

Classification of Chemical Medicine or Drug using K nearest Neighbor (kNN) and Genetic Algorithm

Pradip A. Sarkate¹, Prof. A. V. Deorankar²

¹P.G. Student, Department of Computer Science & Engineering, Govt. College of Engineering, Amravati, India

²Associate Professor, Department of Computer Science & Engineering, Govt. College of Engineering, Amravati, India.

Abstract- Data mining techniques refers to extracting or mine knowledge information from large amount of data. In Data mining classification are supervised learning methods which are used to predict a different design model and describing important a data classes. Data classification in medical field is different from other field. Medical data classification involves multi class classification, heterogeneous and complex data structure. In classification K- Nearest Neighbor are most popular, very simple, highly effective algorithms. Genetic algorithms are most popular technique in evaluation algorithm that used to solve optimization problem. We used k-NN and genetic algorithm to categorize drug data set, which can solve classification and optimization problem which can better optimal result in drug or medicine data set. The k-NN and genetic algorithm techniques which are improved accuracy classification method in drug or medicine data set.

Key Words: data mining; KNN algorithm, genetic algorithm.

1. INTRODUCTION:

Data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining is also called as "knowledge mining in a data", which integral part of KDD (knowledge data mining) are consist series of transformation step from data preprocessing to the post processing in data mining result. There are various multinational pharmaceutical industries which are developed the medicines which are categories to original research and generic medicines. There are basic functionality of data mining such as clustering association, and classification. Classifications which are used to categories the different types of drug, classifying drug data set. The classification technique is used to improving a drug dataset or an medicine data set the classification is an important role in data mining techniques.

The K Nearest Neighbor is a most valuable popular classification algorithms in data mining technique. The genetic algorithm is an evaluation algorithm which solved an optimization problem. We will define the KNN algorithm and genetic algorithm

1.1 K Nearest Neighbor algorithm:

K nearest neighbor classification algorithm is a instance based learning or non general learning, it will simply stores

instance of training data. The nearest neighbor method is to find a predefined number of training samples closest in distance to new point, and predict the label from these instances. The number of samples can be user defined based on local density of point.

K nearest neighbor algorithm is calculated on the basis of value of k, which will define how many nearest neighbors are to be considered to define class of a training sample data point. The training sample data points are assigned weights according to their distances from sample data point.

Nearest Neighbor techniques are classified into two categories 1. Structure less NN technique and 2. Structure based NN technique. This technique is very simple and easy to implement, The K-nearest neighbor lies in first category whole data is classified into training data and sample data point. Structure based NN techniques is based on structures of data like orthogonal structure tree (OST), ball tree, k-d tree, axis tree, nearest future line and central line. The nearest neighbor is to find the K training instances which are closest to unknown instance and pick the most commonly occurring classification for these K instances.

Application of KNN

- Classification and interpretation
- Nearest Neighbor based Content Retrieval
- Protein-Protein interaction and 3D structure prediction

Drawback

- Low efficiency
- Dependency on the selection of values for K

1.2 Genetic Algorithm:

Genetic algorithm is a most important technique in evolutionary computing which is used to solve an optimization problem. To solve these optimization problem evaluation algorithm require a data structure to represent and evaluate solution from old solution's. A solution generated by genetic algorithm is called by chromosome, while collection of chromosome is called as population.

2. METHOD:

Here we proposed combines KNN algorithms and genetic algorithm to improve the classification accuracy of drug data set or medicine. We used to genetic search as better result measure to prune redundant and irrelevant attributes.

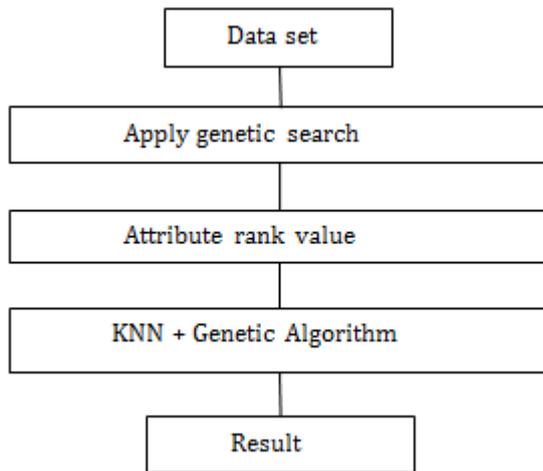


Fig: Proposed System

In this system firstly load the data set, the genetic algorithm which an evolutionary algorithm is useful for search and optimization problem. They are apply genetic search algorithm on the data set are rank based their attributes values , and then select higher ranked of attributes. We are applying two algorithm knn and genetic algorithm which classify better accuracy, the accuracy of the classifier computed as test data is an no of samples correctly classified divided to total number of sample in test data. Knn and genetic algorithms we classify chemical drugs or medicine using training data sample set. The genetic algorithm utilizes basic three operations they are selection, mutation and crossover. These operation have different type of individual properties such as population size, crossover and mutation probabilities in a genetic algorithms. In a medical field there are various application used genetic algorithm such as oncology, radiology, cardiology, endocrinology, obstetrics and gynecology, surgery, infectious, diseases, neurology and orthopedics.

3. CONCLUSION:

Classification of medical data is highly complex structure, knn is most effective classification technique to classify unknown medical data set, better result other algorithm. We used knn and genetic algorithm to optimize problem. Knn algorithm categories unknown type of drugs and genetic algorithm which solve optimal solution, it will improve the classification accuracy of drug data or medicine data set.

REFERENCES:

- [1]. M.Akhil jabbar, B.L Deekshatulu, Priti Chandra," Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm" CIMTA pp.85-94 (2013).
- [2]. Dr saed sayad,"University of toronto <http://chem-eng.utoronto.ca/~data mining>.
- [3]. Nitin Bhatia ,vandana "Survey on nearest neighbor techniques" IJCSIS,Vol 80,no 2(2010).
- [4]. Max bramer,"Principles of data mining" Springer (2007).
- [5]. S.N Sivanandam, S.N Deepa,"Introduction to genetic algorithms" Springer (2008).
- [6]. S.N. Sivanandam,S.N. Deepa, "Introduction to genetic algorithms"Springer (2008) MA. Jabbar, B.L Deekshatulu, Priti chandra. "An evolutionary algorithm for heart disease prediction" CCIS, PP 378-389, Springer(2012).
- [7]. MA. Jabbar, B.L Deekshatulu, Priti chandra, "Prediction of Risk Score for Heart Disease using Associative classification and Hybrid Feature Subset Selection" ,In .Conf ISDA, pp 628-634,IEEE(2013).
- [8]. MA. Jabbar, B.L deekshatulu, priti chandra, "classification of heart disease using ANN and feature subset selection" GJCST, VOL 13,issue 3,version1.0 pp15-25(2013).
- [9]. D.E Goldberg." Genetic algorithm in search optimization and machine learning" Addison wesely (1989)