PATIENT FACING SYSTEMS

An Enhanced Symptom Clustering with Profile Based Prescription Suggestion in Biomedical application

R. Vijayarajeswari[1] · M. Nagabhushan[2] · P. Parthasarathy[3]

Abstract

The application of data mining has been increasing day to day whereas the data base is also enhancing simultaneously. Hence retrieving required content from a huge data base is a critical task. This paper focus on biomedical engineering field, it concentrates on initial stage of database such as data preprocessing and cleansing to deal with noise and missing data in large biomedical data sets. The database of biomedical is huge and enhancing nature retrieving of specific content will be a critical task. Suggesting prescription with respect to identified disease based on profile analysis of specific patient is not available in current system. This paper proposes a recommendation system of prescription based on disease identification is done by combining user and professional suggestion with profile based analysis. Hence this focuses on profile based suggestions and report will be generated. The retrieving of specific suggestion from a huge database is done by hybrid feature selection algorithm. This approach focuses on enabling recommendation based on user profile and implementing Hybrid feature selection algorithm to retrieve specific content from a huge database. Hence it attains better retrieval of required content from a huge database compared to other existing approaches and suggests better recommendation with respect to user profile.

Keywords Recommendation · Profile based analysis · Feature selection · Preprocessing and cleaning

Introduction

Data mining application is enhancing now days and its application is used in variety of way in bio medicine hence it is related to more subfields in biomedical. In general, data mining and machine learning applications were focused on clinical applications like decision making in medical practitioners and interpretation of image and signal data [1–5]. More recent systems like, epidemiology, bio-surveillance and bio informatics have increased the attentions. Current system of disease identification based on symptoms is done through semantic clustering where the prescription is done through recommendation.

This article is part of the Topical Collection on *Patient Facing Systems*

✉ R. Vijayarajeswari
vijayarajeswarir@mahendra.info

[1] Department of Computer Science & Engineering, Mahendra Engineering College (Autonomous), Namakkal, India

[2] Department of Electronics and Communication Engineering, M.S. Ramaiah Institute of Technology, Bangalore, India

[3] School of Electrical Engineering, VIT University, Vellore, India

Recommendation to a product or any other item will provide better suggestion to a user and attains better satisfaction. Hence implementing this one in biomedical field will support better treatment for patient. Where current recommendation is focuses on collaborative filtering, content based and hybrid recommendation approaches are available. But in field of biomedical it should focus on patient details because it should save a life of a person. In collaborative filtering, suggestions are made to new users based on products that are liked by similar users [6]. Content based filtering focuses on user history analysis from those suggestions are provided to users based on their own interest. Hybrid based recommendation is available which combines both content and collaborative filtering approaches to carryover the advantages of both approaches.

To eliminate the irrelevant and redundant features and choosing better subset of features would produce better classification of patterns that were belonging to various classes is done through feature selections [7–9]. Filter and wrapper methods are two major categorization of feature selection. For evaluating the relation of a set of features with respect to criteria has been done by filter algorithms which utilize independent measures (information measures, distance measures and consistency measures). To calculate the value of features

wrapper algorithms utilize particular learning algorithms. In general, compared to filtering methods, wrapper methods consume more cost and it is expensive when dealing with high dimensional data or huge data. Data preprocessing is an important step in data retrieval but most often neglected. Gathering of data are mostly loosely structured hence it result in out of range of values, impossible combination and missing values. If the dataset contains wrong information then the result obtained would also be irrelevant [10]. Here it is focused importantly because utilized datasets are medical datasets. Hence preprocessing is an important factor which avoids the irrelevant and noisy data before considering it for processing. Preprocessing will reduce possibility of results to misleading results. In section 2 the related concepts of biomedical and data mining is explained briefly and proposed work is discussed in section 3. The result and its discussions have been explained in section 4.

Key ideas for proposed system: In biomedical field, suggesting prescription to patients based on identified disease is available. But suggesting prescription with respect to patient details like age, body conditions are not available in current system. Therefore symptom analysis and disease identification has been implemented by semantic clustering however the prescription by recommendation based on user profile like age, Blood pressure will enhance better result and save a person's life.

Related work

Some researchers describes the complexity procedures in retrieving the biomedical data from large database requires several processing steps like such as data acquisition, preprocessing (color processing and image enhancement), classification and feature extraction. For performing these functions tools are available individually however integration is required for complex image analysis tasks [11]. Hence a novel framework based on web services has been presented that provide access to tools and methods for data mining of biomedical image data. Web services were straightly integrated to the TAVERNA or similar workflow management allows its integration to different image processing pipelines. Due to biomedical datasets security is an issue therefore proper authentication and encryption mechanism has been implemented to ensure appropriate security. The working of data mining in bio medical datasets to predict heart conditions was introduced by other researchers in later years. These datasets are more valuable in making valuable medical decisions [12]. In discovering hidden patterns from huge datasets the effective analysis tools are the disadvantage. This approach is to extract the optimal significant features and patterns from a past unstructured data of heart patient. Initially data preparation is done in this structured data of heart patient is extracted from unstructured data. Preprocessing is done in order to apply different data mining techniques. K-means clustering is used to retrieve required content in order to extract hidden patterns which were related to heart patients. Later some others present's recommendations system and its application in social media to suggest people and tags that are related to user. To make a suggestion within the enterprise the relationship information among people, tags and items are collected from different sources within the enterprise [13]. Each recommended item is accompanied by an explanation that includes the people and tags that led to its recommendation, as well as their relationships with the user and the item. Approaches towards data preprocessing techniques and association rules was described in later years. The complexity of data and to obtain better conditions to subsequent analysis data pre-processing is an important step. By preprocessing data analysis is performed more accurately and efficiently. The steps that are included are as follows, sampling, transformation, de-noising, and normalization and feature extraction. Pre-processing technique is also useful for association rules algo. Like- Apriori, Partitioned, Princer-search algo and many more algo [14].

Sony et al. [15] States the recommendation system in data mining techniques. Here suggesting items to a client based on User-based approach, Item based approach, Hybrid recommendation approaches and related research in the recommender system. The aim of recommender system is to suggest online items among users and merchant thereby increases the profit of organization. Hence applying this concept in suggesting prescription to identified disease will attain better result in protecting a life of a person. Hesham et.al. Describes the issues in biomedical research through the field of data analysis and data mining [16]. The current research areas like bioinformatics, medical informatics and biomedical imaging. The main issue in biomedical research is addressing the security concerns in maintaining and accessing medical data with facilitating the ability to integrating and correlating different biomedical database is a major research work.

Proposed system

The suggestion of prescription was available only based on disease and it does not focus on individual patient profile. Therefore current approach provides suggestion based on profile based retrieval. The necessary parameters for profile based recommendation are age, patient current record and finally symptoms of the patient. Collaborative filtering, content based, hybrid and knowledge based recommendation are the approaches available in present system. But in biomedical field recommendation should done based on patient profile without general suggestions. In general product suggestions can attain better results in case of available recommendation methods [17]. Collaborative filtering based suggestions recommend items to particular user that similar users liked.
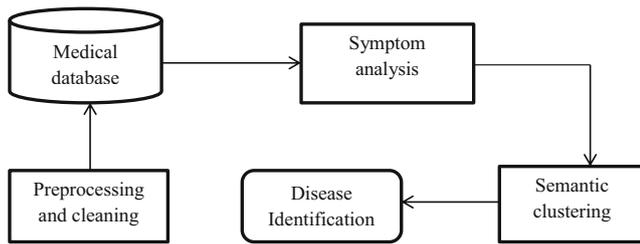
Fig. 1 Medical dataset preprocessing

Similarly content based recommendation suggests items to user based on history. From history an analysis has been done and user interest is identified. Based on identified factor suggestion are done to users from database. Hybrid recommendation is combination of two approaches to overcome disadvantages and attains better results. Hence these approaches won't provide better suggestion in field of medical datasets.

Preprocessing

The issues in real world preprocessing are incomplete due to lacking attribute variables, lacking certain attributes of interest or containing only aggregate data. If datasets contain error or outlier is represented as noisy. The data containing discrepancies in code or names are represented as inconsistent. The following are the task that takes place in preprocessing to produce optimal result from huge database is data cleaning, data integration, data transformation, data reduction and data discretization [18]. Data cleaning is the initial step which includes, filling in missing values, smooth noisy data identify or remove outliers and resolving inconsistencies. In integration process integrating multiple databases, data cubes and files are done. In data reduction, it should produce the similar analytical results while reducing the volume of data. In case of unfilled content will occur as a part of data reduction, the numerical attributes has been replaced with nominal ones. When all this process is done in a database the result output

will be an optimal one. Because searching required content from a huge database is a difficult process. In particular, biomedical database should contain entire field for optimal result extraction. Therefore for this process preprocessing has been done to fill all the fields and attains better results in final. Hence quite changes in database attain better results in final. (Figs. 1, 2, 3, 4 and 5).

Profile based recommendation

In general, recommendation approach is used to suggest best products to a new user. While in medical field suggesting prescription according to user and based on their profile is an issue. Hence considering necessary parameters such as, age, current body condition of the patient and other necessary details will improve in taking decision in optimal way. Hence clustering symptoms with consideration of patient body condition will attain accurate result in prescription identification [19]. While clustering, the symptoms of patient would be clustered and based on it disease identification has been detected accurately. Then based on identified disease the suggestion of prescription has been processed. In our approach the suggestion of prescription has been done to patient by means of considering the personal records.

Initially clustering of symptoms has been done by semantic clustering with ontology to attain better suggestion in medical field in terms of detecting disease. Recommending suggestion has been done from both professional part and user suggestion and not based on user personal detail. In between clustering symptoms and suggesting prescription another factor has been considered is profile based suggestion. Once disease has been identified the next factor is suggestion of prescription previous to this step a new condition will be checked. The patient age, current body condition includes blood pressure, sugar content, major operation, minor surgery and all other details.

Fig. 2 PBPS architecture
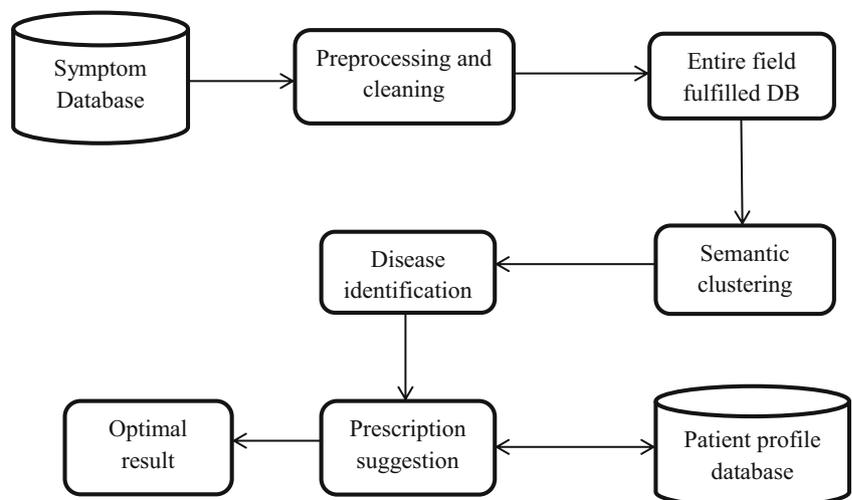
Age	Blood pressure	Sugar level	Hemoglobin	Major surgery	Minor surgery

Fig. 3 Necessary parameters of patient

Semantic clustering in grouping user symptoms is held by following equation, and ontology is another factor considered to overcome the domain word issues while clustering symptoms.

$$Sim_{onto}(Q, DS) = \left\{ |Q_{sym} U\ DS_{sym}|\ if\ |Q_{sym} U\ DS_{sym}|\neq 0 \quad (1) \right.$$

Where Q is a set of symptoms assigned to query Q and DS is a set of symptoms stored already in database. k is small constant, e.g. 0.1. Resulted number represents ontology-based similarity measure. Whole similarity is calculated then computed as multiplication is stated below,

$$Sim(Q, DS) = Sim_{onto}(Q, DS) * Sim_{TF-IDF}(Q, DS) \quad (2)$$

Where,

$$Sim_{TF-IDF}(Q, DS) = (DS \times Q)/|DS\|Q| \quad (3)$$

The Hybrid Feature Selection Method HFSM

HFSM has the ability to reduce the feature space by selecting terms which are statistically and semantically pertinent through a weighting model.

$$HSw = \lambda * r\chi 2w + 1 - \lambda \quad (4)$$

where λ is a weighting parameter between 0 and 1.

To select the most p pertinent terms, three steps are followed:

Calculate the hybrid measure HFSM for each term in the dataset;

Sort the term in descending order of their criterion function;

Select the top p terms from the sorted list.

Once semantic clustering has been done identification of disease will be conducted. After clustering based on patient profile suggestion of prescription will vary with respect to patient personal information. The issue is same prescription could not be given to all patients who are in different condition [18–22]. Therefore with respect to patient current body condition and report suggestion of prescription is implemented. Symptom clustering and patient detail is combined and analyzed to suggest optimal results.

Procedure to obtain optimal result

Step 1: Start
Step 2: Medical datasets stored in database
Step 3: Preprocessing and cleaning process is done to fill missing and wrong information
Step 4: Semantic clustering of symptoms
Step 5: Analyzing patient details and understanding user current body condition
Step 6: Based on identified disease and body condition of patient prescription has been suggested.
Step 7: Optimal prescription is attained from recommendation
Step 8: Stop

Initially symptom database is taken for processing like identification of disease through analyzing symptoms of patient. To attain optimal result by data mining, preprocessing and cleaning, is implemented to fulfill the missing values, irrelevant information in DB. Preprocessing will reduces the misleading results in final and irrelevant and noisy data will be reduced. Once the preprocessing process has been completed semantic clustering has been takes place to cluster the symptoms of user [23–25].

Fig. 4 Accuracy and symptom analysis evaluation
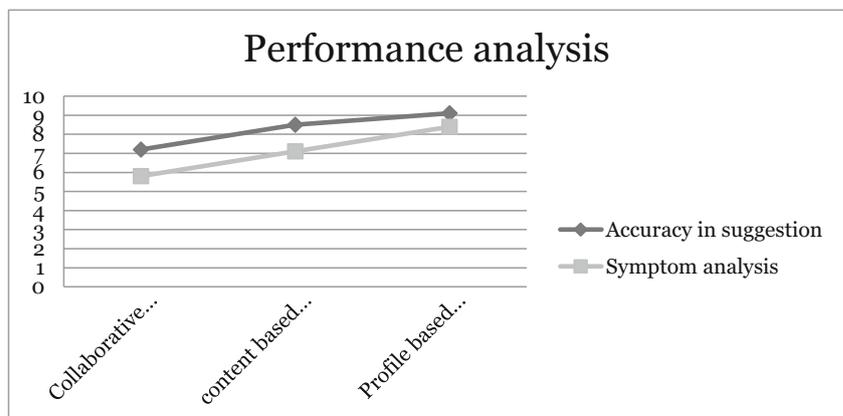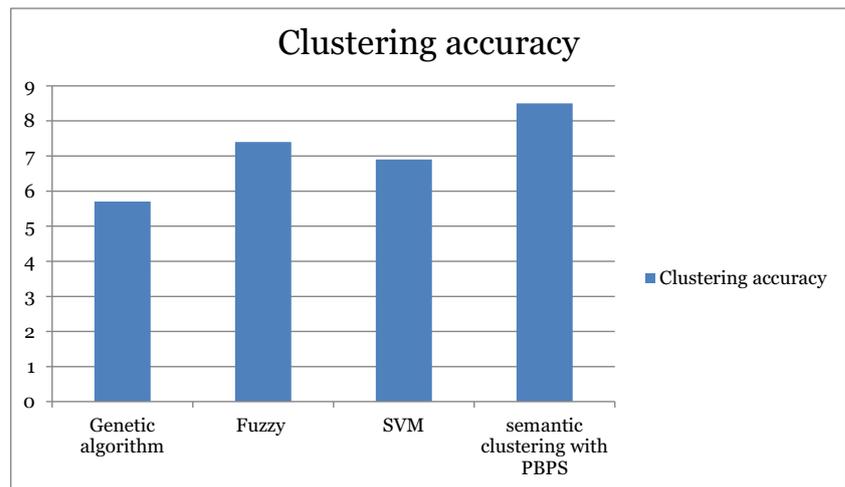
Fig. 5 clustering accuracy with various clustering algorithms

By dint of clustering disease is identified and based on user feedback and professional suggestion prescription to disease has been extracted and suggested. In addition patient profile is considered in separate DB and combining detected disease and current body condition using the parameters like age, blood pressure, sugar, major surgery, minor surgery and other details of the user. From these details suggestion of prescription has been done because similar suggestion of medicine will not attain better results for all patients. Some medicine won't work efficiently if patient current body conditions various due to the factors mentioned above. While extracting suggestions for identified disease with respect to profile based prescription will leads to optimal result. Compared to current recommendation approach in medical database profile based suggestion will attain better result in efficient way and saves patient life.

Result and discussion

The implementation of proposed approach attains better results in terms of efficient suggestion with respect to user profile. Instead of suggesting prescription based on identified disease through symptom analysis in general way will attain result without optimal level of performance. Hence considering user profile and records and suggestion are done based on that will attain better and exact result to individual patients. Existing attains suggestions in similar way to all kinds of patient and it won't work better performance.

The accuracy in suggesting prescription through recommendation system is shown in above graph. Collaborative filtering, content based recommendation suggestion attains better suggestion but not in accurate way. Hence accuracy will obtained when user satisfaction and exact suggestion provided. Therefore keeping it as an objective and it attained by profile based recommendation.

The above graph shows the accuracy in clustering symptoms of user. Where existing approach attains better accuracy but parameters considered is a necessary one. Just focusing on symptoms alone will leads to better identification of diseases but concentrating on user profile will attain optimal result to individual. Hence this could be attained by our proposed approach.

Conclusion

The objective of the work is to attain exact suggestion of prescription based on identified disease through symptoms analysis was implemented through PBPS approach. Profile based prescription suggestion is done by implementation of semantic clustering of symptoms with addition of ontology disease has been identified in effective way. Once disease is identified, prescription suggestion should be provided through database which contains suggestions from professional and user's feedback. In addition to made accuracy in suggestion profile based suggestion would obtain better results by dint of analyzing patient details like age, BP, sugar level, major operation and minor surgery and other parameters in medical record will be used. Understanding user current body condition and prescription suggested would be combined and according to patient condition prescription has been suggested. Hence in the field of medical this approach attains better results compared to existing recommendation system and this method attain optimal results.

Compliance with ethical standards

Conflict of interest   The author's has no conflict of interest in submitting the manuscript to this journal.

Ethical approval   This article does not contain any studies with human participants performed by any of the authors.

References

1. Lokesh, S., Kumar, P. M., Devi, M. R., Parthasarathy, P., and Gokulnath, C., An automatic Tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map. Neural Comput. & Applic.:1–11, 2018.

2. Kumar, P. M., Lokesh, S., Varatharajan, R., Babu, G. C., and Parthasarathy, P., Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Futur. Gener. Comput. Syst. 86:527–534, 2018.

3. Kanisha, B., Lokesh, S., Kumar, P. M., Parthasarathy, P., and Chandra Babu, G., Speech recognition with improved support vector machine using dual classifiers and cross fitness validation. Pers. Ubiquit. Comput.:1–9, 2018.

4. Bellazzia, R., Diomidousb, M., Sarkarc, I. N., Takabayashid, K., Zieglere, A., and McCrayf, A. T., Data Analysis and Data Mining: Current Issues in Biomedical Informatics. Methods Inf. Med. 50(6): 536–544, 2011. https://doi.org/10.3414/ME11-06-0002.

5. Mathan, K., Kumar, P. M., Panchatcharam, P., Manogaran, G., and Varadharajan, R., A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease. Des. Autom. Embed. Syst.:1–18, 2018.

6. Bhatt, B., Patel, PJ, Gaudani, H., A Review Paper on Machine Learning Based Recommendation System. International Journal of Engineering Development and Research (www.ijedr.org) IJEDR | Volume 2, Issue 4 | ISSN: 2321–9939, 2014.

7. Varadharajan, R., Priyan, M. K., Panchatcharam, P., Vivekanandan, S., and Gunasekaran, M., A new approach for prediction of lung carcinoma using back propagation neural network with decision tree classifiers. J. Ambient. Intell. Humaniz. Comput.:1–12, 2018.

8. Doukas, C., Maglogiannis, I., and Chatziioannou, A., An Open Web Services - based Framework for Data Mining of Biomedical Image Data. Proceedings of the 9th International Conference on Info~ation Technology and Applications in Biomedicine, ITAB 2009, Larnaca, Cyprus, 2009.

9. Fatima, M., Basharat, I., and Anjum, D. S. A. K. A. R., Biomedical (Cardiac) Data Mining: Extraction of significant patterns for predicting heart condition. Computational Intelligence in Bioinformatics and Computational Biology, 2014 IEEE Conference. https://doi.org/10.1109/CIBCB.2014.6845499.

10. Parthasarathy, P., and Vivekanandan, S., A comprehensive review on thin film-based nano-biosensor for uric acid determination: arthritis diagnosis. World Review of Science, Technology and Sustainable Development 14(1):52–71, 2018.

11. Parthasarathy, P., and Vivekanandan, S., Urate crystal deposition, prevention and various diagnosis techniques of GOUT arthritis disease: a comprehensive review. Health Information Science and Systems 6(1):19, 2018.

12. Guy, I., Zwerdling, N., Ronen, I., Carmel, D, Uziel, E., Social Media Recommendation based on People and Tags" SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval Pages 194–201.

13. Parthasarathy, P., and Vivekanandan, S., A numerical modelling of an amperometric-enzymatic based uric acid biosensor for GOUT arthritis diseases. Informatics in Medicine Unlocked, 2018.

14. Parthasarathy, P., and Vivekanandan, S., Investigation on uric acid biosensor model for enzyme layer thickness for the application of arthritis disease diagnosis. Health Information Science and Systems 6:1–6, 2018.

15. Hovale, S. V., and Poonam, G., Survey Paper on Recommendation System using Data Mining Techniques. International Journal Of Engineering and Computer Science ISSN: 2319-7242 5(5): 16697–16699, 2016.

16. Ali, H., Dempsey, K., and Patton, R., Introduction to Mini-Track: Innovative Data Analysis and Data Mining Tools in Biomedical Informatics. 45th Hawaii International Conference on System Sciences, 2012. https://doi.org/10.1109/HICSS.2012.363.

17. Parthasarathy, P., and Vivekanandan, S., A typical IoT architecture-based regular monitoring of arthritis disease using time wrapping algorithm. Int. J. Comput. Appl.:1–11, 2018.

18. Priya, S., Varatharajan, R., Manogaran, G., Sundarasekar, R., and Kumar, P. M., Paillier homomorphic cryptosystem with poker shuffling transformation based water marking method for the secured transmission of digital medical images. Pers. Ubiquit. Comput.:1–11, 2018.

19. Varatharajan, R., Preethi, A. P., Manogaran, G., Kumar, P. M., and Sundarasekar, R., Stealthy attack detection in multi-channel multi-radio wireless networks. Multimed. Tools Appl.:1–24, 2018.

20. Basha, A. A., Vivekanandan, S., and Parthasarathy, P., Blood Glucose Regulation for Post-Operative Patients with Diabetics and Hypertension Continuum: A Cascade Control-Based Approach. J. Med. Syst. 43(4):95, 2019.

21. Basheer, S., Gandhi, U. D., Priyan, M. K., and Parthasarathy, P., Network Support Data Analysis for Fault Identification Using Machine Learning. International Journal of Software Innovation (IJSI) 7(2):41–49, 2019.

22. Panchatcharam, P., and Vivekanandan, S., Internet of Things (IOT) in Healthcare–Smart Health and Surveillance, Architectures, Security Analysis and Data Transfer: A Review. International Journal of Software Innovation (IJSI) 7(2):21–40, 2019.

23. Suresh, A., and Varatharajan, R., Retraction Note to: Recognition of pivotal instances from uneven set boundary during classification. Multimed. Tools Appl.:1–1, 2019.

24. Parthasarathy, P., Synthesis and UV detection characteristics of TiO2 thin film prepared through sol gel route. In: *IOP conference series: materials science and engineering* (Vol. 360, No. 1, p. 012056). IOP Publishing, 2018.

25. Selvaraj, P., and Varatharajan, R., Whirlpool Algorithm with Hash Function Based Watermarking Algorithm for the Secured Transmission of Digital Medical Images. Mobile Networks and Applications:1–14, 2018.