a severe negative effect on the accuracy of the predictions that may be produced using these methodologies.

Hassan et al. [17] are the ones responsible for developing the top-N-related prescription medications. The collaborative filtering techniques, in which medication is first clustered as indicated by functional description data, are the source of the proposed framework origins. These approaches can be followed back to their point of origin. Following an investigation of the model inadequacies, a move is made to a strategy that takes use of tensor decomposition and is supported by the cloud. This update is done in order to boost the performance of the procedure that suggests drugs. A few instances of these shortcomings include that it is computationally expensive, that it needs to start from beginning, and that it has limited information.

After reviewing the relevant literature, researchers found that assessing a patient mental status while they are in the midst of a medical emergency is far more challenging than they had originally thought. Multiple factors have been taken into account in the analysis of sentiment. One of the most difficult parts of this job is learning to see things from other people perspectives. We provided a theoretical framework for assessing how individuals with a disease-related condition view themselves.

3. PROPOSED METHOD

In this section, the study unveils our proposed methods and provide a detailed description of each of their components.

3.1 DATA PRE-PROCESSING

In any data classification process, cleansing the raw data is the initial step. This strategy is crucial, yet it also flexible enough to work in many different contexts. Four primary pre-processing steps were applied to our socio-medical data in this study. It helps with better classification and paves the path for more powerful feature creation.

3.1.1 Conversion of Lowercase:

We switched to using only lowercase letters and in order to locate the terms that are all uppercase, a thorough search of the full dataset is required. If a word is found to be capitalized, the Numpy method from the Python library is used to convert it to lowercase.

3.1.2 Eliminate Special Characters:

As we move into the second phase of data preparation, at this point in the process, we will scrub the dataset by deleting all instances of non-alphanumeric characters. Following completion of Step 1, the uppercase free dataset will be required for Step 2.

3.1.3 Eliminate Stop Words:

The text typically contains numerous stop words yet conveys very little meaning. The study used an NLTK library function to get rid of the stop words.

3.1.4 Conversion of Number to word:

We used a Python module called num2words to accomplish the numerical-to-words conversion.

3.1.5 Stemming and Lemmatization:

Stemming is a technique for condensing words by eliminating their peripheral parts (the suffixes and prefixes) until just the core part (the stem) remains. This makes the meaning of the word

simpler. A porter stemmer was used to do this. When a word is lemmatized, it usually becomes a generic noun.

3.2 FEATURE EXTRACTION

Classification systems for sentiment analysis can be created once the text has been preprocessed. For machine learning algorithms to be useful, text must be converted to a numerical format first. More narrowly, the connections between numerical vectors. In this study, we employ the TF-IDF method, a common and straightforward strategy for extracting features from textual data. A model we called the manual feature was employed alongside TF-IDF. By employing feature engineering techniques, we were able to extract features from the reviews column and build this model.

The TF-IDF weighting approach is widely used because it provides meaningful insights beyond simple word counts by giving each word a relative importance. Note that the TF-IDF technique measures relevance, not recurrence, because it gives low weight to items that appear in the dataset more than once. The term frequency of a given document measures how frequently individual words and phrases appear within the text.

$$tf(t,d) = \log(1+\text{freq}(t,d))$$

The formula for determining the frequency of an event at a given time, $tf(t,d)$, is expressed as $\log(1+freq)$ because the frequency of an event is proportional to the logarithm of the product of its frequency and the time dimension.

$$idf(t,d) = \log(N \text{ count}(dD : td))$$

The IDF of a term is the inverse of its frequency of occurrence in the whole corpus. It recognizes that a phrase is affected by its setting.

$$tf\ idf(t, d, D) = tf(t,d).idf(t,D).$$

The importance and worth of a word in a document can be estimated by multiplying its total frequency (TF) by its inverse document frequency (IDF).

3.3 CLASSIFICATION

In this section, we use a voting meta-classifier to evaluate an unsupervised ensemble ML method that integrates the following and an ensemble ML model over various cancer datasets. This evaluation is carried out using an ensemble ML model. An ensemble machine learning model is utilized in the performance of this evaluation.

3.3.1 K-means Clustering:

An excellent illustration of an iterative method that can be utilized for the purpose of data organization is provided by the K-means algorithm. If the model have a dataset that has already been divided into K classes, the model may use distance as a measure to determine where the initial centroid for each of those classes ought to be by determining where it should be for the entire dataset. The Euclidean distance is used as the similarity index, and the clustering goals are to minimize the sum of squares of the different types of data for a given data set X with n multidimensional data points and a category K that needs to be split up. These goals are met by clustering the data in such a way that the Euclidean distance is used as the similarity index.

$$d = \sum_{k=1}^{k}\sum_{i=1}^{n}\|x_i - u_k\|^2$$

where

k - cluster centers,

u_k - kth center, and

x_i - i^{th} point in a cluster.

In order to successfully implement the algorithm, it is necessary to select K points from the sample set that are representative to serve as the nucleus of the first cluster. This is a key phase in the process. It is possible to determine the location of the center of each cluster by first dividing all of the sample points into equal portions and then assigning each portion to the cluster that has the center point that is geographically closest to it. In this way, the location of the center point of each cluster can be calculated. Proceed with the iteration until either the point at the center of the cluster has not changed or the number of times that have been set has been reached. The output of the algorithm can fluctuate due to the fact that it is dependent on the starting position. This makes the result of the algorithm unpredictable. The K value, which is the primary concern of the algorithm, is what determines the center of the cluster and has an immediate influence on the degree to which the clustering procedure achieves both its local and its global optimality.

3.3.2 Principle Component Analysis:

Principal component analysis (PCA) is a well-liked technique for finding patterns in vast amounts of data and extracting features from those datasets. This is due to the fact that it is both user-friendly and effective at compressing images and reducing the dimensions of high-dimensional data. An orthogonal linear model that designs high-dimensional data on a low-dimensional space (the principal component, or PC) while simultaneously maximizing the variance of the data and reducing the mean squared reconstruction error is the goal of principle component analysis (PCA), which seeks to discover such a model. Maximizing the variance of the data and reducing the mean squared reconstruction error are the two main objectives of PCA. Principal component analysis (PCA) is another name for this statistical technique. In order to put this concept into action, it is important to determine not only the eigenvalues of the covariance matrix but also the eigenvectors of the matrix (CM).

In this investigation, the principal component analysis (PCA) is utilized in order to determine the eigenvalues and eigenvectors of a group of curves that are located in a space that has a limited number of dimensions. The goal of this investigation is to determine the eigenvalues and eigenvectors of these curves. Assuming that each database uses an emotional analysis based on a column vector Fi, the length of which is composed of n eigenvectors that lie in the curves of power consumption for electrical appliances in their original spatial context, let say that each database also uses n eigenvectors that lie in the curves of power consumption for electrical appliances. This would be consistent with the previous assumption that each database uses n eigenvectors that lie in the curves of power consumption for electrical It is possible to generate an F-matrix with the dimensions nm for every element of the Fi vector that depicts a power consumption curve. This matrix may be used to analyze the data.

$$F=[F_1,F_2,F_3,\ldots,F_m]$$

Through the application of principal component analysis, the F data matrix can be transformed into a space that has less dimensions.

$$P=H^TF$$

Where F stands for the data matrix, H for the scheme matrix that contains columns that are built using eigenvectors, and H^T for the transpose of H.

The steps of PCA are as follows:

- A rather accurate estimate of the CM,
- The eigen-dissociation of the CM and the selection of the eigenvalue with the *K*-highest rank
- Producing the feature matrix, *I* by making use of the appropriate eigenvectors,
- Mapping the curves onto the *k*-dimensional vector space by applying the I operator

3.3.3 Independent Component Analysis:

The factorization family is a family of matrix factorization techniques, and independent component analysis is a technique that belongs to that family. Each of these processes takes as an input a rectangular matrix X of measurements (in a sufficiently large number of observed samples, N, and with a number of observed features, m), and it approximates the matrix as the sum of products of p pairs of vectors, each of which has lengths N and m. The procedures also take as an output an approximation of the rectangular matrix. To put it another way, each of these processes treats the rectangular matrix of measurements denoted by X as though it were a vector of the dimensions N and m. When the a_k and s_k vectors are multiplied together, the resulting matrix has the same dimension as X and a rank of one. This is a crucial fact to keep in mind. The equation that is shown below is considered to be the fundamental equation for all possible methods of factorizing matrices:

$$\sum_{k=1}^{p} a_k \times s_k \,(*)$$

Therefore, the difficulty of matrix factorization is in locating a set of $\boldsymbol{a_k}$ and $\boldsymbol{s_k}$ values for which all of the following requirements may be satisfied:

$$\left\| X - \sum_{k=1}^{p} a_k \times s_k \right\|^2 \rightarrow \min(**)$$

where ||..|| denotes a valid matrix norm, which is often the sum of the Euclidean norms of the columns of the matrix. Norms of matrices are sometimes referred to as matrix norms.

3.3.4 K-Nearest Neighbors:

The Euclidean distance is used as the primary selection criterion for group membership in the vast majority of implementations of the "k-nearest-neighbor" classifier. The test sample is compared to the training samples that are provided, and the distance between them is determined. Allow x_i to serve as an illustration of an input sample, and stipulate that it possesses p characteristics (xi_1, xi_2,…, xi_p).

Also, let say that n is the total number of input samples (i=1,2,…,n) and p is the total number of features (j=1,2,…,p) that are being used. The following is the formula that can be used to calculate the Euclidean distance between two samples, $\mathbf{x}i$ and $\mathbf{x}l$, where l can range anywhere from 1 to n.

$$d(\mathbf{x}i,\mathbf{x}l)=(xi_1-xl_1)^2+(xi_2-xl_2)^2+\cdots+(xip-xlp)^2$$

Statistical analysis makes use of something called a Voronoi cell, which is a cube with three dimensions.

$$Ri=\{\mathbf{x}\in R_p: d(\mathbf{x},\mathbf{x}i)\leq d(\mathbf{x},\mathbf{x}m),\forall i\neq m\},$$

where Ri is the Voronoi cell for sample $\mathbf{x}_i$, and $\mathbf{x}$ represents all possible points within Voronoi cell Ri. This includes all of the points that are contained within the cell. The primary characteristics of a coordinate system that are reflected in Voronoi tessellations are I the fact that any point within a Voronoi cell is the nearest neighboring point for the sample in question, and (ii) the fact that the edge of any Voronoi cell can be used to determine the location of the sample that is closest to it.

Both of these characteristics can be found in the fact that any point within a Voronoi cell is the nearest neighboring point for the sample in question. Both of these qualities can be observed as a result of the fact that any point that is contained within a Voronoi cell is the point that is immediately adjacent to it. Putting a test sample in the same category as the majority of its k closest training samples is what the k-nearest-neighbor classification method does to a sample.

This rule is predicated on the latter quality, and it uses that quality as the primary basis for its conclusions. It is typical to select an odd value for k in order to remove any potential of a tie developing in the voting process. In categorization theory, the k = 1 rule is sometimes referred to as the nearest-neighbor categorization rule. This is the moniker that is used most frequently for it.

3.4 ENSEMBLE VOTING

The procedure of soft ensemble voting involves imagining the class names combined with the expected probability p for the classifier. This is done in order to reach a conclusion. In the event that the classifiers have been brought up to par, this is a process that may be suggested as a course of action.

$$Y=\arg\max_i \sum_{j=1}^{m} W_j P_{ij}, i=\{0,1\}; j=1,2...,m$$

where W_j – values of j^{th} classifier.

4. RESULTS AND DISCUSSIONS

Datasets are derived from user reviews and ratings submitted to the website http://www.imdb.com/. The training set consists of 2421 reviews, whereas the test set consists of only 500 reviews.

Despite the limited quantity of the training dataset, J48 achieves flawless classification accuracy in some iterations. In Table.1, we see a summary of the 29 iterations and the mean performance of the four classifiers applied to the second and third datasets. J48 has the lowest learning speed of the five classifiers, despite having the highest overall performance.

Compared to the other three classifiers, OneR performance with the J48 classifier is higher in terms of the proportion of correctly classified cases. The J48 algorithm shows some promise for accuracy when both the true positive and false positive rates are taken into account.

Table.1. Class Performance Analysis

Documents	Classes			
	TP	FP	TN	FN
100	24,379	8513	10,374	5954
200	24,680	8439	10,448	5653
300	25,496	8554	10,333	4837
400	19,783	6492	18,766	4179
500	20,127	6435	18,823	3835
600	20,932	6628	18,630	3030

Table.2. Performance Metrics

Documents	Metrics			
	Accuracy	Recall	F-Measure	Precision
100	94.34	90.23	91.51	90.43
200	93.23	91.74	91.46	90.99
300	93.56	90.38	91.22	90.12
400	92.51	91.00	91.01	89.83
500	92.23	90.41	90.54	89.36
600	91.93	90.54	90.23	8.12

Using a variety of variables, we selected the top four estimates and averaged them to arrive at our final forecast. Then, we calculated an aggregate score for the therapies that were successful in curing that condition by multiplying the total number of results by the normalized number of useable units. Generally speaking, the higher the rating, the better the medicine. According to the theory, more respondents will take the poll if they are hoping to get specific prescriptions. This holds true whether or not the feedback is positive. Therefore, we standardized the helpful counts for each condition as we developed the recommender system.

The primary goal is to verify that all four models agree on how to categorize the most often used pharmaceuticals. There must be something wrong with the model if the total rating for a medicine drops because it was incorrectly predicted by one of the models. The total score for each drug was calculated by multiplying the sum of its predictions by the number of times it was deemed effective. A sufficient number of reviewers for the medicine in question was ensured by these actions. Mean scores are calculated by dividing the overall score by the total number of drugs used to treat each illness.

Although all four methods yield encouraging outcomes, this is not yet proof that the recommender architecture is ready for widespread implementation. The situation may be better. It was forecasted that algorithms would be necessary to strike a balance between the positive and negative class metrics. Classification methods can yield more accurate models with some fine-tuning of the hyperparameters.

5. CONCLUSIONS

In this paper, we develop an unsupervised ensemble machine learning (ML) algorithm that includes K-means clustering; Principle Component Analysis; Independent Component Analysis and k-nearest neighbors. The unsupervised ensemble ML model is assessed via voting meta-classifier over various cancer datasets. We also compared our proposed practical framework to a reference approach, which allowed us to show how well it would perform in gauging the emotions of persons whose lives had been upended by an unanticipated illness. One of our long-term aims is to integrate a deep learning strategy with this method so that we can deal with a more complete range of dataset elements.

REFERENCES

[1] W.M. Yafooz and A. Alsaeedi, "Sentimental Analysis on Health-Related Information with Improving Model Performance using Machine Learning", *Journal of Computer Science*, Vol. 17, No. 2, pp. 112-122, 2021.

[2] S.M. Srinivasan, P. Shah and S.S. Surendra, "An Approach to Enhance Business Intelligence and Operations by Sentimental Analysis", *Journal of System and Management Sciences*, Vol. 11, No. 3, pp. 27-40, 2021.

[3] M.H. Song, "A Study on Explainable Artificial Intelligence-based Sentimental Analysis System Model", *International Journal of Internet, Broadcasting and Communication*, Vol. 14, No. 1, pp. 142-151, 2022.

[4] H.S. Saraswathi and C.K. Raju, "Computer-Aided Diagnosis of Pancreatic Ductal Adenocarcinoma Using Machine Learning Techniques", *Proceedings of International Conference on Sentimental Analysis and Deep Learning*, pp. 959-972, 2022.

[5] N. Banerjee and S. Das, "Lung Cancer Prediction in Deep Learning Perspective", Proceedings of International Conference on Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, pp. 237-255, 2021.

[6] S.A. Alanazi, A. Khaliq, F. Ahmad and S. Afsar, "Publics Mental Health Monitoring via Sentimental Analysis of Financial Text Using Machine Learning Techniques", *International Journal of Environmental Research and Public Health*, Vol. 19, No. 15, pp. 9695-9703, 2022.

[7] M. Kentour and J. Lu, "An Investigation into the Deep Learning Approach in Sentimental Analysis using Graph-Based Theories", *Plos One*, Vol. 16, No. 12, pp. 1-8, 2021.

[8] S.K. Pathuri and J. You, "Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model", *Sensors*, Vol. 22, No. 1, pp. 80-95, 2021.

[9] P. Taranath, S. Das and S. Gowrishankar, "Analysis of Healthcare Industry Using Machine Learning Approach: A Case Study in Bengaluru Region", *Sentimental Analysis and Deep Learning*, pp. 1-13, 2022.

[10] L. Lyu, "Lung Cancer Diagnosis Based on Convolutional Neural Networks Ensemble Model", *Proceedings of International Conference on Artificial Intelligence, Networking and Information Technology*, pp. 360-367, 2021.

[11] A.K. Saha and M. Rahman, "An Efficient Deep Learning Approach for Detecting Pneumonia Using the Convolutional Neural Network", *Proceedings of International Conference on Sentimental Analysis and Deep Learning*, pp. 59-68, 2022.

[12] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong and Yanming Xie, "Data-driven Automatic Treatment Regimen Development and Recommendation", *Proceedings of International Conference on Knowledge Discovery and Data Mining*, pp. 1865-1874, 2016.

[13] C. Doulaverakis, "GalenOWL: Ontology-Based Drug Recommendations Discovery", *Journal of Biomedical Semantics*, Vol. 3, pp. 1-14, 2012.

[14] Y. Bao and X. Jiang, "An Intelligent Medicine Recommender System Framework", *Proceedings of IEEE International Conference on Industrial Electronics and Applications*, pp. 1-8, 2016.

[15] K. Shimada, "Drug-Recommendation System for Patients with Infectious Diseases", *Proceedings of IEEE International Conference on Machine Learning*, pp. 1-7, 2005.

[16] J. Li, H. Xu, X. He, J. Deng and X. Sun, "Tweet Modeling with LSTM Recurrent Neural Networks for Hashtag Recommendation", *Proceedings of IEEE International Conference on Neural Networks*, pp. 1570-1577, 2016.

[17] Zhang, Yin and Limei Peng, "CADRE: Cloud-Assisted Drug Recommendation Service for Online Pharmacies", *Mobile Networks and Applications*, Vol. 20, pp. 348-355, 2014.