International Journal of Innovative Research in Technology & Science

Received 6 April 2018, Received in revised form, 25 April 2018, Accepted 20 May 2018, Available online 25 May 2018

ISSN: 2321-1156 Volume VI Issue III, May 2018

Improved Performance Analysis Based on Self-Organizing Map and PAKS Using Health Care Data

Anushka Pandey, M. Tech. Scholar Department of CSE, TITS, R.G.P.V., Bhopal, M.P., India; pandey26anu@gmail.com;
Prof. Rajesh Nigam, A. P., Department of CSE, TITS, R.G.P.V., Bhopal, M.P., India; rajeshrewa37@gmail.com;

ABSTRACT

Improving performance analysis based on self-organizing map and proposed algorithm k-means with self-organizing map (PAKS) Using Health Care Data. The use of self-organizing map in neural network is very wide in data mining due to some feature like parallel performance, Self-organizing adaptive, robustness and fault tolerance. Data mining processing model based on task they achieve some process like association rules, clustering, prediction and classification. Neural network is used to find pattern in data. The grouping of self-organizing map based on neural network model and data mining method can greatly increase the efficiency data mining methods. In this process using general data, health care professionals, pharmaceutical companies, and medical specialty researchers and public health agencies to share, discuss, inquire, and report health care data exploitation early and innovative tools. Background support health care professionals and pharmaceutical corporations to face, observation and coverage the adverse events and proposed scheme can be more efficient than the common ground schemes for health care data recovery and identical time improve health care outcomes. Self-organizing map is a type of mathematical cluster analysis which particularly well suited for recognizing and classifying features in complex, multidimensional data. This paper proposes an improved Self-organizing map clustering algorithm which based on neighborhood mutual information correlation measure. Our propose approach PAKS is better as compare self-organizing map .because PAKS is minimize flat in dataset and improve accuracy of dataset.

Keywords - Data Mining Method ,Clustering, Clustering Algorithm, Neural network**,** Self-organizing Map, Health Care ,K-Mean Clustering, PAKS

I. INTRODUCTION

SOCIAL media, starting from personal electronic communication to measure for as, is providing limitless opportunities for patients to discuss their experiences with medicine and devices. It's conjointly providing limitless opportunities for corporations to receive feedback on their product and services [1]–[2]. Pharmaceutical corporations are already observing social network observation as a high priority inside their IT departments, doubtless making a chance for fast dissemination and feedback of product and services to optimize and enhance delivery, increase turnover and profit, and scale back prices [3]. Recently, methods for harvest home social media for bio-surveillance have conjointly been rumored [4]. Social media allows communication, collaboration, information collection, and sharing within the aid area. It thus provides a virtual social networking atmosphere. Associate in Nursing applicable way to extract information and trends from the knowledge "cloud" would be to model social media exploitation offered network modeling and machine tools (such as network-based analysis methods). underneath this paradigm, a social network (Face book, Twitter, WebMD, etc.) may be a structure manufactured from nodes (individuals or organizations) and edges that connect nodes in numerous relationships such as interests, friendship, kinship, etc. the foremost common thanks to represent the knowledge would be a graphical representation that's terribly convenient for visual image. Network modeling might provide Associate in nursing in-depth understanding of social network dynamics. A network model may be used for simulation studies of assorted network properties like understanding how users diffuse data among themselves (news concerning pandemic or drugs' adverse effects). Another example is finding out the improvement of bound edges of networks and how bound data affects the enhancements (e.g., how bound user communities evolve supported common interests about specific diseases).

The method of neural network is used for feature mining, pattern recognition, Clustering and classification. The model of neural network is dividing into three types such as feed-forward network, feedback network and self-organization network [5]. The huge amount of medical data is available but there is a lack of data analysis tool to extract useful knowledge from it. Unfortunately all doctors are not expert in all field of medical. Clinical results are often ready base on doctor's awareness and International Journal of Database Theory and Application, experience rather than on the

Received 6 April 2018, Received in revised form, 25 April 2018, Accepted 20 May 2018, Available online 25 May 2018

ISSN: 2321-1156 Volume VI Issue III, May 2018

knowledge masked in the database. Data mining have the ability to produce a knowledge-rich situation which can aid to improve the worth of clinical decision, due to this reason automatic medical diagnoses system is very useful by fetching all of them together. Genetic Algorithm is used to reduce the actual size of data which is enough for heart disease [6]

The process of analyzing the data in the form of pictures is called information visualization. This helps and supports decision making in numerous fields, including health-care surveys. Visualizing information from large amounts of heterogeneous survey data in order to find out interesting patterns is a difficult task, but by using data-mining techniques (clustering) coupled with artificial neural networks in the form of SOMs renders it tractable. In particular, clinicians often conduct surveys to better understand their patients. As mentioned earlier, using traditional descriptive statistical methods such as mean, variance, skewness and frequency, may lead to overly simplified conclusions. Hence, clinicians require statistical machine-learning tools that could be deployed as a 'black-box' for carrying out data analysis. For these reasons, we make use of the SOM algorithm for mining correlations and clustering similar responses within the surveys. The clustered responses in the higher dimensions are then visualized in a 2-dimensional grid thereby reducing the complexity within the data. Reducing the complexity in the data reveals more meaningful relationships, enabling understanding of the dependencies among the responses given in the survey. Previously, SOM has been used to visually explore data areas such as health, lifestyle, nutrition [7],

II. RELATED WORK

Courtney D. Corley et al. [8]. "Text and Structural Data Mining of Influenza Mentions in Web and Social Media, Text and structural data processing of internet and social media (WSM) provides a unique illness police work resource and might determine on-line communities for targeted public health communications (PHC) to assure wide dissemination of pertinent data. WSM that mention respiratory disorder are harvested over a 24-week.Theyhave a tendency to conjointly arouse bear a graph-based data processing technique to observe anomalies among respiratory illness blogs connected by publisher sort, links, and user-tags. Text and structural data processing of WSM provides a unique sickness police work resource and technique to spot on-line "flu" topic health data communities. They planned framework of complementary data-mining strategies supports the hypothesis. They have a tendency to comprehensively valuate journal posts containing respiratory disorder topic keywords through text, link, and structural data processing. Results from analysis show sturdy co-occurrence of grippe journal posts throughout the USA 2008-2009 grippe season. Frequency of grippe posts per blogger follows a heavy-tailed distribution, and that they show through graph metrics that the foremost prolific bloggers aren't the foremost authoritative. Pertinent health data ought to have a presence all told known WSM communities. The Girvan-Newman algorithmic program is leveraged to spot clusters of comparable sites as potential target communities for on-line health data campaigns. The result analysis shows that the distinct WSM communities are clustered by various publisher and content sort for e.g. News Corporation &amp, Walt Disney properties, international audiences, or personal blogs. Harvesting WSM may be a continued challenge with the explosive growth of net usage. To enrich the text mining approach to ILI observance, they have a tendency to apply a graph-based data processing technique, Subdue, to observe anomalies and informative substructures among grippe blogs connected by publisher sort, links, and user-tags. this method flags anomalies not discovered with content analysis that correspond to the United Kingdom's worst respiratory disorder season in eight years and therefore the emergence of sturdy personal journal communications throughout the U.S. seasonal respiratory disorder peak incidence. Link analysis reveals communities, clustered by content and in several cases company possession, to assure wide dissemination of pertinent data that ought to be targeted in a very productive public health communications campaign. Text mining of respiratory disorder mentions in WSM is shown to spot trends in grippe posts that correlate to real-world ILI patient reportage information.

Altug Akay et al. [9]. Recently from social media intelligently extracting data has attracted nice interest from the medical specialty and Health science community to at the same time improve tending outcomes and cut back prices exploitation consumer-generated opinion. They tend to propose a two-step analysis framework that focuses on positive and negative sentiment, moreover because the aspect effects of treatment, and identifies user communities (modules) and important users for the aim of ascertaining user opinion of cancer treatment. They tend to use a self-organizing map to investigate word frequency knowledge derived from users' forum posts. They tend to then

Received 6 April 2018, Received in revised form, 25 April 2018, Accepted 20 May 2018, Available online 25 May 2018

introduced a completely unique network-based approach for modeling users' forum interactions and used a network partitioning methodology supported optimizing a stability quality live. This allowed US to see client opinion and establish important users among the retrieved modules exploitation data derived from both word-frequency knowledge and network-based properties. Our approach will expand analysis into showing intelligence mining social media knowledge for client opinion of varied treatments to supply speedy, up-to-date data for the pharmaceutical business, hospitals, and medical workers, on the effectiveness (or ineffectiveness) of future treatments. They tend to born-again a forum targeted on medical specialty into weighted vectors to live client thoughts on the drug Erlotinib exploitation positive and negative terms aboard another list containing the aspect effects. Our strategies were ready to investigate positive and negative sentiment on carcinoma treatment exploitation the drug by mapping the big dimensional knowledge onto lower dimensional area exploitation the Som. Most of the user knowledge was clustered to the realm of the map coupled to positive sentiment, so reflective the final positive read of the users ulterior network primarily based modeling of the forum yielded fascinating insights on the underlying data exchange among users. Modules of powerfully interacting users were known employing a multi scale community detection methodology delineated. By overlaying these modules with content-based data within the sort of word-frequency scores retrieved from user posts, They tend to were ready to establish data brokers that appear to play necessary roles within the shaping the data content of the forum in addition, They tend to were ready to establish potential aspect effects systematically mentioned by teams of users. Such AN approach might be accustomed raise red flags in future clinical police investigation operations, moreover as highlight varied alternative treatment connected problems. The results have opened new potentialities into developing advanced solutions, moreover as revealing challenges in developing such solutions. The accord on Erlotinib depends on individual patient expertise. Social media, by its nature, can bring completely different completely different} people with different experiences and viewpoints. They tend to sifted through the information to search out positive and negative sentiment that was later confirmed by analysis that emerged concerning Erlotinib's effectiveness and aspect effects.

Xiaodong Feng et al. [10]. Identifying cancer risks related to meditative agents plays a crucial role in cancer management and hindrance. Case reports of cancers related to pharmacotherapy are escalating within the Food and Drug Administration Adverse Event coverage System (FAERS). The target of this study is to assess the chance of pancreatic cancer related to anti-diabetic medication of dipeptidyl peptidase four (DPP 4) inhibitors with or while not combination of anti diabetic drug. Using the FAERS public information, the adverse event reports (ADRs) related to wide used DPP 4 inhibitors with or while not combination of Glucophage were generated and evaluated. Standardized pharmacovigilance tools were applied to find the signal for cancer risks by calculative the proportional reporting ratio (PRR) and also the reporting odds ratio (ROR). Among 12618 ADRs related to sitagliptin from 2007 to 2011, there have been 223 cases of cancer. There was a big correlation between the cancer coverage magnitude relation and also the time (R=0.796, P&lt; 0.001). Pancreatic cancers accounted for twenty second of all combined cancer adverse events. Pharmacovigilance assessment from 2007 to 2012 indicated that there was a big risk of carcinoma related to DPP four inhibitors treatment (ROR=5.922). Curiously, negligible risk of carcinoma risk was related to Glucophage (ROR=1.214). Combination of DPP four matter sitagliptin with Glucophage correlates with considerably lower risk of carcinoma compared to sitagliptin treatment while not Glucophage (OR=0.277, 95%CI: 0.210-0.365). There was a big signal of carcinoma risk related to DPP four matter treatments. For the primary time they tend to incontestable that combination with Glucophage considerably reduced the chance signal of duct gland cancer related to DPP four inhibitors in FAERS. Considering the limitation of the FAERS, this study silent the potential strategy for cancer management and hindrance in diabetic patients, and provided directions for future clinical studies.

Juha Vesanto et al. [11]. The Self-Organizing Map (SOM) could be a vector quantization technique that places the image vectors on a regular low-dimensional grid in an ordered fashion. This makes the som a robust image tool. The SOM toolbox is an implementation of the som and its visualization within the Mat lab five computing surroundings. In this article, the som tool chest and its usage are shortly presented. Additionally its performance in terms of machine load is evaluated and compared to a corresponding Program. In this paper, the som toolbox has been shortly introduced. The som is a superb tool within the visualization of high dimensional knowledge. Intrinsically it's most suitable for knowledge understanding section of the knowledge discovery method, though it is often used for data preparation, modeling and classification furthermore. The analysis considers the quantitative analysis of som mappings, especially analysis of clusters and their properties. New functions and graphical program tools are going to be further to the Toolbox to extend its quality in data processing. Also outside contributions to the toolbox are welcome. The som toolbox promotes the utilization of som formula – in analysis furthermore as in industry – by creating its best options additional promptly accessible.

Cai-Hong Yun et al. [12]. Lung cancers caused by activating mutations within the epidemic growth factor receptor (EGFR) square measure at the start alert to little molecule tyrosine kinase inhibitors (TKIs), however the effectualness of those agents is commonly restricted due to the emergence of drug resistance conferred by a second mutation, T790M. Threonine 790 is that the ‘‘gatekeeper’’ residue, a vital determinant of matter specificity within the ATP binding pocket. The T790M mutation has been thought to cause resistance by sterically block binding of TKIs like gefitinib and erlotinib, however this rationalization is tough to reconcile with the very fact that it remains sensitive to structurally similar irreversible inhibitors. Here, they tend to show by employing a direct binding assay that T790M mutants retain low-nanomolar affinity for gefitinib. Moreover, they tend to show that the T790M mutation activates WT EGFR which introduction of the T790M mutation increases the ATP affinity of the oncogenic L858R mutant by additional than an order of magnitude. The increased ATP affinity is that the primary mechanism by that the T790M mutation confers drug resistance. Crystallographic analysis of the T790M mutant shows how it will adapt to accommodate tight binding of various inhibitors, including the irreversible matter HKI-272, and conjointly suggests a structural mechanism for chemical action activation. They tend to conclude that the T790M mutation could be a ‘‘generic’’ resistance mutation that may reduce the efficiency of any ATP-competitive enzyme matter and that irreversible inhibitors overcome this resistance merely through covalent binding, not as a results of an alternate binding mode.

Katherine Faust et. al. [13].This paper demonstrates limitations in utility of the triad census for studying similarities among native structural properties of social networks. A triad census compactly summarizes the native structure of a network using the frequencies of sixteen isomorphic categories of triads (sub-graphs of three nodes). The empirical base for this study may be an assortment of fifty one social networks menstruation totally different relative contents (friendship, advice, agonistic encounters, and victories in fights, dominance relations, and so on) among a range of species (humans, chimpanzees, hyenas, monkeys, ponies, cows, and variety of bird species). Results show that, in combination, similarities among triad censuses of those empirical networks are for the most part explained by nodal and two properties – the density of the network and distributions of mutual, asymmetric, and null dyads. These results prompt us that the vary of doable network-level properties is very forced by the dimensions and density of the network and caution ought to be taken in interpreting higher order structural properties after they are for the most part explained by native network options.

Erwan Le Martelot et al [14], many real systems can be represented as networks whose analysis can be very informative regarding the original system's organization. In the past decade community detection received a lot of attention and is now a very active field of research. Recently stability was introduced as a new measure for partition quality. This work investigates stability as an optimization criterion that exploits a Markov process view of networks to enable multi-scale community detection. Several heuristics and variations of an algorithm optimizing stability are presented as well as an application to overlapping communities. Experiments show that the method enables accurate multi-scale network analysis.

Altug Akay et al [15] a novel data mining method was developed to gauge the experience of the drug Sitagliptin (trade name Januvia) by patients with diabetes mellitus type 2. To this goal, we devised a two-step analysis framework. Initial exploratory analysis using self organizing maps was performed to determine structures based on user opinions among the forum posts. The results were a compilation of user’s clusters and their correlated (positive or negative) opinion of the drug. Subsequent modeling using network analysis methods was used to determine influential users among the forum members. These findings can open new avenues of research into rapid data collection, feedback, and analysis that can enable improved outcomes and solutions for public health and important feedback for the manufacturer.

III. SIMULATION AND RESULT ANALYSIS

(a) 1 MATLAB TOOL

MATLAB (2013a) is the high level language and interactive environment used by millions of engineers and scientists

Received 6 April 2018, Received in revised form, 25 April 2018, Accepted 20 May 2018, Available online 25 May 2018

worldwide. It lets the explore and visualize ideas and collaborate across different disciplines with signal and image processing, communication and computation of results. MATLAB (2013a) provides tools to acquire, analyze, and visualize data, enable you to get insight into your data in a division of the time it would take using spreadsheets or traditional programming languages. It can also document and share the results through plots and reports or as published MATLAB (2013a) code .MATLAB (2013a) (matrix laboratory) is a multi paradigm numerical computing situation and 4th generation programming language. It is developed by math work; MATLAB (2013a) allows matrix strategy, plotting of function and data, implementation of algorithm, construction of user interfaces with programs. MATLAB (2013a) is intended mainly for mathematical computing; an optional tool box uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. It is simulating on mat lab 7.8.0 and for this work we use Intel 1.4 GHz Machine. MATLAB (2013a) is a high-level technical compute language and interactive environment for algorithm development, data visualization, records analysis, and numeric computation Mat lab is a software program that allows you to do data manipulation and visualization, calculations, math and programming. It can be used to do very simple as well as very sophisticated tasks. Database, analysis, visualization, and algorithm development. You can perform efficient data retrieve enhancement. Many functions in the toolbox are multithreaded to take benefit of multicore and multiprocessor computers.

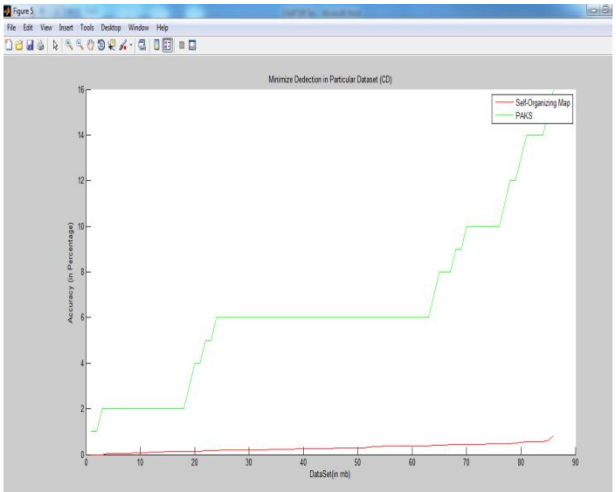

Figure1 MDA Analysis Base Dataset

(b) **Results Graph Analysis:**

(i) Minimize detection analysis using self-organizing map process and proposed process PAKS, but actually data recovery condition are not fulfilled. In this process detection of dataset point values at the time of self-organizing map process and PAKS case where x-axis shows data set and y-axis percentage of data copied minimize and finding accuracy . At the time of self-organizing map process proceed of false data recovery is more than the proposed approach PAKS process take the input data set as dataset or huge dataset so possibility of error is minimize.

(ii) The accuracy analysis shows that the with respect to time the correct information retrieval form the dataset. The graph here represents the percentage of the accuracy comparing the self-organizing map process and proposed process PAKS. The green line represents the proposed process PAKS and another blue line represents the self-organizing map process on traditional dataset. The graph shows with the purpose of PAKS has high accuracy as compared to the normal existing method because the PAKS data set has less redundancy.

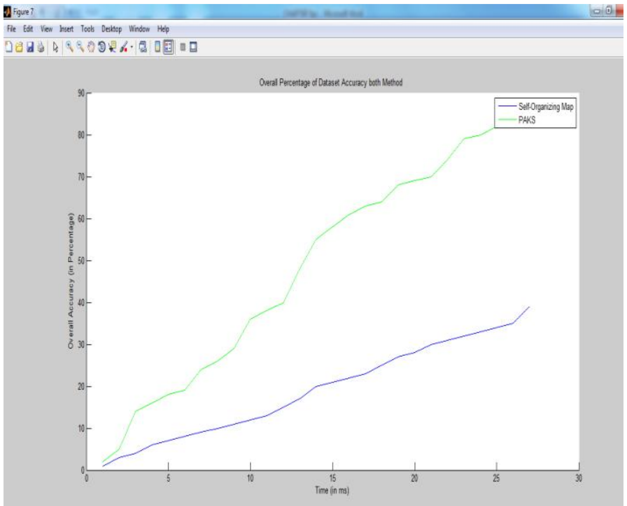

Figure2 Overall Accuracy in Dataset Analysis

VI. CONCLUSION

The proposed workflow of analyzing healthcare dataset and overcome the limitations of large scale healthcare dataset analysis. This work can help the users to be updated with the effectiveness of the medicines and it can even suggest them with few better medications available. This work also provides feedback to the healthcare system organization and pharmaceutical companies for the available treatments and medicines. Specific algorithms will perform network-clustering, one amongst the elemental tasks in network analysis. Finding a community during a healthcare dataset analysis and similar finding social network means those unique nodes that move with every other a lot of oftentimes

Received 6 April 2018, Received in revised form, 25 April 2018, Accepted 20 May 2018, Available online 25 May 2018

ISSN: 2321-1156 Volume VI Issue III, May 2018

than element nodes and optimize the clusters. Community detection will facilitate the extraction of valuable data for the complete aid trade. The pharmaceutical trade could benefit from this for higher targeting their promoting use up. Healthcare dataset may be higher identifying the point in clustering of satisfaction and minimizing identifying the points. Physicians may collect necessary indicator and different doctors and patients that may facilitate them in their treatment recommendations and optimize data analysis of the treatment and show results. Finally, patients might value and control different consumers data before creating better-informed result. Clustering and classification are key tasks of usefully data identification. In good quality of the continuous attributes, our proposed feature clustering algorithm for usefully data selection can directly deal with the continuous data and acquire high accuracy. Experimental results show that this algorithm outperforms than other approach. In the recent research, the cluster configuration is studied by using qualitative analysis. In order to get through understanding about gene expression profiles, and extend our model to improve its generation.

REFERENCES

[1]. A. Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. “Artificial societies and social simulation using ant colony, particle swarm optimization and cultural algorithms,” New Achievements in Evolutionary Computation, P. Korosec, Ed. Rijeka, Croatia: InTech, pp. 267–297, 2010. .

[2]. L. Toldo, “Text mining fundamentals for business analytics,” presented at the 11th Annual Text and Social Analytics Summit, Boston, MA, USA, 2013.

[3]. L. Dunbrack, “Pharma 2.0 – social media and pharmaceutical sales and marketing,” in Proc. Health Ind. Insights, 2010, p. 7.

[4]. C. Corley, D. Cook, A. Mikler, and K. Singh, “Text and structural data mining of influenza mentions in web and social media,” Int. J. Environ. Res. Public Health, vol. 7, pp. 596–615, Feb. 2010.

[5]. P. Gaur, “Neural Networks in Data Mining”, International Journal of Electronics and Computer Science Engineering (IJECSE, ISSN: 2277-1956), vol. 1, (2012), pp. 1449-1453.

[6]. J. Soni, U. Ansari, D. Sharma and S. Soni, “Predictive data mining for medical diagnosis: An overview of heart disease prediction”, International Journal of Computer Applications, vol. 17, 2011, pp. 43-48.

[7]. Y. Mehmood, M. Abbas, X. Chen, and T. Honkela, “Self-organizing maps of nutrition, lifestyle and health situation in the world,” in Advances in Self-Organizing Maps. Springer, 2011, pp. 160–167

[8]. Courtney D. Corley, Diane J. Cook, Armin R. Mikler and Karan P. Singh, “Text and Structural Data Mining of Influenza Mentions in Web and Social Media”, International Journal of Environmental Research and Public Health, ISSN: 1660-4601, Pp. 596-615, 2010.

[9]. Altug Akay, Andrei Dragomir, Bjorn-Erik Erlandsson, “Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care”, IEEE Journal of Biomedical And Health Informatics, Vol. 19, No. 1, January 2015.

[10]. Xiaodong Feng, Amie Cai, Kevin Dong, Wendy Chaing, Max Feng, Nilesh S Bhutada, John Inciardi, Tibebe Woldemariam Feng, “Assessing Pancreatic Cancer Risk Associated with Dipeptidyl Peptidase 4 Inhibitors: Data Mining of FDA Adverse Event Reporting System (FAERS)”, J Pharmacovigilance 2013, http://dx.doi.org/10.4172/2329-6887.1000110.

[11]. Juha Vesanto, Johan Himberg, Esa Alhoniemi and Juha Parhankangas, “Self-organizing map in Matlab: The SOM Toolbox”, Proceedings of the Matlab DSP Conference 1999, Espoo, Finland, November 16–17, pp. 35–40, 1999.

[12]. Cai-Hong Yun, Kristen E. Mengwasser, Angela V. Toms, Michele S. Woo, Heidi Greulich, Kwok-Kin Wong, Matthew Meyerson, Michael J. Eck, “The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP”, Pp. 2070–2075, PNAS, February 12, 2008, vol. 105 no.6,www.pnas.org_cgi_doi_10.1073..

[13]. Katherine Faust, Metodoloski Zvezki, “Comparing Social Networks: Size, Density, and Local Structure”, Vol. 3, No. 2, 2006, 185-216.

[14]. Erwan Le Martelot, “Multi-Scale Community Detection using Stability Optimization”.

[15]. S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. New York, NY, USA: Cambridge University Press, 1994, pp. 825.