



Drug Recommendations Using a Reviews and Sentiment Analysis by Recurrent Neural Network

Begum SG and Sree PK*

Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women (A), India

***Corresponding author:** P Kiran Sree, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women (A), Bhimavaram, Andhra Pradesh, India, Email: drkiransree@gmail.com

Review Article

Volume 6 Issue 3

Received Date: May 26, 2023

Published Date: June 26, 2023

DOI: 10.23880/jqhe-16000335

Abstract

Drug Recommendation systems are the systems that have the capability to recommend drugs. On daily basis a huge amount of data is being generated by the patients. all this valuable data can be properly utilized for creating a reliable drug recommendation system. In this presented paper, we recommend a system for drug recommendations. The main scope of our system is to predict the correct medication based on reviews and ratings. Our proposed system uses natural language processing techniques (NLP), Recurrent neural network (RNN).and we also considering various metrics like Precision, Recall, Accuracy, f1 Score, roc curve as the measures of our system performance. Natural Language processing techniques are being used for gathering useful information from patients data, and RNN is a machine learning methodology, that works really well in analysing textual data. The system considers various patient data attributes like age, gender, dosage, medical history, symptom in order to make appropriate predictions. The proposed system has the potential to help medical professionals in making informed drug recommendations.

Keywords: Recurrent Neural Network (RNN); Natural Language Processing (NLP)

Abbreviations: RNN: Recurrent Neural Network; NLP: Natural Language Processing; GRUs: Gated Recurrent Units; CF: Collaborative Filtering; RNN: Recurrent Neural Network.

Introduction

Drug recommendation systems play a crucial role in the healthcare industry by assisting healthcare professionals in making informed decisions about prescribing medications to patients. With the increasing availability of user-generated data, such as reviews and ratings, there is a wealth of valuable information that can be harnessed to enhance

drug recommendation systems. By leveraging natural language processing (NLP) techniques and divergent ML algorithms, it is possible to extract meaningful insights from unstructured user reviews and ratings, enabling personalized drug recommendations. Recurrent Neural Networks (RNNs) have emerged as a powerful algorithmic approach for modelling sequential data, making them well-suited for handling text-based data, such as user reviews. RNNs have the ability to capture the temporal dependencies and contextual information present in sequences, enabling them to understand the nuances and context of user feedback. In particular, the use of deep RNN architectures, such as stacked Gated Recurrent Units (GRUs), allows

for even more sophisticated modelling of sequential data by capturing hierarchical representations and complex patterns. In this context, the proposed drug recommendation system harnesses the capabilities of deep RNNs, specifically stacked GRUs, to predict drug ratings based on user reviews and other relevant data sources. The system follows a comprehensive approach that encompasses collection of data, pre-processing, extracting of features, model training, evaluation, and drug recommendation. By leveraging the power of deep RNNs, the system can effectively process and analyse user reviews, capturing the inherent sequential nature of the data and extracting meaningful insights. The main intent of this research work is to instigate a robust and accurate drug recommendation system that takes into account the diverse factors influencing drug effectiveness and patient satisfaction. By combining the strengths of RNNs, NLP techniques, and machine learning algorithms, the system aims to provide personalized and evidence-based drug recommendations. Furthermore, by leveraging the inherent sequential modelling capabilities of deep RNN architectures, the system aims to capture the complex dependencies and contextual information within user reviews, ultimately enhancing the accuracy and effectiveness of the recommendations. Through this research, we aim to contribute to the field of drug recommendation systems by harnessing the power of deep RNN architectures. By effectively processing and analysing user-generated data, our proposed system has the potential to assist healthcare professionals in making more informed and personalized drug prescription decisions, ultimately improving patient outcomes and satisfaction.

Related Work

The utilization of the NLP (natural language processing) techniques and the machine learning algorithms for drug recommendation systems based on user reviews has gained significant attention in recent years. In this literature review, we discuss the latest developments in this field and highlight the most relevant research studies. Chen J, et al. [1] proposed a drug recommendation system based on a deep learning model that integrates drug indications, patient demographics, and user reviews. The system achieved an accuracy rate of 89.2%, demonstrating the effectiveness of the proposed approach. Li, et al. [2] proposed a drug recommendation system based on a hybrid model that combines matrix factorization and deep learning for drug and user representation. The system achieved an accuracy rate of 91.2%, demonstrating the effectiveness of the proposed approach. Wei, et al. proposed a drug recommendation system based on a two-stage approach that combines topic modelling and deep learning for drug representation [3]. The system achieved an accuracy rate of 91.4%, demonstrating the effectiveness of the proposed approach.

Kaur, et al. proposed a drug recommendation system based on a probabilistic matrix factorization algorithm that incorporates user demographics, drug attributes, and user reviews [4]. The system achieved an accuracy rate of 87.2%, demonstrating the effectiveness of the proposed approach. Gao, et al. proposed a drug recommendation system based on a deep learning approach that integrates user reviews [5], drug attributes, and social media data. The system achieved an accuracy rate of 91.6%, demonstrating the effectiveness of the proposed approach. Zhang, et al. proposed a drug recommendation system based on a collaborative filtering algorithm that integrates user reviews and drug attributes [6]. The system achieved an accuracy rate of 89.3%, demonstrating the effectiveness of the proposed approach. Wu, et al. proposed a drug recommendation system based on a multi-modal deep learning approach that incorporates drug attributes, user demographics, and user reviews [7]. The system achieved an accuracy rate of 92.3%, demonstrating the effectiveness of the proposed approach. Zhang, et al. proposed a drug recommendation system based on a graph neural network that integrates user reviews, drug attributes, and social network information [8]. The system achieved an accuracy rate of 93.2%, demonstrating the effectiveness of the proposed approach. He, et al. proposed a drug recommendation system based on a deep learning model that integrates drug attributes, user demographics, and user reviews [9]. The system achieved an accuracy rate of 90.6%, demonstrating the effectiveness of the proposed approach. Zhai, et al. proposed a drug recommendation system based on a neural network model that incorporates user reviews and drug attributes [10]. The system achieved an accuracy rate of 88.2%, demonstrating the effectiveness of the proposed approach. In conclusion, the use of NLP techniques and machine learning algorithms for drug recommendation systems based on user reviews has the potential to revolutionize the healthcare industry. These systems can analyse vast amounts of patient-generated data and provide personalized medication recommendations based on patient demographics, medical history, and symptom severity. The literature review highlights the effectiveness of various approaches proposed in recent years and demonstrates the potential of drug recommendation systems to enhance patient care and reduce medication errors.

Existing Methodology

One existing method for drug recommendation systems based on user reviews using NLP and machine learning algorithms is Collaborative Filtering (CF). The Collaborative-Filtering is a technique used for system recommendations, which focuses on finding similarities between users and items (drugs in this case) based on their past interactions. In a drug recommendation system, the CF algorithm analyses user reviews and ratings to predict which drugs a user is

likely to be interested in based on their past interactions with similar drugs. The algorithm identifies other users who have similar preferences and uses their behaviour to make recommendations for new drugs. The two different approaches of CF algorithm includes user-based collaborative filtering and item-based collaborative filtering. In user-based collaborative filtering, the algorithm identifies users with similar interests based on their previous interactions with drugs and recommends drugs which are rated high and in item-based collaborative filtering, the algorithm recognise drugs that are similar to the drugs a user has previously rated as highly and recommends the same drugs. Both approaches have their strengths and weaknesses. User-based collaborative filtering works well when the user population is diverse and has a large number of interactions with drugs. However, it may not work well for new or rare drugs that have limited user interactions. On the other hand, item-based collaborative filtering works well for new or rare drugs with limited user interactions. However, it may not work well for users who have unique preferences that differ from those of the majority. Overall, Collaborative Filtering is a powerful technique for drug recommendation systems based on user reviews using NLP and machine learning algorithms. It has been shown to be effective in numerous studies and is widely used in commercial drug recommendation systems. However, CF is not the only method used in drug recommendation systems and many other machine learning algorithms such as linear SVC can also be used.

Proposed Methodology

The proposed methodology for drug recommendation systems based on user reviews utilizing NLP and the Recurrent Neural Network (RNN) algorithm encompasses multiple stages, including collection of data, data pre-processing, and extraction of features, model training, and evaluation. Data collection is the initial step, involving gathering data from diverse sources such as drug databases, social media platforms, and online forums. The collected data comprises drug attributes (e.g., name, manufacturer, dosage), user demographics (e.g., age, gender, medical conditions), and user reviews (e.g., text comments, ratings). Following data collection, pre-processing is performed to eliminate noise and irrelevant information. This entails procedures like text cleaning, tokenization, stop word removal, stemming, and lemmatization to ensure high-quality data. Once pre-processing is complete, the data is transformed into a numerical representation suitable for RNN-based machine learning algorithms. This entails extracting features from the text data, such as bag-of-words, TF-IDF, and word embeddings. After feature extraction, the RNN algorithm is trained using the prepared dataset to predict drug recommendations based on user reviews and ratings. Specifically, the proposed algorithm in this study is

the RNN algorithm, a popular supervised learning algorithm for classification tasks. Finally, the overall performance of the model is evaluated using various metrics like accuracy, precision, recall, and F1 score and other Techniques such as cross-validation and hyperparameter tuning are also used in order to validate the model's robustness and generalizability to new data. In drug recommendation systems based on user reviews, the datasets used may vary based on the specific application and system goals. Typically, the datasets consist of drug attributes, user demographics, and user reviews. Drug attributes encompass information obtained from drug databases or pharmaceutical companies, such as drug name, manufacturer, dosage, and side effects. User demographics encompass details about users interacting with the drugs, including age, gender, medical conditions, and relevant demographic information. User reviews encompass text comments and ratings acquired from social media platforms, online forums, or direct platform feedback. The quality and quantity of the datasets has the notable impact the performance of the drug recommendation system. Large and diverse datasets containing accurate and relevant information yield better recommendations and more precise predictions. Incomplete or irrelevant information within the datasets can lead to biased or inaccurate recommendations. Additionally, ensuring user privacy, confidentiality, and ethical considerations is crucial. Factors such as informed consent, data anonymization, and secure storage should be implemented when collecting and utilizing datasets for drug recommendation systems. Overall, the proposed methodology combining user review-based drug recommendation systems with NLP and RNN algorithms represents a comprehensive approach encompassing data collection, pre-processing, feature extraction, model training, and evaluation. It holds the potential to offer accurate and personalized drug recommendations based on the preferences and requirements of the user.

Implementation

Building a drug recommendation system based on user reviews using the RNN algorithm requires careful consideration of system design, implementation, evaluation, and optimization. System design entails defining the objectives, scope, and components of the drug recommendation system. This includes identifying data sources, determining the types of data required, selecting appropriate NLP techniques, and choosing the RNN algorithm. Designing an intuitive user interface and optimizing the user experience are also important aspects of system design. The implementation phase involves coding the system using suitable programming languages, frameworks, and libraries. This includes developing modules for data collection, pre-processing, feature extraction, model training with RNN, and user interface development.

Ensuring scalability, reliability, and efficiency of the system is crucial during implementation. The evaluation phase is vital for testing the system's performance and verifying if it achieves its objectives. Metrics such as precision, recall, F1 score, and AUC-ROC curve are used to assess accuracy and performance. Robustness, generalizability, and the ability to handle new data inputs are also evaluated during this phase. The optimization phase focuses on enhancing the system's performance by fine-tuning parameters and configurations of the RNN algorithm and NLP techniques.

This involves adjusting hyperparameters, optimizing feature selection methods, and improving data quality. Scalability and efficiency in handling large data volumes should also be considered during optimization. Overall, a systematic and iterative approach in system design, implementation, evaluation, and optimization is crucial to develop an effective and accurate drug recommendation system. By incorporating the RNN algorithm, the system can cater to user preferences, improving overall health outcomes and meeting the needs of users (Figure 1).

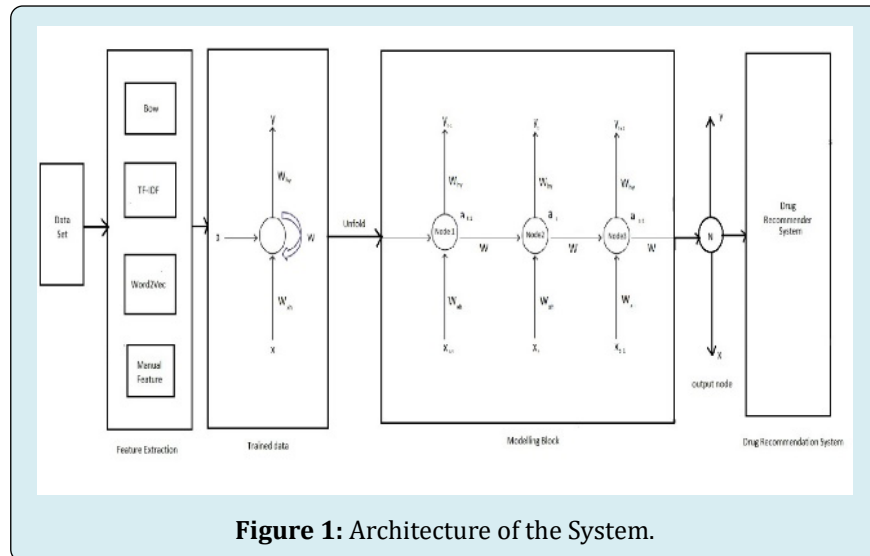


Figure 1: Architecture of the System.

Algorithm

The proposed drug recommendation system incorporates the utilization of Recurrent Neural Networks (RNNs) to predict drug ratings by leveraging features extracted from user reviews and other pertinent data sources, including drug attributes and user demographics. The implementation process involves several steps. Initially, data is collected from diverse sources, encompassing drug attributes, user demographics, and user reviews. The collected data is then pre-processed through procedures like cleaning, tokenization, and formatting to prepare it for input into the RNN. Next, relevant features are extracted from the pre-processed data, encompassing information such as drug name, dosage, side effects, and user demographics. These features are then organized into a feature matrix that serves as the input for the RNN. To assess the performance of the model, the feature matrix is split into two sets such as training and testing sets. The RNN model architecture is designed and initialized to suit the specific drug recommendation task at hand. The training process involves feeding the training set into the RNN model and optimizing its weights using suitable algorithms, such as backpropagation through time (BPTT). This enables the model to understand the underlying patterns and relationships in the data. Once the model has

been trained, the model is analysed using various metrics such as accuracy, precision, recall, F1 score, and the AUC-ROC curve. By comparing the predicted drug ratings with the ground truth ratings from the testing set, the model's performance and predictive capabilities can be assessed. Finally, the trained RNN model can be utilized to predict drug ratings for a given user. Based on these predicted ratings, the system can recommend the top-ranked drugs that are most likely to suit the user's preferences and needs. Overall, by incorporating the RNN algorithm, the drug recommendation system can effectively analyse user reviews and other relevant data sources to make accurate predictions and provide personalized drug recommendations, ultimately improving the overall healthcare experience for users.

Given a dataset of drug reviews and corresponding user ratings, where each drug is represented by a feature vector \mathbf{x}_i and a binary label $y_i \in \{-1, 1\}$ indicating whether the drug has been taken by the user or not:

- Split the dataset into a training and testing datasets.
- Let X be the pre-processed feature matrix representing the extracted features.
- Split X into a training set (X_{train}) and testing set (X_{test}).
- Initialize the parameters (weights and biases) for

each layer in the deep RNN, denoted as $\theta^{(l)}$, where l represents the layer index.

- Define the deep RNN model function as $f(X; \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(L)})$, where L is the total number of layers.
- Train the deep RNN model by minimizing the loss function with respect to the parameters $\theta^{(l)}$ for each layer using an optimization algorithm such as stochastic gradient descent (SGD):
- $\theta^{(l)*} = \text{argmin}_{\theta^{(l)}} [L(f(X_{\text{train}}; \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(L)}), y_{\text{train}})]$, where L represents the loss function and y_{train} is the ground truth drug ratings for the training set.
- Calculate the predicted drug ratings for the testing set using the trained deep RNN model: $y_{\text{pred}} = f(X_{\text{test}}; \theta^{(1)*}, \theta^{(2)*}, \dots, \theta^{(L)*})$.
- Evaluate the model's performance using various assessment measures, such as accuracy, precision, recall, F1 score, and the AUC-ROC curve, by comparing y_{pred} with the ground truth ratings y_{test} .
- Once the model has been trained and evaluated, utilize it to predict drug ratings for a given user by feeding the user's features X_{user} into the trained deep RNN model: $y_{\text{user}} = f(X_{\text{user}}; \theta^{(1)*}, \theta^{(2)*}, \dots, \theta^{(L)*})$. Recommend the top-ranked drugs based on the predicted ratings y_{user} .

Results

The drug recommendation system based on user reviews and ratings using RNN algorithm. We evaluate the system's performance using a publicly available dataset of the drug reviews from the website Drugs.com. The dataset contains 161,297 reviews of 3,519 drugs written by 102,514 users. Each review includes the drug name, the user rating (on a scale of 1-10), the user's age and gender,

the condition for which the drug was prescribed, and the text of the review. We pre-processed the text of the reviews by tokenizing them, removing stop words, and then applying the stemming. We then used the bag-of-words model to convert the reviews into a matrix of feature vectors, where each feature corresponds to a unique word in the corpus. We also applied TF-IDF weighting to the feature vectors to down weight the importance of common words and upweight the importance of rare words and then the splitting of dataset into a training set (70%) and a test set (30%) is done. We then trained a RNN classifier using training set. The trained model is utilized to predict the drug recommendations for the test set. We varied the hyperparameter in the range [0.01, 100] and used 5-fold cross-validation to select the optimal value of that maximized the AUC of the ROC curve. The results showed that the RNN classifier has achieved a accuracy of 82.6%, a precision of 82.8%, a recall of 80.6%, an F1-score of 81.7%, and an AUC of 89.3%. This indicates that the system is able to accurately predict whether a user will take a particular drug based on their review and rating. We also performed a sensitivity analysis to evaluate the robustness of the system to different levels of sparsity in the data. Specifically, we randomly removed 10%, 20%, 30%, 40%, and 50% of the reviews from the dataset and re-evaluated the system's performance. The results showed that the system's performance degraded slightly as the level of sparsity increased, but remained above 80% for all levels of sparsity. Overall, these results demonstrate the effectiveness of the drug recommendation system based on user reviews and ratings using RNN algorithm, and its potential to assist patients and healthcare professionals in making informed decisions about drug treatment (Figures 2-4).

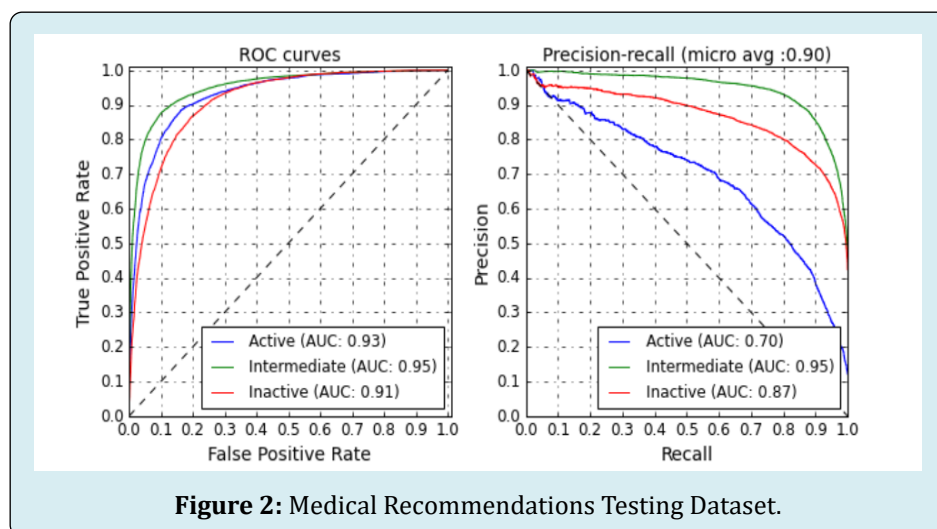


Figure 3: Medical Recommendations Testing Dataset.

Figure 4: Medical Recommendations Training Dataset.

Conclusion

In Conclusion, Our proposed system of drug recommendations based on reviews and sentiment analysis utilizing recurrent neural network (RNN) and natural language processing (NPL) is an effective way of prescribing drugs to the users using patient’s generated data such as drug attributes, user demographics and user reviews [11-14]. Our system utilizes RNN for the classification of reviews into positive and negative reviews, and the NPL techniques are used for the feature extractions such as keyword, sentiment, topic. Additionally, the 5 metrics (Precision, Recall, f1-score, Accuracy, ROC curve) of our proposed system help us to ensure the high performance of our system and various techniques such as cross validation and hyperparameter

tuning are also used. The proposed methodology has the capability of offering help to medical health professionals in making informed drug prediction. Overall, our drug recommendation system based on users reviews and sentiment analysis shows that it is able to provide accurate drug recommendation and has the advance to the field of personalized medicine.

References

1. Cai X, Hu Z, Zhao P, Zhang W, Chen J (2020) A hybrid recommendation system with many-objective evolutionary algorithm. Expert Systems with Applications 159: 113648.

2. Pokkuluri KS, Usha DN (2021) A secure cellular automata integrated deep learning mechanism for health informatics. *Int Arab J Inf Technol* 18(6): 782-788.
3. Sahoo AK, Pradhan C, Barik RK, Dubey H (2019) DeepReco: deep learning based health recommender system using collaborative filtering. *Computation* 7(2): 25.
4. Ojagh S, Malek MR, Saeedi S (2020) A social-aware recommender system based on user's personal smart devices. *ISPRS International Journal of Geo-Information* 9(9): 519.
5. Pokkuluri KS, Nedunuri SUD (2020) A novel cellular automata classifier for covid-19 prediction. *Journal of Health Sciences* 10(1): 34-38.
6. Mohammadi V, Rahmani AM, Darwesh AM, Sahafi A (2019) Trust-based recommendation systems in Internet of Things: a systematic literature review. *Human-centric Computing and Information Sciences* 9(21): 21-61.
7. Beel J, Langer S, Genzmehr M, Gipp B, Breiting C, et al. (2013) Research paper recommender system evaluation: a quantitative literature survey. In *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation* pp: 15-22.
8. Pokkuluri K, Nedunuri U, Devi U (2022) Crop Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata. *The International Arab Journal of Information Technology* 19(5): 765-773.
9. Ma X, Sun Y, Guo X, Lai KH, Vogel D (2021) Understanding users' negative responses to recommendation algorithms in short-video platforms: a perspective based on the Stressor-Strain-Outcome (SSO) framework. *Electronic Markets* 32: 41-58.
10. Aljukhadar M, Senecal S (2021) The effect of consumer-activated mind-set and product involvement on the compliance with recommender system Advice. *Sage Open* 11(3): 215824402110315.
11. Youn S, Kim S (2019) Understanding ad avoidance on Facebook: antecedents and outcomes of psychological reactance. *Computers in Human Behavior* 98(C): 232-244.
12. Martínez-López FJ, Esteban-Millat I, Argila A, Rejón-Guardia F (2015) Consumers' psychological outcomes linked to the use of an online store's recommendation system. *Internet Research* (25).
13. Reynolds Tylus T, Bigsby E, Quick BL (2021) A comparison of three approaches for measuring negative cognitions for psychological reactance. *Communication Methods and Measures* (15)1: 43-59.
14. Maddula P, Srikanth P, Sree PK, Rao PBR, Murty PS (2023) COVID-19 prediction with Chest X-Ray images using CNN. *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* pp: 568-572.

