BOHR

Drug recommendation using recurrent neural networks augmented with cellular automata

S. Gousiya Begum and Pokkuluri Kiran Sree[*]

Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India

[*]**Correspondence:**
Pokkuluri Kiran Sree,
drkiransree@gmail.com

Drug recommendation systems are systems that have the capability to recommend drugs. On a daily basis, a huge amount of data is being generated by the patients. All this valuable data can be properly utilized to create a reliable drug recommendation system. In this 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 networks (RNN), and cellular automata (CA). We also considered various metrics like precision, recall, accuracy, F1 score, and ROC curve as measures of our system's performance. NLP techniques are being used for gathering useful information from patient data, and RNN is a machine learning methodology that works really well in analyzing textual data. The system considers various patient data attributes like age, gender, dosage, medical history, and symptoms in order to make appropriate predictions. The proposed system has the potential to help medical professionals make informed drug recommendations.

Keywords: recurrent neural network (RNN), natural language processing (NLP), machine learning, cellular automata

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 machine learning (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 modeling 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 modeling 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 the collection of data, pre-processing, extraction of features, model training, evaluation, and drug recommendation. By leveraging the power of deep RNNs, the system can effectively process and analyze 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 ML algorithms, the system aims to provide personalized and evidence-based drug recommendations. Furthermore, by leveraging

the inherent sequential modeling 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 analyzing 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

In recent years, there has been a lot of interest in the application of NLP techniques and ML algorithms for drug recommendation systems based on user ratings. We analyze the most recent advancements in this area and highlight the most pertinent research projects in this overview of the literature. A deep learning model-based drug recommendation system that incorporates drug indications, patient demographics, and user evaluations was proposed by Chen et al. (1). The system's accuracy rate of 89.2% demonstrates the potency of the suggested strategy. Li et al. (2) suggested a hybrid model that combines matrix factorization and deep learning for drug and user representation in a drug recommendation system.

The system's accuracy rate of 91.4% proved the value of the suggested strategy. A probabilistic matrix factorization algorithm-based drug recommendation system that takes into account user demographics, drug properties, and user evaluations was proposed by Kaur and Singh (3). The system's accuracy rate of 87.2% showed how successful the suggested strategy is. A deep learning-based drug recommendation system that incorporates user evaluations, drug properties, and social media data was proposed by Gao et al. (4). The system's accuracy rate of 91.6% proved the value of the suggested strategy. A drug recommendation system based on a collaborative filtering algorithm that incorporates user reviews and drug properties was proposed by Zhang et al. (5). The system had an 89.3% accuracy rate.

The system's accuracy percentage of 93.2% proves that the suggested strategy is effective. He et al. (6) suggested a deep learning-based drug recommendation system that combines drug properties, user demographics, and user evaluations. The system's accuracy rate of 90.6% showed how successful the suggested strategy is a neural network model-based drug recommendation system that integrates user feedback and drug characteristics was proposed by Zhai et al. (7). The system's accuracy rate of 88.2% demonstrates the potency of the suggested strategy. In conclusion, the application of ML algorithms and NLP techniques to drug recommendation

systems based on user reviews has the potential to completely transform the healthcare sector.

Existing methodology

One existing method for drug recommendation systems based on user reviews using NLP and ML algorithms is collaborative filtering (CF). CF is a technique used for system recommendations that 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 analyzes 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 behavior to make recommendations for new drugs. The two different approaches to the CF algorithm include user-based CF and item-based CF. In user-based CF, the algorithm identifies users with similar interests based on their previous interactions with drugs and recommends drugs that are rated highly; in item-based CF, the algorithm recognizes drugs that are similar to the drugs a user has previously rated highly and recommends the same drugs. Both approaches have their strengths and weaknesses. User-based CF 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 CF 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, CF is a powerful technique for drug recommendation systems based on user reviews using NLP and ML 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 ML 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 RNN algorithm encompasses multiple stages, including collection of data, data pre-processing, 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, and dosage), user demographics (e.g., age, gender, and medical conditions), and user reviews (e.g., text comments and 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 ML algorithms. This entails extracting features from the text data, such as bag-of-words, TF-IDF, and word embedding. 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, F1 score, and others. 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 interactions 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 have a notable impact on 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 suggested technique, which combines NLP and RNN algorithms with user review-based drug recommendation systems, constitutes a thorough procedure covering data collection, pre-processing, feature extraction, model training, and evaluation. It has the capacity to provide precise and individualized medicine recommendations depending on the needs and preferences 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, as shown in **Figure 1**. 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 system must be programmed during the implementation phase utilizing the appropriate programming languages, frameworks, and libraries. Creating modules for data collection, pre-processing, feature extraction, model training with RNN, and interface creation are all included in this. During implementation, it is essential to guarantee the system's scalability, dependability, and efficiency. 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 the 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 to system design, implementation, evaluation, and optimization is crucial to developing 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.

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
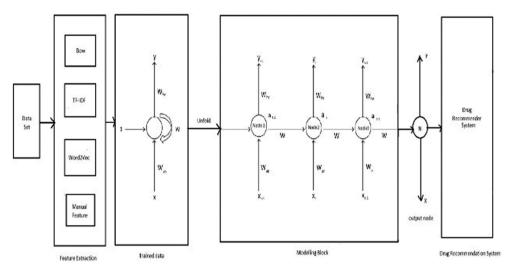
FIGURE 1 | Architecture of the system.

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, it is analyzed 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 analyze 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:

(1) Split the dataset into training and testing datasets.

(2) Let X be the pre-processed feature matrix representing the extracted features.

(3) Split X1 into a training1 set (X1_train) and testing1 set (X2_test).

(4) Initialize the parameters (weights and biases) for each layer in the deep RNN, denoted as $\theta^{(l)}$, where l represents the layer index.

(5) Specify f[X; (1), (2), f(L)] as the deep RNN model function, where L is the total number of layers.

(6) Using an optimization approach like stochastic gradient descent (SGD), train the deep RNN model by minimizing the loss function with respect to the parameters (l) for each layer: $\theta^{(l)*} = \arg\min \theta^{(l)} \{L[f(X\_train; \theta^{(1)}, \theta^{(2)}, \theta^{(L)}], y\_train]\}$, where L represents the loss function and y train is the ground truth drug ratings for the training set.

(7) Calculate the predicted drug ratings for the testing set using the trained deep RNN model: y_pred = f[X_test; $\theta^{(1)*}, \theta^{(2)*}, \theta^{(L)*}$].

(8) By contrasting y_pred with the ground truth ratings y_test, assess the model's performance using a variety of assessment metrics, including accuracy, precision, recall, F1 score, and the AUC-ROC curve.

(9) 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_user = f[X_user; $\theta^{(1)*}, \theta^{(2)*}, ., \theta^{(L)*}$]. Recommend the top-ranked drugs based on the predicted ratings y_user.

Results

The drug recommendation system is based on user reviews and ratings using the RNN algorithm. We evaluate the system's performance using a publicly available dataset of drug reviews from the website Drugs.com, as shown in **Figure 2**. 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 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 downweight the importance of common words and upweight the importance of rare words, and then the splitting of the dataset into a training set (70%) and a test set (30%) was done. We then trained a RNN classifier using a 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 five-fold cross-validation to select the optimal value that maximized the AUC of the ROC curve. The results showed that the RNN classifier has achieved an 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

FIGURE 2 | ROC versus false positives.

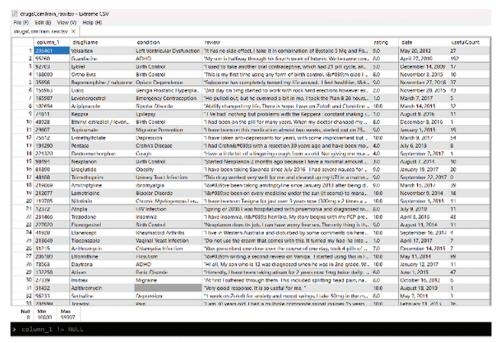**FIGURE 3 |** Medical recommendations testing dataset.

FIGURE 4 | Medical recommendations training dataset.

reviews from the dataset and re-evaluated the system's performance as shown in **Figure 3**. 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 the RNN algorithm, as shown in **Figure 4**, and its potential to assist patients and healthcare professionals in making informed decisions about drug treatment.

Conclusion

In conclusion, our proposed system of drug recommendations based on reviews and sentiment analysis utilizing RNN and NPL is an effective way of prescribing drugs to users using patient-generated data such as drug attributes, user demographics, and user reviews. Our system utilizes RNN for the classification of reviews into positive and negative reviews, and NPL techniques are used for feature extractions such as keyword, sentiment, and topic. Additionally, the five metrics (precision, recall, F1 score, accuracy, and ROC curve) of our proposed system help us 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 health professionals in making informed drug predictions. Overall, our drug recommendation system based on user reviews and sentiment analysis shows that it is able to provide accurate drug recommendations and has the potential to advance the field of personalized medicine.

References

1. Chen et al. (2020)
2. Li et al. (2020)
3. Kaur and Singh. (2020)
4. Gao et al. (2021)
5. Zhang et al. (2019)
6. He et al. (2020)
7. Zhai et al. (2020)
8. Wang L, Zhang W, He X, Zha H. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining.* New York, NY (2018).
9. Pokkuluri KS, Usha DN. A secure cellular automata integrated deep learning mechanism for health informatics. *Int. Arab J Inf Technol.* (2021) 18:782–8.
10. Settouti N, Youbi F. Convolutional and Recurrent Neural Networks for Opinion Mining on Drug Reviews. In: Hong T, Serrano-Estrada L, Saxena A, Biswas A editors. *Deep Learning for Social Media Data Analytics.* Cham: Springer International Publishing (2022).
11. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat rev Drug discov.* (2019) 18:41–58.
12. Pokkuluri KS, Nedunuri SUD. A novel cellular automata classifier for COVID-19 prediction. *J Health Sci.* (2020) 10:34–8.
13. Hines LE, Malone DC, Murphy JE. Recommendations for generating, evaluating, and implementing drug-drug interaction evidence. *Pharmacotherapy.* (2012) 32:304–13.
14. Beel J, Langer S, Genzmehr M, Gipp B, Breitinger C, Nürnberger A. Research paper recommender system evaluation: a quantitative literature survey. *Proceedings of the International Workshop On Reproducibility And Replication In Recommender Systems Evaluation.* Hong Kong (2013).

15. Pokkuluri K, Nedunuri U, Devi U. Crop disease prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata. *Int Arab J Inf Technol.* (2022) 19:765–73.

16. Dumbreck S, Flynn A, Nairn M, Wilson M, Treweek S, Mercer SW, et al. Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines. *BMJ.* (2015) 350:h949.

17. Youn S, Kim S. Understanding ad avoidance on Facebook: antecedents and outcomes of psychological reactance. *Comput Hum Behav.* (2019) 98:232–44.

18. Bhimavarapu U, Chintalapudi N, Battineni G. A fair and safe usage drug recommendation system in medical emergencies by a stacked ANN. *Algorithms.* (2022) 15:186.

19. Reynolds-Tylus T, Bigsby E, Quick BL. A comparison of three approaches for measuring negative cognitions for psychological reactance. *Commun Methods Meas.* (2021) 15:43–59.

20. Maddula P, Srikanth P, Sree PK, Rao PBR, Murty PS. COVID-19 prediction with Chest X-Ray images using CNN. *Proceedings of the 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE).* Manhattan, NY (2023).