MINING ONLINE DRUG REVIEWS DATABASE FOR THE TREATMENT OF RHEUMATOID ARTHRITIS BY USING DEEP LEARNING

Pinar Yildirim

Department of Computer Engineering,
Faculty of Engineering and Natural Sciences,
Istanbul Okan University, Istanbul, Turkey

ABSTRACT

In this paper, a research study for online patient reviews is introduced. Rheumatoid arthritis is a long-term and disabling autoimmune disease. Today, a huge amount of people have rheumatoid arthritis in the world. Considering the importance of the medication of rheumatoid arthritis, we aimed to investigate patient reviews in WebMD database and get some useful information for this disease. Our results revealed that etanercept treatment has the highest number of reviews. Data analysis was applied to discover knowledge on this drug. Deep learning approach was used to predict the effectiveness of etanercept and classification results were compared with other traditional classifiers. According to the comparison of classifiers, deep neural network has better accuracy metrics than others. Therefore, the results highlight that deep learning can be encouraging for medical data analyses. We hope that our study can make contributions to intelligent data analysis in medical domain.

.

KEYWORDS

Classification, Deep Learning, Etanercept, Online Drug Reviews.

1. INTRODUCTION

Digital technologies provide many opportunities for healthcare treatment and research [1]. Thanks to these technologies, patients can communicate with other patients and type reviews about their medications in some social media sites. These reviews are important for both healthcare experts and drug companies who goal to follow the results of medications and increase the efficiency of them. In this study, WebMD medical website was used and an analysis was conducted using the patients' reviews on the medication of rheumatoid arthritis (RA). RA is a long-term, growing, and disabling autoimmune disease [2]. Considering the importance of the treatment of RA, we explored WebMD reviews and aimed to get some useful information for this disease [3]. Recently, the introduction of deep learning techniques is a promising trend in intelligent data analysis. Still, there are few studies based on these techniques for online patient reviews. We used these approaches in our study and targeted to make contributions to intelligent data analysis in medical domain.

2. RELATED WORKS

Several studies related to patient reviews exist in the literature. Bordes et al. investigated the patient with RA acceptance of social networking sites for the self management of disease. They performed a qualitative study by using interviews in patients. They found that the patients usually use Internet for health information but have limited aspect of social networks for their disease

David C. Wyld et al. (Eds): ARIA, SIPR, SOFEA, CSEN, DSML, NLP, EDTECH, NCWC - 2022
pp. 105-113, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121509

management. This reveals that an Internet based tool is needed to help the management of RA [4].

Kanzaki et al., created an Internet based system to gather data. Women patients with RA participated in this study and used this management system and communicated with the researchers. Their study showed that the use of web based system can positively affect on symptom management of RA [5].

Ellis et al., investigated arthritis patients' health literacy through their social network. They developed qualitative study based on semi-structured interviews. According to their results, patients have limited literacy capabilities and little information about their medications. Further, this study shows the patients with higher education level are more apt to health information searching behaviour [6].

3. METHODS

3.1. Data Sources

Data was gathered from the patient reviews for the treatment of RA on the WebMD website. This network provides web based system for patients to share their reviews of medication. In WebMD system, patients can rate effectiveness, ease of use, drug satisfaction from 1 to 5 stars and choose why they use the drug. Patients can also enter their age, gender, medication duration data and free text comments (Figure 1).

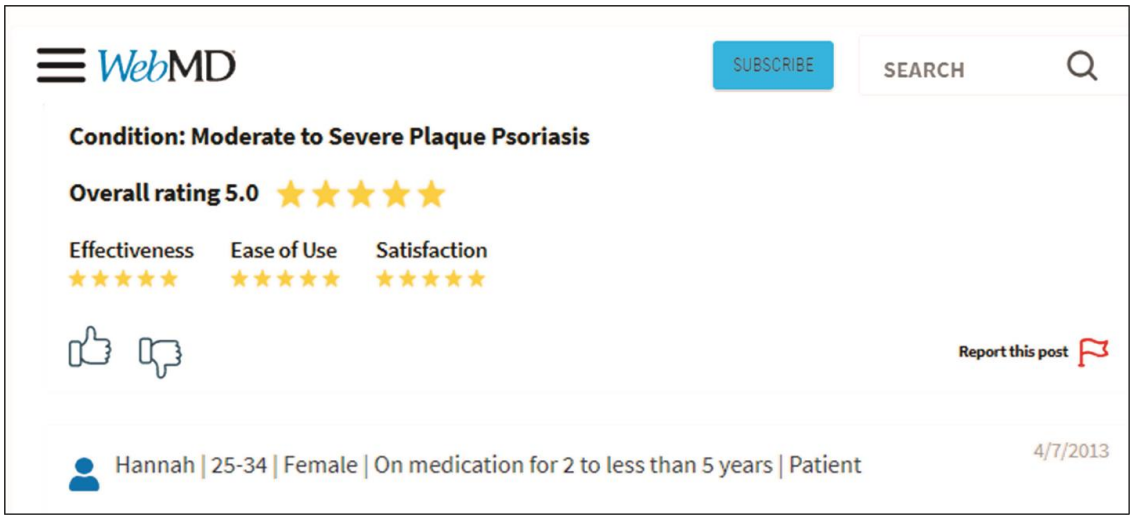

Figure 1. An example of patient review on WebMD website.

3.2. Deep Learning

Deep learning is a neural network approach which has many concealed layers in the network. The network operates huge amount of data through multiple layers and it can easily learn complex features at each layer (Figure 2). Thanks to this feature, deep neural network can handle and analyze many different types of data and decrease the drawbacks of traditional machine learning algorithms [7]. The main idea of deep learning is to search a function that generates the expected output for given inputs. Deep Learning requires intensive computational complexity and graphics processing units (GPUs) are important for the performance of network.

There are several deep learning models developed for different tasks. A simple Deep Neural Network (DNN) model is the auto encoder (AE) that consists of encoder and decoder functions for input and output layers. Convolutional neural network (CNN) has an interleaved set of feed forward layers containing convolutional filters, reduction, rectification or pooling layers. For each layer the CNN generates high-level abstract feature. CNN have been broadly applied image recognition and natural language processing. Another deep learning model is recurrent neural network (RNN). RNN exhibits dynamic structure and neurons in the network are related to time steps. Therefore, RNNs can easily handle sequential data. Deep belief network (DBN) is another type of deep neural network which consists of multiple layer of graphical model having both directed and undirected edges. A Deep Boltzmann Machine is a type of a Deep Neural Network formed from multiple layers of neurons with nonlinear activation functions. The architecture of a Deep Boltzmann Machine allows it to come to know very complicated relationships between attributes and provides advanced performance in learning of high-level representation of attributes [8]. Some types of deep learning techniques have been used in biomedical area such as biomedical imaging, medical diagnoisis, electronic health records and biomedical singals [9]. Some concepts are explained below;

Activation function

An activation function decides the output of each node in an artificial neural network basically and it can be called a transfer function that is used to map the output of one layer to another. Activation functions are important components of artificial neural networks (ANN), they affect on the performance of network. There are several activation functions such as sigmoid, hyperbolic tangent, softmax, rectified linear unit (ReLU) and softplus used in neural networks [9].

Learning rate

The learning rate is a hyper parameter that checks how much to vary the network in reply to the expected error each time the weights are changed. Small learning rate can cause a long training, whereas large rates may result in learning a sub-optimal set of weights too fast. The learning rate may be a significant hyper parameter for the architecture of neural network and it affect on the performance of the network [10].

Epoch

Epochs can be defined as how many iterations of the data the network will be used to train a model [11].

Loss function

Loss functions are used to measure how accurate the prediction is performed. If the prediction is obtained far away from true value i.e. prediction deviates more from real value, then the loss function generates high numeric value. In order to get good prediction, it must have low loss function values.

Regularization

In neural network studies, sometimes training data cannot be enough and the network model can face overfitting and under fitting problems. To handle these problems, some regularization techniques have been generally used for data analysis [12].

Batch size

Batch size *refers* the number of training examples in one iteration in the neural network.

Optimization

A learning task can be defined as an optimization problem, to find the minima of the objective function by choosing hyper parameters [9]. Stochastic gradient descent approach repeatedly makes small regulations to neural network configuration to reduce the error of the network. In deep learning networks, this method and its variants are widely used to achive optimization [13].

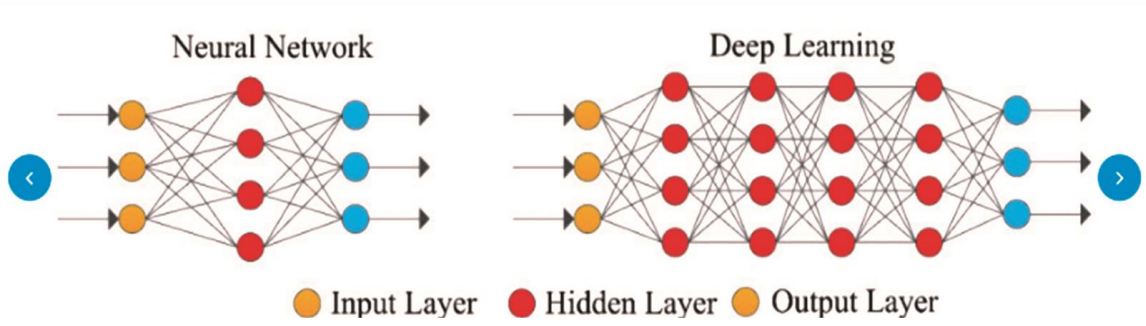

Figure 2. Architecture of Deep Learning [14].

3.3. Traditional Classification Algorithms

IBK

K-nearest neighbour algorithm is knowned as IBK in Weka software. When a new sample is given, a k-nearest neighbour classifier investigates the patterns for the training instances that are nearest to the unknown sample. The new sample is classified by its k nearest neighbours [15].

RandomTree

Random Tree is a kind of classifier; it is a type of ensemble learning algorithm that produces many individual learners. It uses a bagging idea to generate a random set of data for building a decision tree. In standard tree each node is split using the best split among all features [15,16, 17].

Random Forest

A Random Forest is a kind of ensemble of classification trees, where each tree makes contributions with a vote for the assignment of the most frequent class to the input data [18].

Naive Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problem. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object [19].

Kstar

Kstar is an instance-based classifier, that is the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function. It differs from other instance-based learners in that it uses an entropy-based distance function [20].

Logistic Regression

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables [21].

4. EXPERIMENTAL RESULTS

4.1. Data Analysis

We processed RA related patient reviews from WebMD website [8]. WebMD contains both structured data and free text comments. We collected patient reviews and processed these data. After data preprocessing, we converted WebMD data into structured form and created a MySQL database. According to analysis, etanercept has the highest number of reviews (248), we selected this drug for analysis [22-23]. Table 1 shows etanercept related attributes used for classification. These attributes are user's gender, user's age group, the time on the drug, user rating of ease of use, user rating of several satisfaction and user rating of effectiveness.

4.2. Classification Results

We selected six attributes of etanercept dataset (Table 1) to predict user rating of drug effectiveness. We implemented DNN for classification. The network was designed as dense layer (35 nodes) and output layer 6 nodes). We used some DNN parameters and these parameters were kept fixed in all experiments performed in this study:

(a) ActivationSoftmax was selected as a activation function,

$$g(a) = \frac{e^{a_i}}{\sum_j e^{a_j}}$$

The softmax output, which a_n be considered as a probability distribution over the categories, is commonly used in the final layer [9].

(b) Stochastic Gradient Descent was used as an optimization method which an iterative method for optimizing an objective function with suitable smoothness properties [24].

(c) LossMCXENT, a type Multi-Class Cross Entropy loss function, was selected as a loss function

The performance of the DNN model was affected by hyper parameters which results in optimal classification results. We tried to get optimal values for the DNN model by changing the numbers of epoch and mini batch size. The same classification experiment was run in increasing steps by varying the number of epoch between 10 and 100 and the value of mini batch size between 1 and 10 (Table 4-5). WekaDeeplearning4j software was used. The software is a deep learning package

in Deeplearning4j. This software ensures a graphical user interface (GUI) for deep learning applications [25].

Table 1. Attributes used for classification

Attributes	Data Type
User's gender	**Categorical** Male,Female
User's age group	**Categorical** 3-6, 7-12, 13-18,19-24, 25-34, 35-44, 45-54,55-64,65-74,75 or over
User rating of ease of use	**Categorical** Rating: 1,2,3,4,5
User rating of several satisfaction	**Categorical** Rating: 1,2,3,4,5
The time on the drug	**Categorical** Less than 1 month 1 to 6 months 6 months to less than 1 year 1 to less than 2 years 2 to less than 5 years 5 to less than 10 years 10 years or more
User rating of effectiveness	**Categorical** Rating: 1,2,3,4,5

Performance of classification algorithms are evaluated by some accuracy measures.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - measure = \frac{2TP}{2TP + FN + FP}$$

TP: The number of True Positives
TN: The number of Negatives instances
FP: The number of False Positives
FN: The number of Negatives instances

The evaluation analysis by root mean squared error is also widely used where n is the number of data, $y_{p,m}$ shows the predicted, $t_{m,m}$ is the measured value of one data point m and $\bar{t}_{m,m}$ is the mean value of all measure data values. Root Mean Squared Error (RMSE) can be shown as follows [26]:

$$\text{RMSE} = \sqrt{\frac{\sum_{m=1}^{n} (y_{p,m} - t_{m,m})^2}{n}}$$

Table 2. Performance evaluation for different epoch values

Epoch	Precision	Recall	F-measure	RMSE
10	0.700	0.677	0.680	0.2884
20	**0.722**	**0.702**	**0.707**	0.276
30	0.704	0.702	0.700	0.2735
40	0.701	0.710	0.703	0.2738
50	0.697	0.706	0.699	0.2746
60	0.684	0.694	0.687	0.2755
70	0.679	0.690	0.683	0.2763
80	0.672	0.681	0.675	0.2771
90	0.667	0.677	0.671	0.2777
100	0.667	0.677	0.671	0.2783

Table 3. Performance evaluation for different mini-batch sizes.

Mini-batch size	Precision	Recall	F-Measure	RMSE
1	**0.700**	**0.677**	**0.680**	**0.2884**
2	0.680	0.633	0.645	0.3059
3	0.650	0.585	0.606	0.3193
4	0.647	0.556	0.586	0.3298
5	0.636	0.524	0.563	0.3376
6	0.630	0.504	0.548	0.3433
7	0.610	0.472	0.521	0.3492
8	0.608	0.464	0.514	0.3534
9	0.573	0.395	0.452	0.3573
10	0.568	0.387	0.445	0.3596

Table 4. Performance evaluation for different classification algorithms

Classification Algorithm	Precision	Recall	F-measure	RMSE	Execution Time
Deep Neural Network	**0.700**	**0.677**	**0.680**	0.2884	5.04
Random Forest	0.668	0.673	0.670	0.2791	0
IBK	0.585	0.593	0.587	0.3128	0
Random Tree	0.615	0.605	0.607	0.3257	0
Kstar	0.582	0.625	0.592	0.2836	0

Table 2-3 show the performance evaluation for DNN with varying epoch values and mini-batch sizes. According to Table 2, the highest precision values were got for the dataset with epoch 20. For instance, the recall of DNN with epoch 20 is 0.702 in the Table 2. These results reveal that smaller epoch values may produce higher accuracy values. Similarly, small mini batch sizes generates good accuracy values for classification tasks. For example, mini batch size 1 has highest F-measure with 0.680 and smallest RMSE with 0.2884 in Table 3. However, there is no general decision for these results. Smallest epoch and mini batch values may not result in small RMSE values in all studies.

We also compared the performance of DNN with other traditional algorithms. Table 4 shows the accuracy metrics for different classification algorithms. According to results, DNN has better accuracy values than other algorithms. For example, the precision of DNN is 0.700 which is the highest value in the table. On the other hand, DNN has the longest execution time with 5.04.

5. CONCLUSIONS

The biomedical research aims to investigate unknown and useful knowledge to make contributions for healthcare. Drug satisfaction is one of the most important issue for the medical area. In this study, we carried out data analysis for the treatment of RA and mostly reviewed etanercept drug. We analyzed WebMD database and searched the patient satisfaction of etanercept. We implemented deep learning approach to predict the relationships between drug effectiveness and other features such as gender, age and the time on the drug. The performance of DNN was observed by epoch and mini batch size to find optimum parameters. A comparative experiment was also performed on classification algorithms to evaluate them. The results highlight that deep learning is promising technique and it can result in high accuracy of classification with optimum epoch and mini batch size parameters. In conclusion, our study can make contributions for both medical experts and data scientists.

ACKNOWLEDGEMENTS

We would like to thank Alkan Kaya for his help and contributions.

REFERENCES

[1] Yildirim P. Association patterns in open data to explore ciprofloxacin adverse events. Applied Clinical Informatics 2015, 6(4): 728-747.
[2] https://www.rheumatoidarthritis.org.
[3] Boshu R, Charles W-H, Yao L. Identifying serendipitous Drug Usages in Patient Forum Data-A Feasibility Study. BIOSTEC 2017, 2017, page 106-118.
[4] Bordes JKA. des, Gonzales E., Lopez-Olivo M., Nayak P, Suarez-Almazor M.E. Assessing information needs and use of online resources for disease self-management in patients with rheumatoid arthritis:a qualitative study, Clinical Rheumatology 2018; 37;1791-1797.
[5] Kanzaki H, Makimoto K, Takemura T, Ashida N. Development of web-based qualitative and quantitativee data collection systems: study on Daily symptoms and cop,ng strategies among Japanese rheumatoid arthritis patients, Nursing and Health Sciences 2004; 6, 229-236.
[6] Ellis J, Mullan J, Worsley A, Pai N. The Role of Health Literacy and Social Networks in Arthritis Patients' Health Information-Seeking Behavior: A Qualitative Study, International Journal of Family Medicine Volume 2012, Article ID 397039, 6 pages.
[7] Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L. Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations, JMIR MHealth and UHealth 2009, vol.7, iss.8, p1-36.
[8] Taherkhani A, Cosma G, Mcginnity T M, Deep-FS. A feature selection algorithm for deep Boltzmann Machines, December 2018, Neurocomputing Vol 322, page 22-37.
[9] Cao C, Liu F, Tan H, Song D, Shu w, Li W, Zhou Y, Bo X, Zie Z. Deep Learning and its applications in biomedicine, Genomics Proteomics Bioinformatics 2018, 16, 17-32.
[10] https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/ [20 December 2021].
[11] Maxwell A, Li R, Yang B, Weng H, Ou A, Hong H, Zhou Z, Gong P, Zhang C, Deep learning architectures for multi-label classification of intelligent health risk prediction, BMC Bioinformatics, 18(Suppl 14);523, 2017
[12] Nusrat I, Jang SB. Comparison of Regularization Techniques in Deep Neural Networks, Symmetry 2018, 10, 648.

[13] Koutsoukas A, Monaghan KJ, Li X, Huan J, Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data, J Cheminform,9:42, 2017.
[14] https://www.researchgate.net/figure/Deep-learning-diagram_fig5_323784695.
[15] Han J Kamber M. Data Mining Concepts and Techniques.Morgan Kaufmann, 2011 .
[16] Kalmegh SR. Comparative Analysis of WEKA Data Mining Algorithm RandomForest, RandomTree and LADTree for Classification of Indigenous News Data. Int. Journal of Emerging Technology and Advanced Engineering 2015; 5;1. 13.
[17] Pfahringer B. Random model trees: an elective and scalable regression method. Working Paper Series, 2010, ISSN 1177-777X.
[18] Breiman L, Friedman J, Stone CJ, Olshen RA, Classification and regression trees (1st ed.), Chapman and Hall/CRC, Belmont, CA (1984).
[19] https://www.javatpoint.com/machine-learning-naive-bayes-classifier
[20] https://weka.sourceforge.io/doc.dev/weka/classifiers/lazy/KStar.html
[21] https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-logistic-regression
[22] Rastegar-Mojarad, Majid, Liu,Hongfang, Nambisan, Priya. (2016) "Using Social Media Data to Identify Potential Candidates for Drug Repurposing: A Feasibility Study." JMIR RESEARCH PROTOCOLS ,vol. 5, iss. 2, e121 p.
[23] http://www.drugs.com
[24] http://www.wikipedia.org
[25] Lang S, Bravo-Marquez, Beckham C, Hall M, Frank E, WekaDeeplearning4j: A deep learning package for Weka based on Deeplearning4j, Knowledge-Based Systems, Volume 178, 15 August 2019, Pages 48-50).
[26] Kuçuksille EU, Selbas R, Şencan A. Prediction of thermodynamic properties of refrigerants using data mining. Energy conversion and management 2011; 52: 836-848.

AUTHORS

Pinar Yildirim is an associated professor at the Department of Computer Engineering of Istanbul Okan University in Istanbul.

She received her B.S. degree Yıldız Technical University, M.S. degree from Akdeniz University and PhD degree in the Department of Health Informatics of the Informatics Institute at Middle East Technical University in Turkey.Her research areas include biomedical data mining,.machine learning,.