Available online at www.sciencedirect.com

Journal of Applied Research and Technology

Journal of Applied Research and Technology 15 (2017) 311–319

www.jart.ccadet.unam.mx

Original

Patient opinion mining to analyze drugs satisfaction using supervised learning

Vinodhini Gopalakrishnan *, Chandrasekaran Ramaswamy

Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar 608002, India

Received 4 August 2015; accepted 8 February 2017
Available online 2 August 2017

Abstract

Opinion mining is a very challenging problem, since user generated content is described in various complex ways using natural language. In opinion mining, most of the researchers have worked on general domains such as electronic products, movies, and restaurants reviews but not much on health and medical domains. Patients using drugs are often looking for stories from patients like them on the internet which they cannot always find among their friends and family. Few studies investigating the impact of social media on patients have shown that for some health problems, online community support results in a positive effect. The opinion mining method employed in this work focuses on predicting the drug satisfaction level among the other patients who already experienced the effect of a drug. This work aims to apply neural network based methods for opinion mining from social web in health care domain. We have extracted the reviews of two different drugs. Experimental analysis is done to analyze the performance of classification methods on reviews of two different drugs. The results demonstrate that neural network based opinion mining approach outperforms the support vector machine method in terms of precision, recall and *f*-score. It is also shown that the performance of radial basis function neural network method is superior than probabilistic neural network method in terms of the performance measures used.

Keywords: Sentiment; Opinion; Health; Drugs; Classification

1. Introduction

Natural language processing (NLP) is a set of computational techniques for analyzing natural language texts that allows computers to understand human language. Opinion mining discipline places itself at the crossroads of information retrieval and computational linguistics. These are the two fields from which opinion mining gathers and combines many concepts, ideas and methods (Yu & Hatzivassiloglou, 2003; Wilson, Wiebe, & Hoffmann, 2005; Xia, Zong, & Li, 2011; Ye, Zhang, & Law, 2009). The primary unit of NLP is the language term. Each language term has various linguistic features, such as grammatical category, meaning, sense, co-occurrence similarity, and contextual relationships that are employed for term classification and subjectivity analysis. The NLP tasks require a knowledge base for information extraction and analysis. While some techniques are necessary for building a knowledge base, other techniques use existing knowledge bases to analyze documents. Researchers have presented great contributions in this area and diverse approaches have been employed to accomplish the opinion mining (Miao, Li, & Dai, 2009; Tang, Tan, & Cheng, 2009).

Opinion mining has led to development of a significant number of tools that are used to analyze consumer opinion for most major business environments including travel, housing, consumables, materials, products, services, and many others. But not much work was carried out on health care domain (Tang et al., 2009; Vinodhini & Chandrasekaran, 2014; Xu, Liao, Li, & Song, 2011). Drug surveillance is a major factor of drug safety once a drug has been released to the public use. Drug trials are often done in limited number of test subjects where the probability to detect uncommon adverse effects is minimal. Also the volunteers or patients participating in drug trials are also different from those receiving licensed medications, differing in age, co-morbidity and poly-pharmacy. Thus it is necessary to study

* Corresponding author.
E-mail address: g.t.vino@gmail.com (V. Gopalakrishnan).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

http://dx.doi.org/10.1016/j.jart.2017.02.005

the safety of marketed drugs on an epidemiological scale. It is also important to understand how the general population uses a particular drug, perceive its safety, reactions and efficiency. The objective of this work is to use neural network methods for opinion classification in health care domain.

This paper is outlined as follows. Section 2 narrates the related work. Section 3 discusses the methodology used to develop the models. The data source used is reported in Section 4. The various classification methods used are introduced in Section 5. Sections 6 and 7 discuss the results. Section 8 concludes the work.

2. Literature review

Many studies on opinion classification have used machine learning algorithms with support vector machine (SVM) being the most commonly used. SVM has been used extensively for opinion miming on online reviews (Baccianella, Esuli, & Sebastiani, 2009; Jian, Chen, & Wang, 2010; Pang, Lee, & Vaithyanathan, 2002; Tan & Zhang, 2008; Tan & Zhang, 2008; Xia et al., 2011; Ye, Lin, & Li, 2005; Ye et al., 2009). Some research works exists for opinion mining in medical domain using machine learning techniques. Greaves, Ramirez-Cano, Millett, Darzi, and Donaldson (2012) analyzed patient's opinions about different performance aspects of hospitals in the United Kingdom. Ali, Sokolova, Schramm, and Inkpen (2013) used a subjectivity lexicon and machine learning algorithms to sentiment analysis of messages posted on forums dedicated to hearing loss. Sokolova and Bobicev (2011) used supervised machine learning methods to analyze sentiments and opinions expressed in health related user written texts. They analyzed opinions posted on a general medical forum. Sokolova and Bobicev (2011) presented a mining method for personal health information in Twitter. Wang, Chen, Tan, Wang, and Sheth (2012) proposed a combination of machine learning and rule based classifiers for opinion classification in suicide notes.

There are also very few studies specifically on opinion mining in the drug domain. Na et al. (2012) proposed a rule-based system for polarity classification of drug reviews. Yalamanchi (2011) developed a system called SidEffective to analyze patient's sentiments about a particular drug. Goeuriot et al. (2012) built a health-related sentiment lexicon and used it for polarity classification of drug reviews. Wiley, Jin, Hristidis, and Esterling (2014) used SentiWordNet and word-emotion lexicons for opinion classification of drug reviews of online social networks. Sharif, Zaffar, Abbasi, and Zimbra (2014) developed a framework that extracts important semantic and sentiment, that are better able to analyze the experiences of people when they discuss adverse drug reactions as well as the severity and the emotional impact of their experiences. Xia, Gentile, Munro, and Iria (2009) proposed a multi-step approach, where in the initial step a standard topic classifier is learned from the data and the topic labels, and in the ensuing step several polarity classifiers, one per topic, are learned from the data and the polarity labels.

2.1. Motivation and contributions

In recent years, we witnessed the advance in neural network methodology, like fast training algorithm for deep multilayer neural networks (Chen, Liu, & Chiu, 2011; Ghiassi, Skinner, & Zimbra, 2013; Jian et al., 2010; Luong, Socher, & Manning, 2013; Moraes, Valiati, & Neto, 2013; Noferesti & Shamsfard, 2015; Sharma & Dey, 2012). Neural network techniques have found success in several NLP tasks recently such as opinion mining (Bobicev, Sokolova, Jafer, & Schramm, 2012; Socher, Lin, Manning, & Ng, 2011). Socher et al. (2011) also introduced prediction architecture based on recursive neural networks that can be used to provide a competitive syntactic parser for natural language sentences from the Penn Treebank.

However, the literature does not contribute much work in opinion classification using neural networks, especially probabilistic neural network (PNN) and radial basis function neural network (RBFN). But many researchers have proved that PNN and RBFN model is more effective than other models for data classification in various other domains (Adeli & Panakkat, 2009; Hajmeer & Basheer, 2002; Savchenko, 2013).

This motivates us to investigate the effectiveness of neural networks in multi class opinion classification. Based on the literature survey done, the major contributions of our work are as follows, PNN and RBFN based on function approximation are used for opinion classification which is not considered so far in opinion mining literature. The results obtained for neural network based method are compared with SVM as baseline method. We have applied opinion mining to health care domain, particularly focusing on public opinions on drugs.

3. Problem design

The following is the summary of our methodology for developing and validating the prediction models (Fig. 1).

i. Perform data preprocessing.
ii. Construct the vector space representation for the drug reviews selected for drug I and drug II.
iii. Apply the following classification methods using the respective training data set
   a. SVM.
   b. The neural network model using PNN.
   c. The neural network model using RBFN.
iv. Predict the class (positive or negative) of each review in the test data set.
v. Compare the prediction results with actual values.

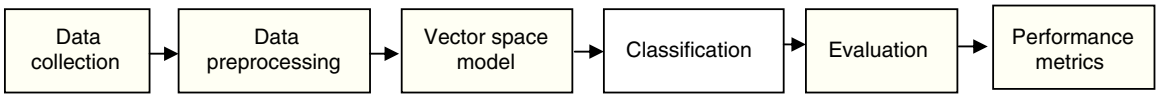

Fig. 1. Problem design.

Rating	Reason	Side effects	Comments	Sex	Age	Duration	Date
4	Depression fibromyalgia	Heart arrhythmia (skipping beats) Insomnia	Was amazed and delighted how quickly symptoms disappeared in the first week. I have struggled to lose since menopause.	F	59	6 weeks 30mg	5/23/2015

Fig. 2. Sample review from www.askapatient.com.

vi. Compute the various quality parameters and compare the prediction results.

4. Data source

We constructed the drug review dataset from the popular web resource www.askapatient.com. Figure 2 shows the sample review format from www.askapatient.com. Opinion mining is proved to be domain specific due to the nature of the language used. But the focus of this work is on opinion mining in single domain i.e. drug reviews. We constructed two datasets (Dataset I and Dataset II) by collecting the reviews of two popular drugs such as cymbalta and depo-provera. Cymbalta is used to treat major depressive disorder in adults. Depo-Provera is a form of progesterone, a female hormone used as contraception to prevent pregnancy. These two drugs differ in their nature of use. The natural language and medical terms used may also differ. Two different dataset are used for evaluation so as to provide a more solid evidence of the evaluation to be carried out. The choice of these drugs was chosen randomly from the list of the most frequently rated drugs in the review website. We manually constructed two different datasets (Dataset I and Dataset II) for the drugs chosen.

We automated the extraction of the comments from each user review along with rating score using Java. Our automated system extracts only the comments and rating details for each review. Each user comment is stored as separate text document, which are considered as samples for training and testing in learning process. For each of the textual content, we initially employed the Stanford Core NLP to detect sentences. For supervised machine learning approaches, need for labeled data is necessary in classifying the opinion of reviews. The text reviews are labeled based on the user ratings. The text review contains written text to explain the rating score of the review. Since the reviews are provided by various individuals with ranging backgrounds the meaning of a 5 star review changes from person to person. Then the document extracted is assigned opinion class based on the review score in the data format The opinion class is positive if review score is >3, if review score is <3 the class label is assigned negative and else the class label is assigned neutral. Among 2379 reviews of drug cymbalta, 620 review documents are neutral, 890documents are positive reviews, and 869 documents are negative reviews. Similarly for the second drug (Depo-Provera) chosen, out of 2247 ratings, 602 reviews are neutral, 902are positive reviews and 743 are negative reviews. To have a balanced data distribution we have randomly selected 600 positive, 600 negative reviews and 600 neutral reviews for each drugs to establish the corpus. The description datasets used is shown in Table 1.

To create the vector space representation for classifiers, the review documents are pre processed. We used Rapid Miner for preprocessing of review documents (Fig. 3(a)). Figure 3(b) represents the sub process flow for vector creation of the process document component in Figure 3(a). Previous studies have demonstrated that these preprocessing steps can improve the performance of text retrieval, classification and summarization (Xu et al., 2011). The steps involved in data preprocessing are tokenization, transformation, stop words removal and stemming.

Subsequently POS tagging is performed for each sentence. By adopting the POS tagging mentioned in Hu and Liu (2004), we selected all noun and noun phrases. The identified noun and noun phrases are grouped as ngrams. The next step is mining frequent n-grams from the processed texts. We employ a priori association mining technique in text reviews to identify frequent ngrams. The resulting unigrams, bigrams and trigrams are identified as features for vector space representation. Term frequency inverse document frequency (TF-F) is used as document weighting method.

5. Methods

This section gives a brief discussion of the methods used in this work to develop the opinion classification system. SVM is used as the baseline method. SVM is employed using Rapid Miner tool. Function approximation, which finds the underlying relationship from a given finite input output data is the fundamental problem in a vast majority of real world applications, such as prediction, pattern recognition, data mining and classification. There exist multiple methods that have been established as function approximation tools, where

Table 1
Properties of data source.

Drug	Data model	No. of reviews	Positive reviews	Negative reviews	Neutral reviews
Cymbalta	Dataset I	1800	600	600	600
Depo-provera	Dataset II	1800	600	600	600

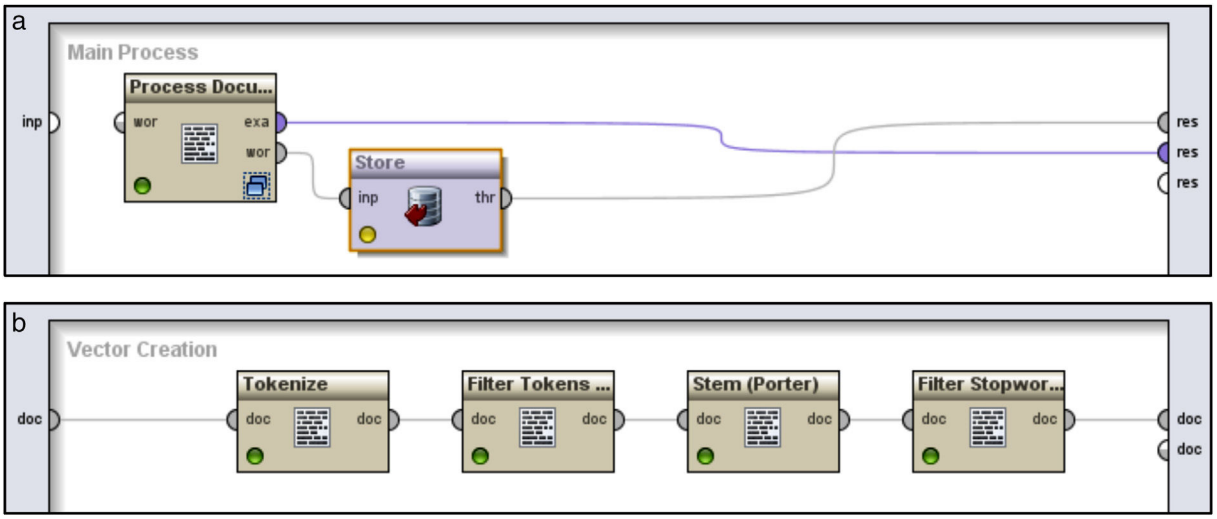

Fig. 3. (a) Rapid miner main process flow for preprocessing. (b) Rapid miner subprocess flow for vector creation.

an artificial neural network (ANNs) is one of them. In this paper, two different neural network based methods such as PNN and RBFN have been applied as function approximation tools.

SVM is powerful classifier arising from statistical learning theory that has proven to be efficient for various classification tasks in text categorization (Zhang, Ye, Zhang, & Li, 2011). SVM model is employed using weka tool. The SVM is employed with default values for most of the parameters available in the Weka tool. The kernel type chosen is a polynomial kernel. The size of the cache (kernel) is 1. The exponent value of kernel is 1.0 and the complexity parameter (C) is set to be 10.0.

5.1. Probabilistic neural network (PNN)

A PNN is based on statistical bayesian classification algorithm. The functions are organized into multilayered feed forward network with four layers such as input layer, pattern layer, summation layer and output layer. The input layer consists of input nodes, which are the set of measurements. The pattern layer is fully connected to the input layer, with one neuron for each pattern in the training set. The pattern layer outputs are selectively connected to the summation units depending on the class of patterns. Although SVM performs well in opinion classification in most cases, its drawback is high computational cost in finding the best parameter combinations. But PNN is simple and is easily trained, because it has only one parameter to be optimized in the experiment. Therefore, selecting PNN as classifier is a best choice from the viewpoint of parameter optimization (Adeli & Panakkat, 2009; Hajmeer & Basheer, 2002; Savchenko, 2013) Given an unknown sample x, we can compute its posterior probability $P(c_i|x)$ to determine which class label the sample x belongs to. According to Bayesian rule, $P(c_i|x)$ is proportional to the multiplication of all prior probability π_i by probability density function $f_i(x)$. That can be represented as: $P(c_i|x)\alpha\pi_i f_i(x)$. Let m be the number of training samples, n the number of genes, x_{ij} the jth training sample for class i, and k_i the number of samples of class i. The Parzen estimate probability density function for class i can be written as (Eq. (1))

$$f_i(x) = \frac{1}{(2\pi)^{m/2}\sigma^m}\frac{1}{k_i}\sum_{j-1}^{k_i}\exp\left[-\frac{(x-x_{ij})^{\mathrm{T}}(x-x_{ij})}{2\sigma^2}\right] \quad (1)$$

where σ is a smoothing parameter. The choice of σ has a great effect on the estimation error of the PNN classifier and is determined experimentally by comparing their corresponding classification accuracy rates. If there is no way to obtain the prior probability, the prior probability can be evaluated by the following formula

$$\pi_i = \frac{k_i}{\sum_{j-1}^{c} k_j} \quad (2)$$

where C denotes the number of subclasses in dataset. The datasets used in this work has only three sub classes: positive, neutral and negative (Savchenko, 2013).

PNN is implemented using the KNIME tool. The size of the training dataset is the number of neurons in the hidden layer. The smoothing factor value is 1 for the Datasets I and II. The cross validation meta node encapsulates an inner workflow shown in Figure 4 which is executed several times. The results of each iteration are collected and presented in aggregate through the output ports of the cross validation meta node. The first output port reports the predicted class values for each row while the second output port reports the error rates for each of the iteration.

5.2. Radial basis function neural networks

The radial basis function (RBF) neural network is a feedforward network with its architecture consists of three layers: an input layer of neurons, a hidden radial basis layer of neurons and an output linear layer of neurons. The information of the input neurons will transfer to the neurons in the hidden layer. The RBF in the hidden layer will response to input information, and then generates the outputs in the neurons of the output layer. The advantage of the RBF network is that the hidden neurons

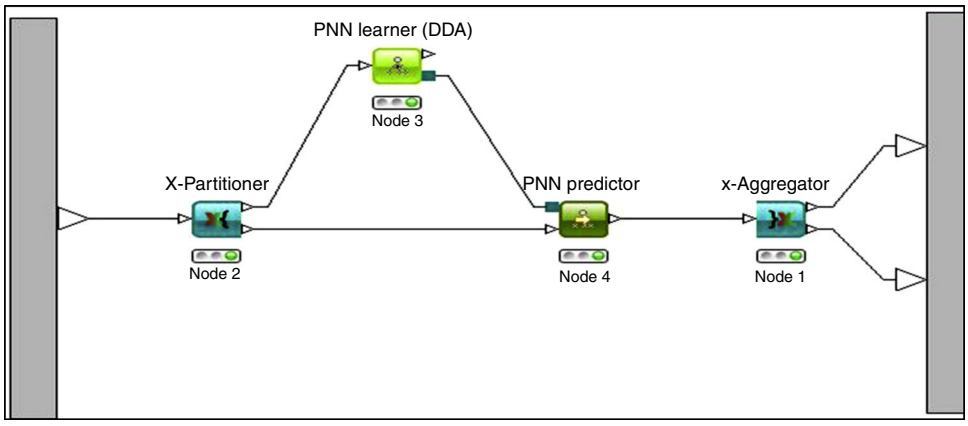

Fig. 4. Work flow of PNN (Knime).

will have non-zero response if the minimum of a function is in the pre defined limited range of the input values, otherwise, the response will be zero. Therefore, the number of active neurons is smaller and the time required in training the network is less. Hence, the RBF network is also referred to as the local range network. In the RBF neural network, the transfer function of the hidden layer is a Gaussian function

$$A_j = \exp\left[-\frac{(P - C_j)^{\mathrm{T}}(P - C_j)}{2\sigma_j^2}\right], \quad j = 1, 2.\ldots..., S^1 \quad (3)$$

where A_j is the output of the jth neuron in the hidden layer, p is the input mode, c_j is the center of the jth neuron Gaussian function. σ_j^2 is the unitary parameter, and S^1 is the neurons in hidden layer 1. There are two steps in the training of the RBF network. In the first step, depending on the information contained in the input samples, the neurons in the hidden layer S^1; the center of Gaussian function c_j and the unitary parameter σ_j^2 will be determined. The most frequently used method to determine the Gaussian function is the K-means aggregation algorithm. Assume that the K-means aggregation algorithm aggregates the input samples, θ_j are all the samples of the jth group, then the center of the jth aggregation c_j is given by the following Eq. (4):

$$C_j = \frac{1}{M_j}\sum_{x\in\theta_j} x \quad (4)$$

And the jth unitary parameter σ_j^2 is given by the following equation:

$$\sigma_j^2 = \frac{1}{M_j}\sum_{x\in\theta_j} (x - C_j)^{\mathrm{T}}(x - C_j) \quad (5)$$

In Eq. (5), M_j is the mode of θ_j and σ_j^2 is the estimate of samples that distributed the center of the jth aggregation, c_j. In the second step, according to the parameter in the hidden layer, input samples and the target values, the weight w_j will be determined and adjusted by the principle of least squares. The RBF network has a strong capability of approximation to the kernel vector in a limited part of the whole net, These networks have the advantage of being much simpler than the perceptrons while keeping the major property of universal approximation of functions. RBFN model is employed using weka tool with default values for the parameters. The knowledge flow model of RBFN is shown in Figure 5.

6. Results

Many approaches for evaluating the quality of the opinion classification methods are used. The methods are validated using 10-fold cross validation. In this work, the results obtained for the test dataset are evaluated using various model quality parameters on each single class label. In this particular domain of drug reviews, the positive, negative and neutral reviews are equally important in making a decision on the drug usage. So, it is necessary to measure the classification performance of all positive, negative and neutral reviews independently. Thus the precision, recall and f-score of the classifiers on each single class label (positive/negative/neutral) are measured. Tables 2–5 summarize the results obtained as confusion matrices of all the classification methods used on each single class label. The various measures used for evaluation are as follows (Briand, Wüst, Daly, & Porter, 2000)

6.1. Misclassification rate

Misclassification rate is defined as the ratio of number of wrongly classified reviews to the total number of reviews classified by the prediction method. The wrong classification falls into three categories. C1 represents number of negative and neutral reviews classified as positive. C2 represents number of positive and neutral reviews classified as negative. C3 represents number of positive and negative reviews classified as neutral.

$$\text{Type I error} = \frac{\text{C1}}{(\text{Total no of positive reviews})}$$

$$\text{Type II error} = \frac{\text{C2}}{(\text{Total no of negative reviews})}$$

$$\text{Type III error} = \frac{\text{C3}}{(\text{Total no of neutral reviews})}$$

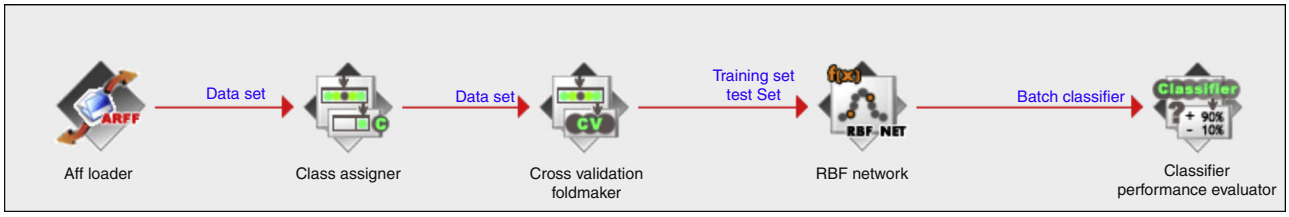

Fig. 5. Knowledge flow model of RBFN (Weka).

Table 2
Confusion matrix for SVM.

Actual	Dataset I (predicted)			Dataset II (predicted)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Positive	483	62	55	495	57	48
Negative	68	478	54	63	487	50
Neutral	54	65	481	52	57	491
Error	TypeI	TypeII	TypeIII	Type I	Type II	TypeIII
	20.3%	21.2%	18.7%	19.2%	19.0%	16.3%
	Misclassification rate: 19.9%			Misclassification rate: 18.2%		

Table 3
Confusion matrix for PNN.

Actual	Dataset I (predicted)			Dataset II (predicted)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Positive	513	43	44	530	32	38
Negative	46	509	45	36	517	47
Neutral	42	37	521	32	44	524
Error	Type I	Type II	Type III	Type I	Type II	TypeIII
	14.7%	13.3%	14.8%	11.3%	12.7%	14.2%
	Misclassification rate: 14.3%			Misclassification rate: 12.7%		

$$\text{Misclassification rate} = \frac{C1 + C2 + C3}{(\text{Total no of reviews})}$$

6.2. Precision

Precision (also known as correctness) is the percentage of classified instances that are correct (Briand et al., 2000).

$$\text{Precision} = \frac{\text{No. of correctly classified reviews}}{\text{Total no. of classified reviews}}$$

Precision of positive reviews is defined the number of reviews correctly classified as positive to the total number of reviews classified as positive. Low Precision means that a high percentage of the classes being classified as positive, which is not actually positive. So there is a waste of developers' effort in correcting them. Hence, Precision is expected to be high always.

6.3. Recall

Recall is defined as the ratio of number of positive reviews classified correctly to the total number of reviews.

$$\text{Recall} = \frac{\text{No. of correctly classified reviews}}{\text{Total no. of reviews}}$$

Table 4
Confusion Matrix for RBFN.

Actual	Dataset I (predicted)			Dataset II (predicted)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Positive	538	23	39	545	21	34
Negative	21	526	53	16	534	50
Neutral	24	49	527	20	43	537
Error	Type I	Type II	Type III	Type I	Type II	TypeIII
	7.5%	12.0%	15.3%	6.0%	10.7%	14.0%
	Misclassification rate: 11.6%			Misclassification rate: 10.2%		

Table 5
Precision of classifiers.

	Dataset I (%)			Dataset II (%)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
SVM	79.8	79.0	81.5	81.1	81.0	83.4
PNN	85.4	86.4	85.4	88.6	87.2	86.0
RBFN	92.3	88.0	85.4	93.8	89.3	86.5

6.4. *f*-Score

f-Score is a measure that combines precision and recall. It is the harmonic mean of precision and recall

$$\text{f-Score} = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}}$$

The classification results are presented as confusion matrix in Tables 2–4.

Tables 2–4 summarize the misclassification results of all classification methods for datasets (Dataset I & II). Table 2 gives the classification results in terms of error measures for SVM method. PNN prediction results are presented in Table 3. Table 4 shows the results of RBFN classification. The type I, type II and type III errors are calculated and tabulated. The overall misclassification rate is also calculated from these errors. For PNN, the overall misclassification rate is less for dataset II than dataset I. This is due to lesser types I, II and III error of the PNN compared to the SVM classification method. Thus depicts that PNN method classifies better than SVM. The possible reason for better performance of PNN could be due deep architectures with hidden layers which represent intelligent behavior more efficiently than shallow architectures like SVMs. However, the reduced overall misclassification rate of RBFN shows better performance for data Datasets than PNN. The results show that RBFN is advantageous to use for classification compared to PNN. It is also evident from the results that neural networks based methods dominates SVM classifier.

The results obtained for positive, negative and neutral precision are shown in Table 5 for the datasets used. From the results, it is observed that the performance of neural networks based method is better than SVM in classifying the positive, negative and neutral reviews with high precision. Among the neural methods investigated, RBFN performs better in terms of positive precision (92.3% for dataset I & 93.8% for dataset II), negative precision (88% for dataset I & 89.3% for dataset II) and neutral precision (85.4% for dataset I & 86.5% for dataset II). It is also evident from Table 5 that negative and neutral precision had a lesser value for RBFN method than positive precision. The significant increase in positive precision of RBFN compared to other classifiers specifies that the RBFN classifies positive reviews with more probability when compared to the other classifiers used. Recall of the classification methods is given in Table 6. All the methods predict positive reviews more completely than negative and neutral reviews.

The results obtained for positive, negative and neutral recall is shown in Table 6 for the datasets used (dataset I and dataset II). From the results, it is found that, dataset II of RBFN is able to identify the maximum number of positive, negative and neutral reviews present compared to other methods used. It is also noted that the performance of neural networks based methods is highly significant than SVM in classifying the positive, negative and neutral reviews with high recall.

Table 6
Recall of classifiers.

	Dataset I (%)			Dataset II (%)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
SVM	80.5	79.7	80.2	82.5	81.2	81.8
PNN	85.5	84.8	86.8	88.3	86.2	87.3
RBFN	89.7	87.7	87.8	90.8	89.0	89.5

Table 7
f-Score of classifiers.

	Dataset I (%)			Dataset II (%)		
	Positive	Negative	Neutral	Positive	Negative	Neutral
SVM	79.7	78.8	81.8	80.8	81.0	83.7
PNN	85.3	86.7	85.2	88.7	87.3	85.8
RBFN	92.5	88.0	84.7	94.0	89.3	86.0

Among the neural methods investigated, RBFN performs better in terms of positive recall (89.7% for dataset I & 90.8% for dataset II), negative recall (7.78% for dataset I & 89.0% for dataset II) and neutral recall (87.8% for dataset I & 89.5% for dataset II). It is also evident from Table 6 that negative and neutral recall had a lesser value for RBFN method than positive recall. This specifies that the RBFN classifies positive reviews with more completeness when compared to the other review classes. The results obtained for positive, negative and neutral *f*-score is shown in Table 7 for the datasets used (dataset I and dataset II). Since *f*-score is a combined measure of precision and recall, similar observation of precision and recall is noted. From the results, it is found that, dataset II of RBFN is having higher *f*-score for positive, negative and neutral reviews present compared to other methods used. It is also noted that the performance of neural networks based methods is highly significant than SVM in terms of *f*-score. Dataset II of RBFN gets 94% *f*-score, it was due to the 549 actual positive review documents classified as the positive reviews.

7. Discussion

From the results it is found that among the three methods, RBFN based prediction models perform well in all aspects. RBFN serve as an excellent role to make the learning process because they are universal approximations. Among the two datasets, dataset II based classification performance is high almost. RBFN performs better than PNN for all review classes in both datasets used. This is due to the fact that PNN itself has certain shortcomings. PNN tend to get trapped in undesirable local minima in order to reach the global minimum of a very complex search space. Also PNN methods fail to converge when high nonlinearities exist. Thus, these drawbacks deteriorate the

Table 8
Comparison with previous works.

S. No.	Author	Domain (review website)	Methods	Performance
1	Sharif et al. (2014)	Askapatient.com	Ensemble approach	Accuracy = 78.2%, precision = 78.4, recall = 79.8%
2	Asghar et al. (2014)	Askapatient.com	Incremental model	Accuracy = 78%, precision = 56%, recall = 66%
3	Yazdavar, Ebrahimi, and Salim (2016)	Askapatient.com	Fuzzy methods	Precision = 81%, recall = 64%, *f*-score = 72%
4	Goeuriot et al. (2012)	Askapatient.com	Lexicon based	Precision = 76%, recall = 52%, *f*-score = 62%
5	Our approach	Askapatient.com	PNN	Precision = 88.6%, recall = 88.3%, *f*-score = 88.7%
			RBFN	Precision = 93.8%, recall = 90.8%, *f*-score = 94.0%

accuracy of PNN in function approximation. RBFN overcomes these obstacles by replacing the global activation function in PNN with a localized radial basis function to perform better in function approximation. From the results in Table 5, it is found that the SVM lead to low negative precision, which implies that a large number of documents that are not actually negative would have been inspected. In general among the two datasets in each method, dataset II classification results are good. This indicates that the performance of the classifier depends on the nature of dataset (content of the document) irrespective of the size of the dataset. Among the methods, RBFN classify the reviews very accurately. Our results are comparative to other related studies for opinion classification of drug reviews (Table 8).

Our employed approach shows a better result in terms of various performance measures compared to other existing works on drug reviews collected from review website (askapatient.com). However, there is some room for improving the performance of opinion mining. The results could be increased further by eliminating the major cause for errors. From the analysis, it is found that errors were mainly caused due to misspelling words which caused some entities not to be recognized by tagger and errors in dependency and parse tree which led to errors in relation extraction between entities.

Although this investigation is based on a large review samples, there are a number of threats to its validity. This work is carried out with a balanced dataset and results of datasets may vary if class distribution is unbalanced. Though most of the opinion mining work on drug reviews has been carried out using reviews collected from Askapatient.com, few other review sites are also available in different format.

8. Conclusion

This paper presents function approximation by using ANNs based methods. RBFN perform better in function approximation and yields higher accuracy in prediction. In general, neural network based approaches perform better than the statistical based approach. Among them, RBFN was highly robust in nature which was studied through the various quality parameters. Our experimental analysis shows that RBFN could be a possible solution for increasing the classification performance. However, in domains such as medical, it is very common for users to express their opinions indirectly. In the drug domain, patients usually write about their experiences of drug effectiveness or side effects instead of expressing a direct opinion. So work need to be improved, to identify the implicit opinion present. To test the limitations of the proposed method, future works could use different datasets obtained from various forms of social web. Handling imbalanced data distribution need to analyzed in future. We believe this work is paving the way for a better understanding of patient reviews and opens interesting future research directions.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Adeli, H., & Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, *22*(7), 1018–1024.

Ali, T., Sokolova, M., Schramm, D., & Inkpen, D. (2013). Opinion learning from medical forums. In *RANLP* (pp. 18–24).

Asghar, M. Z., Khan, A., Kundi, F. M., Qasim, M., Khan, F., Ullah, R., et al. (2014). Medical opinion lexicon: An incremental model for mining health reviews. *International Journal of Academic Research*, *6*(1), 295–302.

Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Multi-facet rating of product reviews. In *European conference on information retrieval* (pp. 461–472). Berlin Heidelberg: Springer.

Bobicev, V., Sokolova, M., Jafer, Y., & Schramm, D. (2012, May). Learning sentiments from tweets with personal health information. In *Canadian conference on artificial intelligence* (pp. 37–48). Berlin, Heidelberg: Springer.

Briand, L. C., Wüst, J., Daly, J. W., & Porter, D. V. (2000). Exploring the relationships between design measures and software quality in object-oriented systems. *Journal of Systems and Software*, *51*(3), 245–273.

Chen, L. S., Liu, C. H., & Chiu, H. J. (2011). A neural network based approach for sentiment classification in the blogosphere. *Journal of Informetrics*, *5*(2), 313–322.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, *40*(16), 6266–6282.

Goeuriot, L., Na, J. C., Min Kyaing, W. Y., Khoo, C., Chang, Y. K., Theng, Y. L., et al. (2012). Sentiment lexicons for health-related opinion mining. In *Proceedings of the 2nd ACM SIGHIT international health informatics symposium* (pp. 219–226). ACM.

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2012). Machine learning and sentiment analysis of unstructured free-text information about patient experience online. *Lancet*, *380*, S10.

Hajmeer, M., & Basheer, I. (2002). A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. *Journal of Microbiological Methods*, *51*(2), 217–226.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177). ACM.

Jian, Z. H. U., Chen, X. U., & Wang, H. S. (2010). Sentiment classification using the theory of ANNs. *The Journal of China Universities of Posts and Telecommunications*, *17*, 58–62.

Luong, T., Socher, R., & Manning, C. D. (2013). Better word representations with recursive neural networks for morphology. In *CoNLL* (pp. 104–113).

Miao, Q., Li, Q., & Dai, R. (2009). AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications*, *36*(3), 7192–7198.

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621–633.

Na, J. C., Kyaing, W., Khoo, C., Foo, S., Chang, Y. K., & Theng, Y. L. (2012). Sentiment classification of drug reviews using a rule-based linguistic approach. *The Outreach of Digital Libraries: A Globalized Resource Network*, 189–198.

Noferesti, S., & Shamsfard, M. (2015). Resource construction and evaluation for indirect opinion mining of drug reviews. *PLoS One*, *10*(5), e0124993.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing, Vol. 10* (pp. 79–86). Association for Computational Linguistics.

Savchenko, A. V. (2013). Probabilistic neural network with homogeneity testing in recognition of discrete patterns set. *Neural Networks*, *46*, 227–241.

Sharif, H., Zaffar, F., Abbasi, A., & Zimbra, D. (2014). *Detecting adverse drug reactions using a sentiment classification framework*.

Sharma, A., & Dey, S. (2012). A document-level sentiment analysis approach using artificial neural network and sentiment lexicons. *ACM SIGAPP Applied Computing Review*, *12*(4), 67–75.

Socher, R., Lin, C. C., Manning, C., & Ng, A. Y. (2011). Parsing natural scenes and natural language with recursive neural networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 129–136).

Sokolova, M., & Bobicev, V. (2011). Sentiments and opinions in health-related web messages. In *RANLP* (pp. 132–139).

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applications*, *34*(4), 2622–2629.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, *36*(7), 10760–10773.

Vinodhini, G., & Chandrasekaran, R. M. (2014). Opinion mining using principal component analysis based ensemble model for e-commerce application. *CSI Transactions on ICT*, *2*(3), 169–179.

Wang, W., Chen, L., Tan, M., Wang, S., & Sheth, A. P. (2012). Discovering fine-grained sentiment in suicide notes. *Biomedical Informatics Insights*, *5*(Suppl. 1), 137.

Wiley, M. T., Jin, C., Hristidis, V., & Esterling, K. M. (2014). Pharmaceutical drugs chatter on online social networks. *Journal of Biomedical Informatics*, *49*, 245–254.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347–354). Association for Computational Linguistics.

Xia, L., Gentile, A. L., Munro, J., & Iria, J. (2009). Improving patient opinion mining through multi-step classification. In *International Conference on Text, Speech and Dialogue* (pp. 70–76). Berlin, Heidelberg: Springer.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, *181*(6), 1138–1152.

Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision Support Systems*, *50*(4), 743–754.

Yalamanchi, D. (2011). *Sideffective-system to mine patient reviews: Sentiment analysis* Doctoral dissertation. Rutgers University-Graduate School-New Brunswick.

Yazdavar, A. H., Ebrahimi, M., & Salim, N. (2016). Fuzzy based implicit sentiment analysis on quantitative sentences. *Journal of Soft Computing and Decision Support Systems*, *3*(4), 7–18.

Ye, Q., Lin, B., & Li, Y. J. (2005). Sentiment classification for Chinese reviews: A comparison between SVM and semantic approaches. In *International conference on machine learning and cybernetics, ICMLC 2005* (pp. 2341–2346). IEEE.

Ye, Q., Zhang, Z. Q., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert System with Application*, *36*(3), 6527–6535.

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the conference on empirical methods in natural language processing*.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, *38*(6), 7674–7682.