International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Random Forest Algorithm in Drug Selection in Medical Field

Aleena Susan Mathew[1], Lashma. K.[2]

[1]PG Scholar, Computer Science and Engineering, Sree Buddha College of Engineering, Kerala, India
[2]Assistant Professor, Computer Science and Engineering, Sree Buddha College of Engineering, Kerala, India

---

Abstract - *For drug selection decision-making, datasets are taken from healthcare databases. Fuzzy recommender system for drug selection uses this data as the input. Drug selections are performed based on some features that are extracting from the dataset. Relevant features are eliminated using a random forest algorithm.*

Key Words: **Drug selection, Fuzzy recommender system, Random forest algorithm.**

1. INTRODUCTION

The knowledge of individual clinicians about clinical effects of applied drugs in everyday practice reaches a statistical significance in time with the number of drug (co-)applications to individual patients. Moreover, data are generally statistically significant among the clinicians because the numbers of treated patients and drug application occurrences are high. Although these data are available in healthcare databases of insurance companies, they are not systematically summarized in an available database, so it is not possible to establish rules for any expert system. Moreover, the setting up of such a database is problematic also for unrealistic way to ask each clinician for the reason why she/he decided to apply a selected drug (or drug combination) in a selected case [1].

II. FUZZY RECOMMENDER SYSTEM

The fuzzy logic mechanism have IF – THEN rules and which is easily interpreted by doctors/clinicians. The set of such rules called fuzzy rule base: IF (a set of conditions are satisfied) THEN (a set of consequences are inferred). For Example, IF (antibiotic is applied during 2 days in case of pneumonia) THEN (temperature is normalized). The antecedents and consequents of this IF – THEN rules are associated with fuzzy Concepts [2]. In the proposed system, the datasets are collected from healthcare databases. That External dataset contains patient, drug, diagnosis, etc are imported to the current application. The patients attending doctors are diagnosed and various medicines suggested. These information forms a real dataset which also act as the input of the proposed system. Here, fuzzy based rule mining algorithm i.e. FP-Growth tree algorithm, is applied to find the relation value of each combination. The most relations with most confidence value are identified. In FP-Growth, there is no candidate item set generation.

III. RANDOM FOREST ALGORITHM

Random forest algorithm is a supervised classification algorithm. Using number of trees, the random forest algorithm creates the forest. The more trees in the forest the more robust looks like. The random forest classifier used higher number of trees in the forest and gives the high accuracy results. Random forest algorithm or the random forest classifier can use both classification and the regression task. Random forest classifier will handle the missing values. When more trees in the forest, random forest classifier won't over fit the model.

In random forest algorithm, in the first stage, randomly selecting "k" features out of total "m" features. In the next stage, using the randomly selected "k" features to find the root node by using the best split approach. The next stage, calculating the daughter nodes using the same best split approach. The first 3 stages until form the tree with a root node and having leaf node as the target node. Finally, repeat 1 to 4 stages to create "n" randomly created trees. This randomly created trees forms the random forest [3].

In proposed system, diagnostic features affecting the drug selection are used for the feature selection. In an entire database, some of the features are found to be relevant which can be eliminated using a feature selection algorithm. A random forest based feature selection is applied. Random forest algorithm used in weka tool. Weka is the library of machine learning used to solve various data mining problems. In weka, Java programming language is used so, it can be used in any platform. Here, information based feature selection is used. Calculating the information gain for each attribute for the output variable. Weka supports feature selection via information gain using InfoGainAttributeEval Attribute Evaluator [4]. The weka supports only ARF(Advanced Recording File) file format. So, CSV(Comma-Separated Values) file must be converted to ARF file format. In the proposed system, the features are taken from the diagnosis and forms a dataset, which is in CSV file format. That file is converted to ARF file format. Attribute selection feature selection algorithm is used. Then, got relevant features. By using that features proper medicines are prescribed.

REFERENCES

[1] Monika Fedorová1, Daniela Perduková , Zdenko Pirnik Viliam Fedák, Ondrej Sukel, And Padmanaban Sanjeevikumar, (Senior Member, IEEE) , "The Fuzzy

System as a Promising Tool for Drugs Selection in Medical Practice", *IEEE Access*, pp. 27294-27301, June 5, 2018.

[2] K. Pagava and T. Kiseleva, ``New approach to estimate different drugs and/or other medical interventions effectiveness based on fuzzy logic principles,'' *Georgian Med. News*, vol. 156, pp. 65_68, Mar. 2008.

[3] http://dataaspirant.com/2017/05/22/random-forest algorithm-machine-learning

[4] https://machinelearningmastery.com/perform-feature-selection-machine-learning-data-weka