Predicting Drug Data And Similarity Searching In Drug Database

Pradip Sarkate[1], Prof. A. V. Deorankar[2]
[1]Dept of Computer Sci. & Engg
[2]Associate Professor, Dept of Computer Sci. & Engg
[1, 2] Govt. College of Engineering, Amravati, India

Abstract- *The drug database or drug information is publicly available on the internet; DrugBank is a resource that detailed drug data. To gather drug data information available on drugbank data set it will identify the new drug, it will predict of a new drug based on training sample data set. Predicting a new drug or target similarity searching in chemical drug is an important to identified new drug. It will define drug molecular descriptor or the content of drug that identified the drug descriptor. Molecular similarity searching is identified chemical drug to target drug and searching chemical drug. In computational molecular or content similarity searching technique used KMP algorithm in chemical drug.*

Keywords- Similarity searching, predicting drug, KMP algorithm.

I. INTRODUCTION

The drug information includes drug name, indication, molecular structure and physical and chemical properties. Calculating similarity between medicines which is depends on chemical structure and physical properties of a chemical drug. The similarity searching measure into two distinct classes which exact match searching database and partial match searching in database record. They are defined structure searching and substructure searching which identify drug and target molecules. The content of similarity searching or molecules similarity searching is based on similar properties and to predict molecules based on knowledge derived from measured properties of other molecules. Comparison of two items, it compare of molecules structure, drug information, indication, feature selection and properties of drugs.

In molecular similarity searching, relationship between two molecules measure with large number molecule descriptor are closely related and small number of molecules descriptor unrelated. The similarity searching based on number substructure fragment are common molecules and associate most similar coefficient.

II. RELATED WORK

In similarity searching the chemical drug which represent as fragment based descriptor that define substructure of drug database [1]. They will define structure of chemical molecular will divided into substructure. In structure based which calculate similarity between medicines as feature selection method used as fingerprint calculated the similarity coefficient. Tanimoto coefficient calculates similarity between drug molecules which find similarity and dissimilarity in chemical database [2]. Similarity coefficient is obtain an quantization similarity between pair structure which defined as similar coefficient, association coefficient, distance coefficient, correlation and probabilistic coefficient [3,4,5].

In cheminformatics similarity measure in document database, it will retrieve and manage document database. Retrieve the text information from chemical database using textual information technique to retrieve the document database [6].

METHOD

Similarity measured in chemical drug dataset which exact search in dataset and the partial search in we used KMP algorithm of string matching in text and pattern.

KMP algorithm:

The KMP (Knuth Morris Pratt) is string searching algorithm searched text within main text string. It will search text and pattern searching on document or database, the KMP will match the exact character in document or database. The KMP algorithm discovered liner time string matching which analysis naïve algorithm. In drug molecules or content of drug similarity measurement based on KMP algorithm which matching drug depend on properties of chemical drugs.

Steps of KMP algorithm:

Step 1: initialize the input length variable such as text, pattern, prefix function of pattern and number of character match.
Step 2: define the variable beginning of the match.

Step 3: compare the first character of pattern match to first character of text, if found increment character match value by 1 otherwise substitute the character match.
Step 4: check all pattern match in text, if not repeat search process.
Step 5: look for the next match.

Matching drug molecules or descriptor based on drug information, name, indication and the characteristic of drug structure. The KMP algorithm used similarity searching in drug search in content of drug molecules to other drug which are exact match in drug dataset. The weighted schema is used to different feature in molecules determining the similarity searching to other molecules.

III. CONCLUSION

Similarity searching in chemical drug database, we used string matching KMP algorithm which measure similarity in drug database. It will comparing drug data to other data sets and identified new drug. Predicting drug molecules based on similarity searching in drug dataset, its identified new drug which are compared to another drug. It will identify similarity calculation of chemical which is predicting the new drug. The KMP algorithm discovered first linear time string matching algorithm by analysis of the naïve algorithms.

REFERENCES

[1] P.A. Bath, C. A. Morris and P. Willett. “Effect of standardisation of fragment-based measures of structural similarity”. Journal of Chemometrics, 7, pp. 543, 1993.
[2] Fligner, M.A.; Verducci, J.S.; Blower, P.E. A modification of the Jaccard-Tanimoto similarity index for diverse selection of chemical compounds using binary strings. Technometrics 2002, 44, 110-119.
[3] P. H. A. Sneath and R. R. Sokal. “Numerical Taxanomy”. Freeman, San Francisco, 1973
[4] P. Willett. “Similarity And Clustering In Chemical Information Systems”, Research Studies Press, Letchworth, (1987)
[5] D. Ellis, J. Furner-Hines and P. Willett. “Measuring the degree of similarity between objects in text retrieval systems”. Perspective in Information Management. 3:128-149, 1993
[6] P. Willett. “Chemoinformatics: an application domain for information retrieval techniques”. In Proceedings of the 27th Annual international ACM SIGIR Conference on Research and Development in information Retrieval SIGIR '04. ACM, New York, NY, 393-393, 2004
[7] M. A. Miller. “Chemical Database Techniques in Drug Discovery”. Nature Reviews Drug Discov.,1, pp. 220-227, 2002
[8] M. A. Johnson and G. M. Maggiora. “Concepts and Application of Molecular Similarity”, John Wiley & Sons, New York (1990)
[9] Concepts and Applications of Molecular Similarity; Johnson, A.M.; Maggiora, G.M., Ed.; Wiley: New York, 1990.
[10] Y. Zou, Y. Zhang, and D. Lord, ``Analyzing different functional forms of the varying weight parameter for _nite mixture of negative bino- mial regression models," Anal. Methods Accident Res., vol. 31, no. 1, pp. 39_52, 2014.
[11] P. Vos et al., ``AFLP: A new technique for DNA fingerprinting," Nucl. Acids Res., vol. 23, no. 21, pp. 4407_4414, Jan. 1995.
[12] P. Willett, J. M. Barnard, and G. M. Downs, ``Chemical similarity searching," J. Chem. Inf. Comput. Sci., vol. 38, no. 6, pp. 983_996, 1998.
[13] Y. Zhang, ``GroRec: A group-centric intelligent recommender system integrating social, mobile and big data technologies," IEEE Trans. Services Comput., vol. 9, no. 5, pp. 786_795, Sep. 2016.