IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Comparative Study of Supervised and Semi-Supervised Learning for Enhanced Drug Prediction

V. Jagadeesan[1*], K. Palanivel[2]

[1,2]Dept of Computer Science, AVC College, Mayiladuthurai, India

[*]*Corresponding Author: jagadeesanv0@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 18/Aug/2018, Published: 30/Sep/2018

***Abstract*—**Several precautions should be taken in using pharmaceutical drugs, for both healthcare professionals, who prescribe and administer drugs, and for drug consumers. Factors such as interactions among the prescribed drugs, interactions with the patient's current medication, side effects to be avoided, and contraindications, need to be carefully considered. Additionally, the presence of some drug properties, such as side effects and effectiveness, depends on characteristics of patients, such as age, gender, lifestyles, and genetic profiles. The goal is to provide a system to assist medical professionals and drug consumers in choosing and finding drugs that suit their needs. And develop an approach that allows querying for drugs that satisfy a set of conditions. The approach allows users to specify side effects and tailors the answers based on user specification. Finally utilize drug data from multiple data sources. However, drug data are usually noisy and incomplete as they are either manually curated or automatically extracted from text resources such as drug labels. To cope with incomplete and noisy data, data mining techniques were designed and implemented which include clustering and classification algorithms. Then the developed system was used for comparative analysis of supervised and semi-supervised learning using performance metrics. The result shows that Semi-supervised method provided 40% improved response time in comparison with Supervised method in Drug Retrieval System.

***Keywords*—** Drug query system, Data mining, Clustering, Classification, Semi-supervised learning.

I. Introduction

Predictive analytic is a branch of data mining concerned with the analysis of data to identify underlying trends, patterns, or relationships to predict future probabilities and trends. It encompasses statistics, data mining and game theory that analyze current and historical facts to make predictions about future events of interest. In predictive modeling, data is collected, a statistical model is formulated, predictions are made and the model is validated or revised as additional data becomes available. Clinical data mining is based on strategic research to retrieve, analyze and interpret both qualitative and quantitative information available from medical datasets or records. Predictive data mining automatically create classification model from training dataset, and apply such model to automatically predict other classes of unclassified datasets. Predictive data mining deals with learning models to support clinicians in diagnostics, therapeutic, or monitoring tasks. It learns from past experience and apply knowledge gained to future situations, by applying machines learning methods to build multivariate models from clinical data and subsequently make inferences on unknown data. Machine learning model is related to the exploitation of supervised classification approaches. Prior to applying the learning model, the data is pre-processed to remove noise and ensure data mining principle is applied on real data. Predictive data mining is the most common type of data mining that has the most application in business and real life that is centered on data pre-processing, data mining and data post-processing collectively referred to as Knowledge Discovery in Databases. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies. Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

For medical practitioners, prescribing drugs requires careful considerations of several factors, such as interactions among the prescribed drugs, interactions with the patient's current medication, and contraindications. In some cases, according to patients' conditions and lifestyles, there are particular side effects that should be avoided as they could cause serious health conditions or injuries. The process is further complicated by the fact that the presence of some drug properties, such as side effects, depends on characteristics of the patients, such as age, gender, and genetic profiles. Having to consider all these complicated factors can be a huge burden to medical practitioners. The basic data mining is shown in fig 1. The data mining process includes reporting, dash boards, planning and modeling system. Finally provide decision support system in many applications.

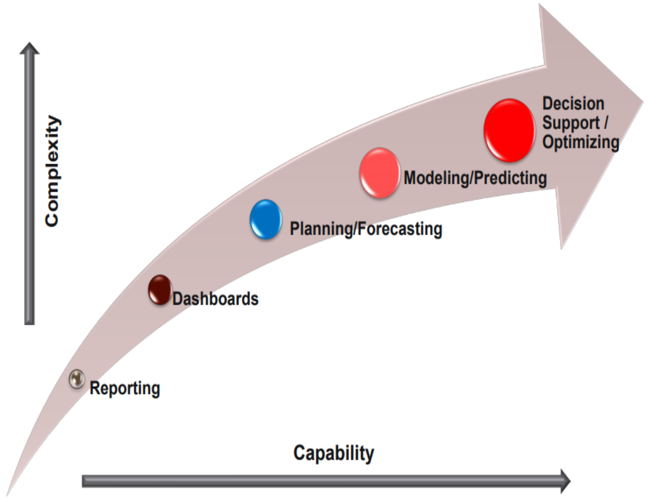


Fig 1.Data mining

II. Related Work

E. Bressoet al [1] proposed the models obtained for a set of selected system with these two machine-learning methods are then evaluated by cross-validation and tested directly with new drugs. Finally, some elements are provided for model interpretation. Single table datasets designed for DT learning represent each drug by an attribute-value vector. Four types of descriptors retrieved from NetworkDB are used to generate these attributes: the first is the class information, i.e. the studied SEP, the second one includes drug categories, the third one lists all drug targets with for each target, three attributes referring to the type of action of the drug (activation, inhibition and other) and the fourth concerns clusters of similar drugs. Relational datasets designed for Inductive Logic Programming (ILP) consist in a set of tables extracted from Network-DB describing drugs properties and background knowledge. Drugs properties are the same as in the single-table dataset, i.e. categories, targets and clusters. Background knowledge includes GO annotations, domain composition, interactants and pathways of each drug target. Relationships between GO terms constitute an additional table.

T. Liu, et.al, [2] presented an approach for predicting novel associations between drugs and diseases that can operate on both drugs with approved indications and on novel molecules with no indication information. Given a query association, we measure the similarity of the pertaining drug and disease to drug–disease pairs that are known to be associated, and rank the accumulative evidence for association using a logistic regression scheme. The prediction process is aided by a comprehensive drug–disease association data set that we have compiled and a collection of novel drug–drug similarity measures. Importantly, we show the potential utility of approach also in a personalized medicine setting, in which a disease name is replaced by a gene expression signature; and consequently, disease–disease similarity is measured via the similarity of the corresponding signatures. For each query drug–disease association, we constructed features expressing its similarity to the closest known drug–disease association, using the scoring scheme.

D. S.Wishartet al [3] proposed a key feature that distinguishes DrugBank from other online drug resources is its extensive support for higher level database searching and selecting functions. In addition to standard data viewing and sorting features, DrugBank also offers a generic text search, a local BLAST search (SeqSearch), a higher level Boolean text search (TextQuery), a chemical structure search utility (ChemQuery) and a relational data extraction tool (Data Extractor). Each of these search utilities has a number of useful bioinformatics or cheminformatic applications, many of which were described in the first DrugBank publication. For the latest release of DrugBank, we have added a number of improvements to both the generic text search and ChemQuery. In particular, the generic text search has been enhanced so that users now have the option of clicking on check boxes to limit their search to a drug's common name, its synonyms/ brand names or all text fields. Because the vast majority of queries to DrugBank are related to drug names/synonyms, the default query always has these two boxes checked off.

J. Bowes et al [4] evaluated potential side effects about drugs which is important in rational drug design and development, as well as successful marketing. Binding of drugs to their on- and off-targets modifies the functions of these targets and therefore is believed to account for their efficacies as well as side effects. Traditionally, properties of a drug such as binding fingerprint and chemical structure are evaluated to anticipate side effects. Moreover, in vitro assays or phenotypic tests in model organisms may not be able to capture the same spectrum of side effects in human. Recently, an increasingly accepted view is that integrating biological networks would provide unique insights into understanding disease mechanisms and identifying novel drug targets. Network-based methods have been explored and successfully applied in finding disease-associated genes and inferring underlying molecular mechanisms. Similarly, phenotypic responses to drugs can be better rationalized by considering their overall effects in the context of molecular networks.

X. Wang, et.al [5] implemented the minimal number of targets, and deciding which targets to include, in an in vitro pharmacological profiling assay is an exercise in judgment and experience, and also depends on budgetary and technical constraints. In summary, in vitro pharmacological profiling is a valuable tool that can allow the early identification of off-target pharmacological interactions that could cause safety liabilities in the clinic, and this early identification of safety liabilities could improve decision-making by discovery project teams. The use of the minimal panel of targets recommended in this article might help to reduce safety-related attrition of drug molecules during drug discovery and development. Further precompetitive knowledge management of this data could lead to the development of in silicon tools that more accurately predict pharmacological activity and integration of these data with robust in vivo models could enable efficient and cost-effective early decision-making based on accurate predictions of the exposures at which a safety liability may be expected in the human population. We hope that this article is a first step towards establishing a broad initiative to work closely on improving drug safety from early stages of drug discovery through to clinical development and at the post-marketing stage.

III. DATA RETRIEVAL USING SUPERVISED LEARNING

Data mining refers to set of tools and strategies to explore information in an automatic method to extract semantically meaningful data. The retrieval process represents a query to the system and extracts the information based on the user request such mechanism referred to as query-by-example and it requires the definition of a data representation a fixed of descriptive functions and of some similarity metrics to examine question and goal photographs. Machine Learning (ML) can be considered as a subfield of Artificial Intelligence since those algorithms can be seen as building blocks to make computers learn to behave more intelligently by somehow generalizing rather than just storing and retrieving data items like a database system and other applications would do. Machine learning has got its inspiration from a variety of academic disciplines, including computer science, statistics, biology, and psychology. The core function of Machine learning attempts is to tell computers how to automatically find a good predictor based on past experiences and this job is done by good classifier. Classification is the process of using a model to predict unknown values (output variables), using a number of known values (input variables).The additional mechanisms have been introduced to achieve better performance and relevance results proved to be a powerful tool to iteratively collect information from the user and transform it into a semantic bias in the retrieval process. Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way (see inductive bias).The parallel task in human and animal psychology is often referred to as concept learning. Supervised learning is a type of machine learning algorithm that uses a known dataset (called the training dataset) to make predictions. The training dataset includes input data and response values. From it, the supervised learning algorithm seeks to build a model that can make predictions of the response values for a new dataset. A test dataset is often used to validate the model. Using larger training datasets often yield models with higher predictive power that can generalize well for new datasets. Supervised learning includes two categories of algorithms:

- **Classification**: for categorical response values, where the data can be separated into specific “classes”
- **Regression**: for continuous-response values

.

The steps of supervised learning as follows

- Prepare Data
- Choose an Algorithm
- Fit a Model
- Choose a Validation Method
- Examine Fit and Update Until Satisfied
- Use Fitted Model for Predictions

Supervised learning system only possible to the trained data. The user can’t be having knowledge about labeled data at the time of search query in drug prediction system.

IV. DISTANCE AND ITS SIGNIFICATIONS

Classification is an important data mining technique that has a wide range of applications in many areas like biology, medicine, image analysis and market research etc. which is the process of partitioning a set of objects into different subsets such that the data in each subset are similar to each other. In classification analysis Distance measure and learning algorithm plays an important role. In order to measure the similarity or regularity among the data-sets, distance metrics plays a very important role. It is necessary to identify, in what manner the data are interrelated, how various data dissimilar or similar with each other and what measures are considered for their comparison. The main purpose of metric calculation in specific problem is to obtain an appropriate distance /similarity function. Many distance measures have been proposed in classification algorithm. Some measurements are listed below in fig 2.

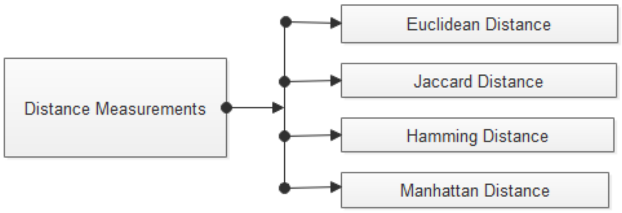


Fig 2. Distance Measurement Techniques

We can elaborate these distances in following sections.

Euclidean Distance

The Euclidean distance function measures the distance. The formula for this distance between a Point X (x_1, x_2, ….,x_n) and a Point Y (y_1, y_2,…, y_n) is specified in Equation(1):

$$\mathrm{Dist}(X,Y) = \sqrt{\sum_{i,j=1}^{m,n}(x_{ij} - y_{ij})^2}\text{------Eqn.(1)}$$

Developing the Euclidean distance between two data points involves computing the square root of the sum of the squares of the differences between corresponding values.

Jaccard Distance

The Jaccard coefficient, which is every other similarity measure, additionally known as the Tanimoto coefficient, is recycled to degree the comparison in the intersection divided by the union of the objects.. The Jaccard distance, which measures dissimilarity between sample groups, is matching to the Jaccard coefficient and is obtained by subtracting the Jaccard coefficient from 1 or equivalently by dividing the difference of the sizes of the union and the intersection of two sets by the size of the union can be defined Equation 2:

$$\mathrm{Dist}(X,Y) = 1 - J(X,Y)\text{------Eqn.(2)}$$

X, Y is the image points and J is Jaccard Co-efficient and specified in Equation 3.
Where as,

$$J(X,Y) = \frac{|(X\cup Y)| - |(X\cap Y)|}{|(X\cup Y)|}\text{----------Eqn.(3)}$$

For photo, the use of Jaccard coefficient is to make a assessment of the sum weight of shared terms and the sum weight of terms presented in either of the two documents but in condition that they are not the shared terms

Manhattan Distance

Manhattan distance is also named as image block distance because it is a distance the image pixel would drive in a image put out in square blocks like Manhattan with specified in Equation 4.

$$Dist(X,Y) = \|x_{ij} - y_{ij}\|\text{-----------Eqn.(4)}$$

X, Y are image points

Manhattan distance is also known as L1 distance. The distance between two points is the absolute difference between the points. Absolute value distance gives more robust result whereas Euclidean influenced by unusual values.

Hamming Distance

The Hamming distance which refers to difference between strings of equal period is the quantity of positions for which the corresponding symbols are different

$$\mathrm{Dist}(X,Y) = \sum_{i=1}^{n}|x_i - y_i|\text{----------Eqn.(5)}$$

$$x = y \rightarrow \mathrm{Dist} = 0$$

$$x \neq y \rightarrow \mathrm{Dist} = 1$$

X,Y are image points and specified in Equation 5.

The Hamming distance may be interpreted as the variety of bits which want to be changed (corrupted) to show one string into other. Sometimes the range of characters is used in location of the number of bits. Hamming distance can be visible as Manhattan distance among bit vectors. The merits and demerits of the distance measurement are shown in Table 1.

Table 1. Merits and Demerits of Distance measures

Distance measure	**Merits**	**Demerits**
Euclidean Distance	Flexible to support all data	Does not work on large datasets
Jaccard Distance	Reduce complexity in computation	Does not work for small values
Manhattan distance	Provide generalized model in distance measurements	Can't be implemented in large number of drugdatasets
Hamming Distance	Easy to calculate distance values	Only support labeled clustering

V. PROPOSED FRAMEWORK

Nowadays medicine consuming has become day to day activities for the people who were suffering from diseases. Hence many of the people are not aware of the medicine which is prescribed by doctors or pharmacies. Once they are affected by diseases they are approaching doctor and they are in taking the medicines prescribed by them, without having any knowledge about it and gets affected by its side effects. Due to the advancement in medical field many different approach are proposed about using medicines.

Many medicines are available for one disease in which some causes side effects. Some patients have a side effect in which the medicine causes same side effects which results in increase in disease. In this proposed system is that first we will cluster the drugs based on patient side effects, age, gender and do analysis in that and we will predict drugs based on side effects based on trained using semi-supervised learning approach. Semi-supervised learning is a class of supervised learning tasks and techniques that also make use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The acquisition of labeled data for a learning problem often requires a skilled human agent or a physical experiment. The cost associated with the labeling process thus may render a fully labeled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value. Semi-supervised learning is also of theoretical interest in machine learning and as a model for human learning. The proposed system makes full use of the unlabeled instance. This method goes with this assumption that the change of all the labels should be smooth on the graph.

The graph, showing the correlation of each pair of instances, is G=(V,E) where V is the vertex set and E is the edge set. Each vertex in V represents an instance in data set and each edge in E gives the similarity for each datum pair with a non-negative weight. Then GSSL tries to find a function F which satisfies two conditions :(1)the result F is closed to the original labels, and(2)all the labels should change very smoothly on the graph G. The first assumption can be expressed like this:

$$\varphi_1(F) = \frac{1}{2}\Sigma_{j=1}^{n}\Sigma_{j=1}^{m}\left(f_{ij} - y_{ij}\right) = \frac{1}{2}\|F - Y\| \text{----Eqn.(6)}$$

From Equation (6) where Y is the original label matrix. The second assumption can be expressed like this:

$$\varphi_2(F) = \frac{1}{2}\Sigma_{i,j=1}^{n}\left\|\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_i}{\sqrt{D_{jj}}}\right\|^2 W_{ij} = tr(F^T(I - D^{-\frac{1}{2}}W\,D^{-\frac{1}{2}})F) = tr(F^T LF) \quad \text{---------Eqn.(7)}$$

From Equation (7) where F=[f_1,f_2,....f_n] D is a diagonal matrix whose element is $D_{ii} = \sum_{j=1}^{n} W_{ij}$, I is identity matrix, L is Laplace matrix. Then the proposed model tries to minimize the sum of two parts showed above:

$$argmin_F\left(\frac{1}{2}\|F - Y\| + tr(F^T LF)\right) \quad \text{-------Eqn.(8)}$$

From Equation (8), extract the drugs details based unlabeled data matrix. For side effect target prediction problem we follow the assumption that similar drugs always interact with the same targets and similar targets always interact with the same drugs, so in our method we make the prediction from the two points of views: the drug view and the side effect view. We also find that the similarity information is key to prediction. The original way of measuring similarity of drugs based on side effects. Figure 3 shows drug query system based on side effects using semi-supervised learning approach.

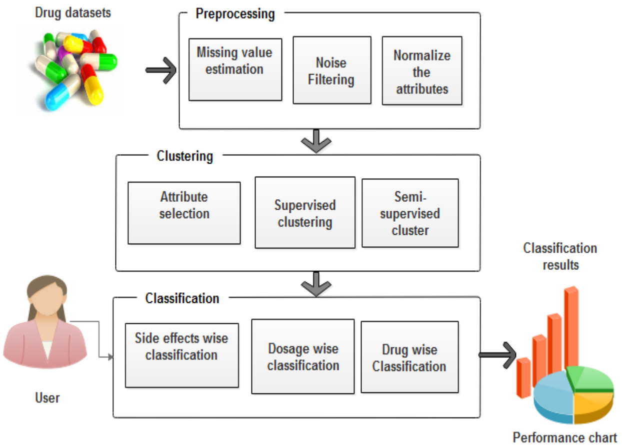
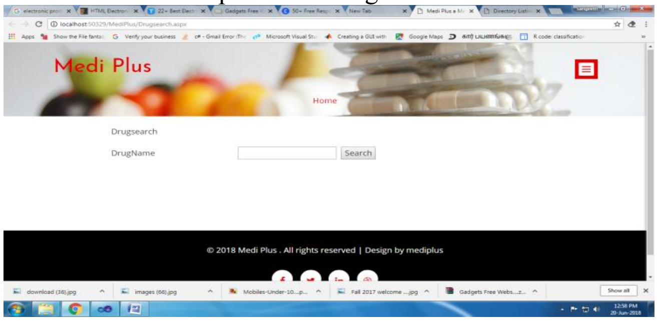


Fig 3. Proposed framework

VI. Experimental Results

The proposed work can be evaluated using C#.net as Front End and SQL SERVER as Back end. We can analyzed two types of learning such as supervised and semi-supervised learning. These results of proposed screen are shown in Fig 5 and Fig 6. The dataset of the drug details can be collected from website as "https://www.drugs.com/".


localhost:50329/MediPlus/Drugsearch.aspx

Medi Plus

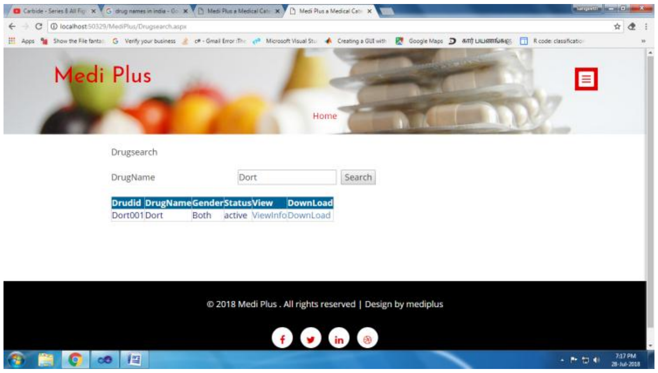
Home

Drugsearch

DrugName Search

© 2018 Medi Plus . All rights reserved | Design by mediplus

Fig 4: Home page



localhost:50329/MediPlus/Drugsearch.aspx

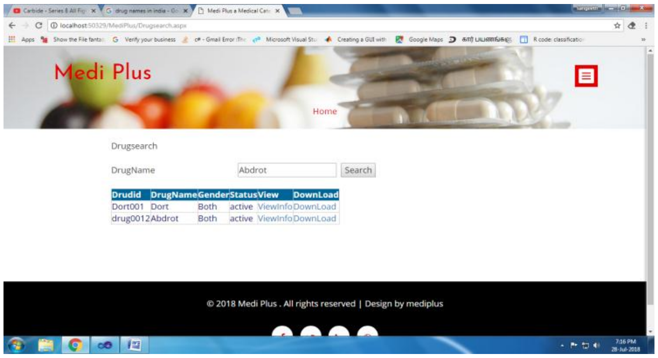
Drugsearch

DrugName Dort Search

Drudid	DrugName	Gender	Status	View	DownLoad
Dort001	Dort	Both	active	ViewInfo	DownLoad

© 2018 Medi Plus . All rights reserved | Design by mediplus

Fig 5: Supervised Result



localhost:50329/MediPlus/Drugsearch.aspx

Drugsearch

DrugName Abdrot Search

Drudid	DrugName	Gender	Status	View	DownLoad
Dort001	Dort	Both	active	ViewInfo	DownLoad
drug0012	Abdrot	Both	active	ViewInfo	DownLoad

© 2018 Medi Plus . All rights reserved | Design by mediplus

Fig 6: Semi-supervised Search Results

Based on the above figures, supervised and semi-supervised learning provide drug names for both labeled and unlabeled data with improved accuracy rate. There are a number of evaluation matrices are used to evaluate the retrieval performance. Here we used response time. Response time is calculated for both supervised and semi-supervised learning system. It can be shown in fig 7.

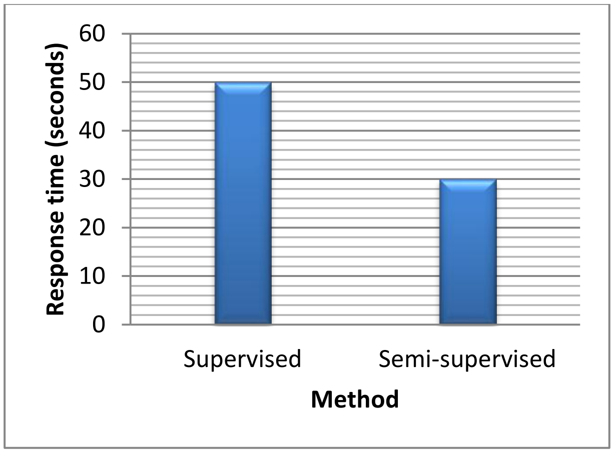


Fig 7: Performance report

And also analyze the performance of the system using various time periods, which can be shown in fig 8.

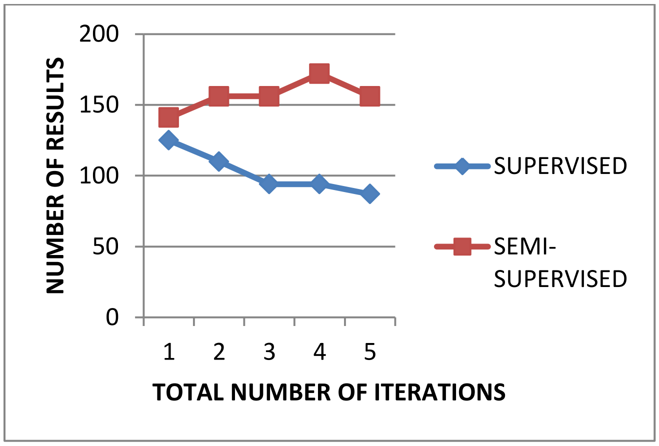


Fig 8: Retrieved Results

From the above graph, semi-supervised learning provide search results with reduced response time

For matrix evaluation, we are using precision and recall. Where, precision measure the availability of relevant answers from the retrieved answers in drug query system and recall measure the availability of relevant answers from the retrieved answers over the total number of relevant answers in the database.

$$\text{Precision} = \frac{\text{No of relevant answers extracted}}{\text{Total no of answers extracted}}$$

$$\text{Recall} = \frac{\text{No of relevant answers extracted}}{\text{Total no of answers in database}}$$

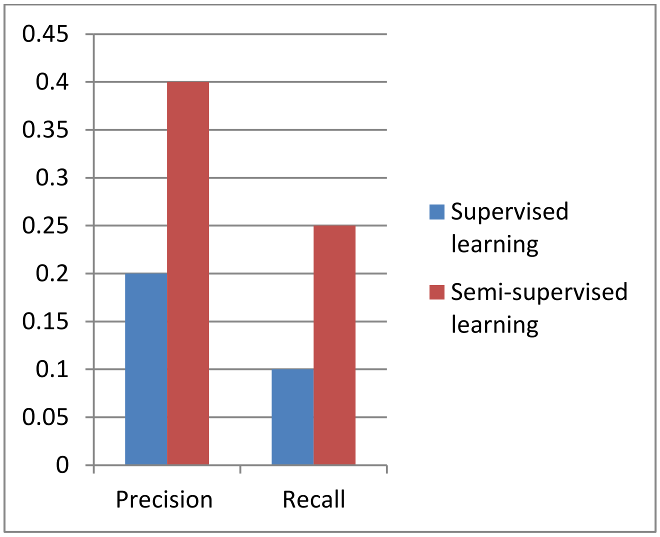


Fig 9: Precision and Recall measure

From the above graph, semi-supervised learning provide relevant answers higher than the supervised approach

VII. CONCLUSION AND FUTURE WORK

In this paper, supervised and semi-supervised learning can be described. The main purpose of this paper to analyze the performance of supervised and semi-supervised method in Drug Query system. The user can search the side effects which can be considered as labeled or unlabeled data. Then extract the similar drugs. And find that semi-supervised learning produced 40% improved response time in comparison with supervised method in drug retrieval system. The proposed drug query system needs large number of datasets to train in the database to retrieve relevant answers. In future, combine query system with Google query system for improvise the results.

REFERENCES

[1] E. Bressoet al., “Integrative relational machine-learningfor understanding drug side-effect profiles," BMC Bioinf.,vol. 14, Issue 2 Jun. 2013.

[2] T. Liu and R. B. Altman, ``Relating essential proteins to drug side effects using canonical component analysis: A structure-based approach," J. Chem. Inf. Model., vol. 55, no. 7, 2015.

[3] D. S.Wishartet al., ``DrugBank: A knowledgebase for drugs, drug actions and drug targets," Nucl. Acids Res., vol. 36, Issue 1, Nov. 2007.

[4] J. Bowes et al., ``Reducing safety-related drug attrition: The use of in vitro pharmacological profiling," Nature Rev. Drug Discovery, vol. 11, no. 12, 2012.

[5] X. Wang, B. Thijssen, and H. Yu, ``Target essentiality and centrality characterize drug side effects," PLoSComput. Biol., vol. 9, no. 7, p. 2013.

[6] M. Duran-Frigola and P. Aloy, ``Analysis of chemical and biological features yields mechanistic insights into drug side effects," Chem. Biol., vol. 20, no. 4, 2013.

[7] T. Liu and R. B. Altman, ``Relating essential proteins to drug side effects using canonical component analysis: A structure-based approach," J. Chem. Inf. Model., vol. 55, no. 7, 2015.

[8] S. Jamal, S. Goyal, A. Shanker, and A. Grover, ``Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models," Sci. Rep., vol. 7, Issue 2 Apr. 2017

[9] J. Scheiberet al., ``Mapping adverse drug reactions in chemical space," J. Med. Chem., vol. 52, no. 9, 2009.

[10] Y. Yamanishi, M. Kotera, M. Kanehisa, and S. Go to, ``Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework," Bioinformatics, vol. 26, no. 12, 2010.

[11] A. F. Fliri, W. T. Loging, P. F. Thadeio, and R. A. Volkmann, ``Analysis of drug-induced effect patterns to link structure and side effects of medicines," Nature Chem. Biol., vol. 1, no. 7, 2005.

[12] J. Scheiberet al., ``Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis," J. Chem. Inf. Model., vol. 49, no. 2, 2009.

[13] F. Wang, P.Zhang, N. Cao, J. Hu, and R. Sorrentino, ``Exploring the associations between drug side-effects and therapeutic indications,"J.Biomed.Inform.,vol. 51, Oct.2014.

[14] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, and Y. Yamanishi, ``Relating drug protein interaction network with drug sideeffects," Bioinformatics, vol. 28, no. 18, 2012.

[15] Y. Yamanishi, E. Pauwels, and M. Kotera, ``Drug side-effect prediction based on the integration of chemical and biological spaces," J. Chem. Inf.Model., vol. 52, no. 12, 2012.

[16] M. Liu et al., ``Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning," J. Amer. Med. Inform.Assoc., vol. 21, no. 2, 2014.

[17] F. Cheng et al., ``Adverse drug events: Database construction and in silicoprediction," J. Chem. Inf. Model., vol. 53, no. 4, pp. 744-752, 2013

[18] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, ``Predictingpotential side effects of drugs by recommender methods and ensemble learning," Neurocomputing, vol. 173, pp. 979-987, Jan. 2016.

[19] Y.-G. Chen, Y.-Y. Wang, and X.-M.Zhao, ``A survey on computational approaches to predicting adverse drug reactions," Current Topics Med.Chem., vol. 16, no. 30, 2016.

[20] D. P. Williams and B. K. Park, ``Idiosyncratic toxicity: The role of toxicophores and bioactivation," Drug Discovery Today, vol. 8, no. 22, pp. 1044-1050, 2003.

Authors Profile

V. Jagadeesan has completed his M.Sc., Computer Science degree at Bharathidasan University (CDE), Trichy. Currently he is doing M.Phil in computer Science at A.V.C College(Auto), Mayiladuthurai. He is doing research in the area of data mining.

Dr. K. Palanivel received his M.Sc. (Computer Science) degree from Bharathidasan University, M.Phil. (Computer Science) degree from Manonmaniam Sundaranar University and Ph.D. degree from Bharathidasan University. He is currently working as Associate Professor in the Department of Computer Science at AVC College (Autonomous), Mayiladuthurai. He has published many research papers in international journals. His research area includes Human Computer Interaction, Machine Learning, Recommender systems and Data mining.