International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.06, June 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Classify and Analyze of Chemical Drug and Similarity in Searching 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

Multinational companies have ability to research and produce new drugs or medicine. Developing new drugs or medicines it must be increased cost in our daily life. The innovation of drug or medicine, they are divided into two categories; original research drug and generic drugs. Only large multinational industries have ability to research new drug and they are very costly. Thus it required to reduce research and development cost of new drugs. In data mining, classification technique is used to classify different categories of drugs. They are classified unknown type of drug based on training sample drugs data set. We used k- Nearest Neighbor classification technique to categories the drug data based on similarity of drugs or medicine. Similarity searching in chemical drugs and the classification technique is used to classify the unknown type of drugs and provide assistance for drug screening during the development process.

Keywords:*chemical drug classification, similarity between medicines, data mining, K-NN algorithm.*

I. INTRODUCTION:

The drug data is obtained publicly available data on the internet, Pharmacodia is a big platform that focus on pharmaceutical research and developing new drugs. The drug data includes physical and biological properties of drugs that specify drug information, drug name and indication. The drug data can be classified chemical properties and biological properties of drugs. There is large number of chemical and biological drugs; it must be automate the classification of drugs. The classification technique used to predict the unknown type of drugs and supporting role in drug screening during the development process.

Searching is very important in large number of drug dataset. Similarity searching in drug dataset is used to find the information of drug or medicine. It will provide the drug content and the drug information which identified the chemical drug data.

II. RELATED WORK:

There are many classification algorithms which are used to classify data with different categories. Random forest algorithm is used to predict drug and drug target. The kernel regression method is used to characterize the

chemical structure [10]. Rhodes et at. [11] Purposed predictive model of the interaction between proteins by selecting chemical parameter, structural selection and other characteristics as attributes.

The vector space model calculates the similarity between medicines by implementing a cosine formula [3]. Feature selection has two approaches for calculating the similarity between medicines, target method [4] and structure method [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].

III. METHOD:

K-NN model:

K- Nearest Neighbor is one of the most popular supervised classification algorithms. K-NN classification algorithm is used to predict the target label by finding nearest neighbor class. K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.

Classification of drugs data based on physical and chemical properties of drugs. The drugs data is used to categories the different types of drugs depends on properties of drugs such as molecular mass, hydrogen bond donors, bond acceptors, flexible rotation key, surface area and hydrophobic constant. Prediction of drug data are made for new instance by searching through the entire training set for the K most similar instances.

The drug data which represent mathematical description of drugs data defined as X denote the vector of the chemical properties of the drugs, where each drug represents a sample. Thus, X={*x1, x2, x3, x4, x5, x6*}, where x1 denote molecular mass, x2 denotes hydrogen bond donors, x3 denotes bond acceptors, x4 denotes flexible rotation key, x5 denotes polar surface area, and x6 denotes hydrophobic constant. There are n samples in the drug dataset $X=\{(x_1,y_1), (x_2,y_2),\ldots,(x_n,y_n)\}$, where each sample is a vector (x_n,y_n). x_n denote the chemical properties of drugs and y_n denote the category of drugs. Suppose y^t_n is the category in y_n which has value of drug category, and y^p_n is the category in y_n which no value of category. Then the drug datasets can be divided into two independent subsets x_t and x_p, where $x_t=\{(x^t_1, y^t_1), (x^t_2, y^t_2),\ldots,(x^t_n, y^t_n) \}$and $x_p=\{(x^p_1, y^p_1), (x^p_2, y^p_2),\ldots,(x^p_n, y^p_n) \}$.It will train the classification model from the dataset x_t to predict the category of drug in x_p, and then provide predictive value for the null values in y^p_n.

IV. CONCLUSION:

We proposed classification technique is used to classifydifferent categories of drugs. Large no of drugs and their physical and chemical properties of drugs, thus newclassification algorithm based on the k-NN chemical is proposed. We classified various drugs with known physical and chemical properties according to k-NN classificationmodel. There are classified drug and drug target which is used to predict a new drug. Itwill identify similarity calculation between drug and target chemical datasets. The classification method based on chemical similarity depends entirely based on available trainingdata and the similarity searching in chemical drug are identified similar type of drugs and provideinformation of drugs.

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. 232fi240, 2008.

[9]. D. R. Rhodes *et al.*, ``*Probabilistic model of the human protein-protein interaction network*," *Nature Biotechnol.*, vol. 23, no. 8, pp. 951fi959, 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. 786fi795, 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. 983fi996, 1998.

[12]. C. Burges *et al.*, ``*Learning to rank using gradient descent*," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 89fi96.