International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.07, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Classification and Analysis of a Chemical Drug and SimilaritySearching in Chemical Drug

Pradip A. Sarkate[1], Prof. A.V.Deorankar[2]

P.G. Student, Department of Computer Sci.&Engg, Govt. College of Engineering, Amravati, India[1]

Associate Professor, Department of Computer Sci.&Engg., Govt. College of Engineering, Amravati,India[2]

ABSTRACT:

The pharmaceutical industries drug divided two categories original research and generic drug. Only multinational companies have ability researched a new drug, it will increased drug cost in our daily life. So it will important to reduce the research and developing cost of new drug. We will classify unknown type of drug and provide backing for drug screening during developed the new drugs. It will reduce the original drug cost and assistance drug development process. Classify unknown typed of drugs and calculating similar type of drugs to identifying its parameter of drug data set. We are used k nearest neighbor to classified unknown drugs and categories the drug data based on similarity of a medicine.k - NN method improve to manageoverfitting problem, the classification method based on chemical similarity and training data set.

Keywords: *chemical medicine classification data mining,k-NN search, similarity between medicines.*

I. INTRODUCTION:

In the multinational pharmaceutical companies will produce research and developing new drug, they are categories mainly two aspects original research and generic drug. Only large multinational industry have ability developing new research drugs, small companies produced generic drugs during the development of original drug. The pharmaceutical industries control drug price during patent protection period which cost is very high developing new drug. While in a market enter a new drug, they are approved the Food and Drug Administration (FDA).Classification of drug it will reduce cost of drugs and drug approval rate. We used *k*-NN classification algorithm to classify unknown type of drug on Map reduced architecture.

II. RELATED WORK:

Our primarily work concentrate on two aspects, similarity between drugs and predict categories of a new drug.

1. Similarity between drugs or medicine:

Calculating the similarity between medicine vector space model implementing cosine formula which calculate similar type of drugs based on input vector consisting various features of drugs [4]. Categories of drugs have many features depends upon chemical properties, function,color, structure and physical properties. We choose its different features of chemical drugs to calculate the similarity between drugs or medicines. So feature selection is an important in classification a chemical drugs.

While feature selectionhas two important methods used calculate similar type of drugs,Target method and Structural method. The target methods which are used to drug target areclassifying the chemical products to calculate the similarity medicines [5]. Drug target is highly accurate and provide credible theoretical basis for drug development. The structural method which used to structure as a feature in drug classification calculates the similarity between medicines [6]. DNA molecular fingerprints drugs are specified substructure of molecules DNA fingerprinting [7]. It will specify molecules as function of a quantization [9].

2. Predicting drug classification algorithm:

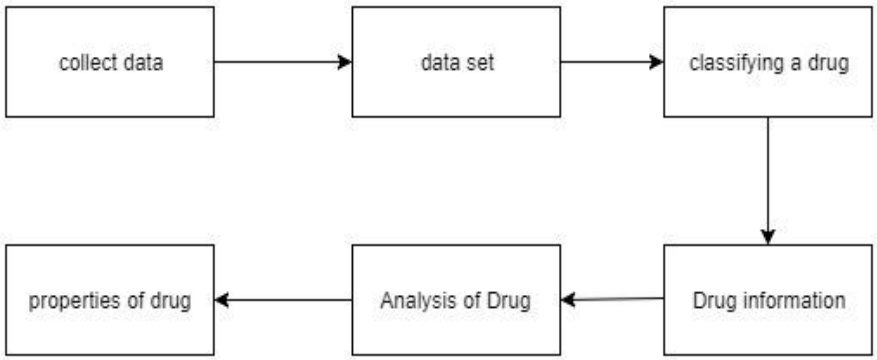
In data mining, there are different classification algorithm which are divided into supervised learning, unsupervised learning or semi supervisedlearning algorithm depending upon category of a sample data.Classify of chemical drugs which used to random forest algorithm [7] to predict drug target. Predicting unknown type of drug depend on chemical structure information used abipartite graph method to learn correlation between the chemical drugs [10].

III. PROPOSED SYSTEM:

In a proposed system classification of a chemical drugs and analysis chemical medicines we proposed k NN classification algorithm to categories the drug data set. We classify the drug with distance using k NN algorithms. We classify unknown type of drug and calculating the similarity between medicines.

A. Drug data description:

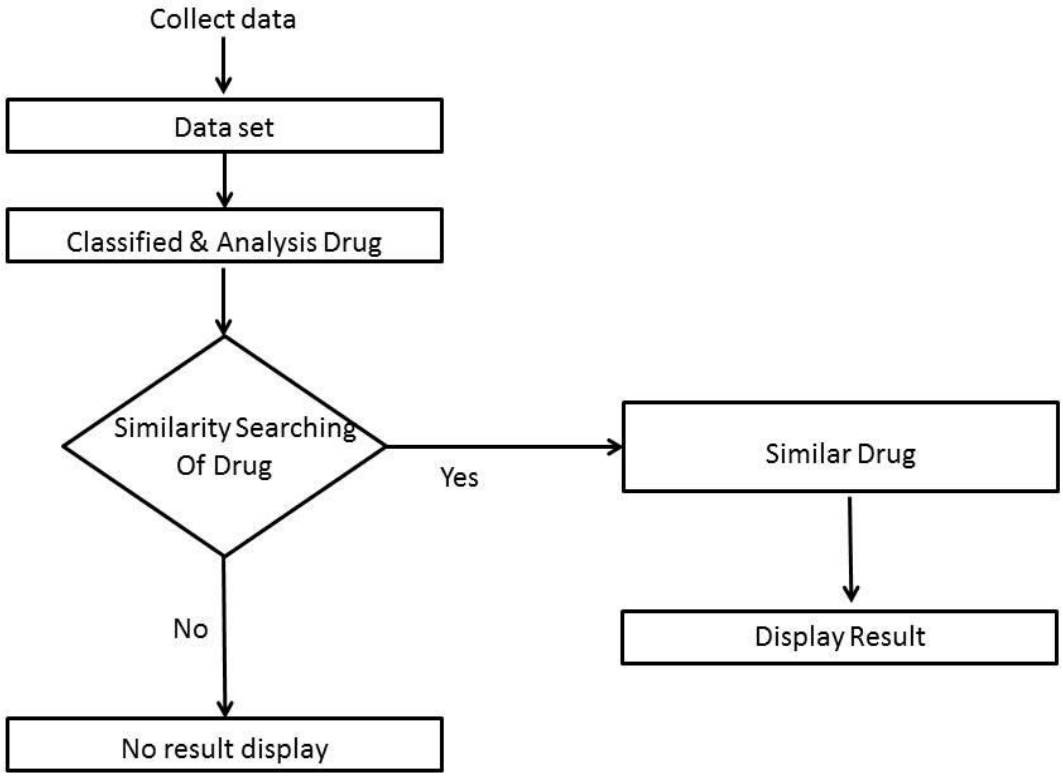
The drugs are classified chemical and biological medicine according to their characteristic. Description of drug have various information includes medicine name, molecules structures, physical and chemical properties of a drug. Pharmacodia is a big platform which is focus on pharmaceutical research and development drug which classifiedchemical drugs and provide assistance for drug screening during the development process. The fig.shows the classification of chemical drug.



In this fig. shows classify unknown type of drug and analysis chemical drugs. Drug data obtained from publicly available on the internet. Drug information includes drug name, indication, molecular structure, physical and chemical properties, chemical synthesis, route and drug related patent information.

B. Chemical classification model:

There are number of samples in drug dataset, each dataset sample is vector denotes physical and chemical properties of drug data set regarding to feature selection. We categories drug samples according to level drug dataset. They are divided into substructure drugs and classify a model from data subset to provide a predictive value of drug classification.



Similarity between chemical drugs depends on its physical and chemical properties of drugs. Regarding to the feature selection which we are calculating the similarity between medicines.

C. Classification of chemical drugs:

In data mining, classification of chemical drug we used to k-NNalgorithm to classify unknown type of drugs, calculate the distance between the drugs. The k-NNalgorithm used k samples with the smallest distance from given samples and assign to most frequently for k samples. It is k sample of training dataset sample which categories unknown type of drugs.

IV. CONCLUSION:

Classification and analysis of a chemical drug we proposed k-NNalgorithm have better accuracy to other method. We classify various types of drugs with physical and chemical properties of chemical drug according to classification model. We used k-NNalgorithm have good balance between accuracy and time complexity.It will perform good performance classification model as compared to other random forest and

decision tree model. We classify various drugs according to physical and chemical properties according to classification model.

REFERENCES:

[1]. D. G. Huang, L. Guo, H. Y. Yang, X. P. Wei, B. Jin "*Chemical Medicine Classification Through Chemical Properties Analysis",* IEEE Access, vol 5, March 2017.

[2]. Chen M, Ma Y, Song J, Lai CF, Hu B. *Smart clothing: Connecting human with clouds and big data for sustainable health monitoring*. Mobile Networks and Applications. 2016 Oct 1;21(5):825-45.

[3]. Zhang Y, Chen M, Huang D, Wu D, Li Y. *iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization*. Future Generation Computer Systems. 2016 Jan 12.

[4]. Mao W, Chu WW. "*ree-text medical document retrieval via phrase-based vector space model." In Proc.Amer. Med. Informat. Assoc. Symp.AMIA Symposium* 2002 (p. 489). American Medical Informatics Association.

[5]. A. L. Hopkins, ``*Network pharmacology: The next paradigm in drug discovery,'' Nature Chem. Biol.*, vol. 4, no. 11, pp. 682 690, Nov. 2008.

[6]. Y. Zhang, M. Chen, S. Mao, L. Hu, and V. C. Leung, ``*Cap: Community activity prediction based on big data analysis*,'' *IEEE Netw.*, vol. 28, no. 4,52 57, Jul. 2014.

[7]. H. Yu et al., ``*A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data*,'' PloS ONE, vol. 7, no. 5, p. e37608, 2012.

[8]. Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kane- hisa, ``*Prediction of drug-target interaction networks from the integration of chemical and genomic spaces*,'' *Bioinformatics*, vol. 24, no. 13, pp. 232_240, 2008.

[9]. D. R. Rhodes *et al.*, ``*Probabilistic model of the human protein-protein interaction network*,'' *Nature Biotechnol.*, vol. 23, no. 8, pp. 951_959, 2005.

[10]. Y. Zhang, ``*GroRec: A group-centric intelligent recommender system integrating social, mobile and big data technologies*,'' *IEEE Trans. ServicesComput.*, vol. 9, no. 5, pp. 786_795, Sep. 2016.

[11]. P. Willett, J. M. Barnard, and G. M. Downs, ``*Chemical similarity searching*,'' *J. Chem. Inf. Comput. ci.*, vol. 38, no. 6, pp. 983_996, 1998.

[12]. C. Burges *et al.*, ``*Learning to rank using gradient descent*,'' in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 89_96.