

Measuring Similarity by Prediction Class between Biomedical Datasets via Fuzzy Unordered Rule Induction

Simon Fong¹, Osama Mohammed², Jinan Fiaidhi³, Sabah Mohammed³
and Chee Keong Kwoh⁴

¹*Department of Computer and Information Science
University of Macau, Macau SAR*

²*Department of Computer Science, University of Victoria, BC, Canada*

³*Department of Computer Science*

Lakehead University, Thunder Bay, Canada

⁴*Biomedical Engineering Research Centre (BMERC)*

Nanyang Technological University, Republic of Singapore

*ccfong@umac.mo, smohamme@uvic.ca, sabah.mohammed@lakeheadu.ca,
jfiadhi@lakeheadu.ca, ASCKKWOH@ntu.edu.sg*

Abstract

The need of similarity measures in life science is ever paramount given the modern biotechnology in producing and storing biomedical datasets in large amounts. This paper presents a novel scheme in measuring similarity of two datasets by prediction class, namely SPC. SPC offers an alternative approach to traditionally used ones such as pairwise correlations which assume every attribute carries equal importance. The unique advantage of SPC is the use of a machine learning model called Fuzzy Unordered Rule Induction to infer the similarity between two datasets based on their common attributes and their degrees of relevance pertaining to a predicted class. The method is demonstrated by a case of comparing lung cancer dataset and heart disease dataset.

Keywords: *Similarity measure; Classification; Fuzzy Unordered Rule Induction; Data Analytics*

1. Introduction

Modern technological advances in life science instruments have given rise to easy and rapid production of biomedical data. The biomedical datasets however may come in different variables and sizes due to a wide variety of sources. Many biomedical datasets in heterogeneous formats are hence produced and electronically stored in distributed locations. For example, biomedical researchers work with many different types of image and signal data, as well as electronic patients' records that cover diagnostic and prognosis data from health-care institutes all over the world. This makes biomedical analytics difficult because so far there is no universal similarity measuring method agreed to be best over heterogeneous data structures. Nevertheless researchers from both computer science communities and biomedical research communities formulated methods for measuring similarities between medical data items over the years. Their methods however vary greatly in theory based on different computational principles.

One of the fundamental criteria for the characterization of similarities between measured biomedical data items is correlation. It is popularly used in ranking attributes that describe the

dataset with relations between the attributes and the predicated class; and also in computing the correlations between attributes owned by different data objects in order to infer the similarity. For instance, Strickert, *et al.*, [1] extended the formal derivative of Pearson correlation for gradient-based optimization of data models, by rating the individual data attributes according to their impact on pairwise data relationships. Then the high-dimensional space is scaled by maximizing the correlation between distances of static source data and adaptive target vectors. The method was tested successfully on mass spectroscopy data.

Analogous to Pearson correlation, the variance measure in hyperspace such as Euclidean Distance, Minkowski Distance and the like, is favoured by another group of researchers. In general, they refer this type of similarity measure as clustering-based similarity models [2, 3]. An affinity coefficient [4] is computed by measuring how close each pairs of attributes are apart via clustering on multivariate data analysis. Recently, Chanchala, *et al.*, [5] proposed a similarity measure that based on the multivariate hypergeometric distribution for the pairwise comparison of images and data vectors. Testing on large-scale biological sample datasets are enabled by their method of piecewise approximation, such as mass spectrometry imaging data and gene expression microarray data.

Along with the concept of pairwise comparison between the data features, a number of similar techniques have been studied but for specific applications such as measure of contextual similarity for biomedical terms [6] via Edit Distance for approximate string matching, biomedical image retrieval via case-based reasoning over image features [7], measuring the relatedness and similarity of biomedical concepts [8] or reports [9] by dictionary-like Unified Medical Language System and some ontology standards respectively; just to name a few.

In this paper we present an alternative similarity measure called Similarity by Prediction Class (SPC). The method is extended from our previous work [10] called Dependency Network that displays out all the attributes and their respective predictive strengths to a disease, also inter-relations between symptoms across different diseases can be inferred. Using functions feature selection and information gain in inducing a predictive model, a Dependency Network is built by assigning the attributes of some disease significance values. The Dependency Network has the advantage of loading multiple medical history datasets so that dependencies can be traced across multiple diseases. This feature is useful for factors exploration especially those that were not previously known. The implicit link could be traced across a chain of diseases provided that they have common attributes in the forms of factors and symptoms in a sense of causality by investigating their relationships towards some related diseases. The underlying logic is a set of formula for quantitatively deriving a relational measure for this indirect dependency across diseases. Readers may find the mathematical definitions in [10]. One technical challenge in implementing a Dependency Network is the need of merging two or more medical datasets that have different dimensions in columns and rows though they may share some common attributes. This is known as schema matching which is a classical problem in information integration. A number of automated methods have been attempted in the past [11], such as matching the missing values by textual similarity, guessing from the mean values, by most frequently appearing numbers and so on. For the sake of obtaining the highest possible accuracy however, in medical data analysis, we resort to the most accurate yet computational-intensive method by building a RIPPER decision tree [12] for estimating the blank values. As long as the two medical datasets have sufficient amount of common attributes and the attributes have fairly good predictive powers to the diseases, the decision-tree-per-missing-attribute method works satisfactorily. A pioneer work on applying decision-trees for estimating missing values demonstrated its feasibility [13].

While Dependency Network was designed for informatics visualization, SPC improves its precursor by qualitatively and quantitatively inferring which are the most relevant common attributes and the similarity measure in numbers, between a pair of medical datasets in question. When studying diseases and genetic disorders, researchers would like to quantify the amount of presence of a condition as well as comparing a particular dataset or sample to a reference dataset. So the objective is mainly classification. For example, cases of a new disease surfaced, which is possibly mutated, unseen and undiscovered before. Not only is it interested to find which other disease it is related to by their common properties, but to know in quantitative value how similar they are. The major difference between our proposed SPC model and the existing ones including Dependency Network, is that SPC is measured based on common attributes and their specific relevance to the predicted class in concern. Instead of measuring similarity merely by pairwise correlations assuming every attribute is equally important (which is clearly not true in medical contexts), SPC infers the predictive powers (in terms of relevance) of the commonly owned attributes of two datasets with respect to the predicted class. Hence a more accurate similarity measure can possibly be achieved. Taking a layman analogous example, two persons can be related by screening their bodily and demographics attributes which they own in common; person A and person B may be similar in terms of body built for the classification of hobby – sports-type, A is a soccer player and B is a hockey player. However, they are dis-similar once the classification is changed to, let's say – profession, A is an accountant and B is a bus driver. The overlapped attributes of the two persons that sustain strong predictive power to the class of 'hobby' could be body height, physical fitness features, etc. But in the context of classification of 'profession', these physical attributes may become less relevant compared to other attributes like education background and work experiences.

2. SPC Method

Three computational steps are involved in estimating a result from the SPC model: Step (1) Data-preprocessing – merge two datasets into a combined dataset which contains a meta-prediction class. Step (2) Classification Model Induction – from the combined dataset that contains training records from the two datasets in comparison, mapping to categories of meta-class, induce a classification model by supervised learning via Fuzzy Unordered Rule Induction algorithm (FURIA). Step (3) Rule Analysis – from the induced rules, calculate the similarity and the respective performance indicators.

2.1. Data-preprocessing

The medical dataset is a matrix of instances collected from the historical records or specimen samples. Each instance (row) is characterized by different attributes or features (columns) and it comes with a priori known label called class (in the last column). It is assumed that the classes of two datasets would have a common ontology which leads to a meta-class. For example, a lung cancer dataset may contain a collection of patients' records, each of which are characterized by attributes such as age, gender, whether he is a smoker, how many cigarettes are smoked daily, how long he has been a smoker, diet habits etc. Each record has a numeric class value having 0 means he is free of the disease, and other positive integers indicate different stages of lung cancer. Likewise for the other dataset heart disease with which the lung cancer dataset compares, it has attributes describing the conditions of the patient in each record and a class label of 0 meaning normal and other values representing different types of heart diseases. Over these two datasets, a meta-class could be simply

binary, 1-sick and 0-not sick. The two datasets, which come in different numbers of instances (n, j) and attributes (m, k) may take the form:

$$A = \begin{Bmatrix} i_{1,1} & i_{1,2} & \dots & i_{1,m} \\ i_{2,1} & i_{2,2} & \vdots & i_{2,m} \\ \vdots & \vdots & \dots & \vdots \\ i_{n,1} & i_{n,2} & \dots & i_{n,m} \end{Bmatrix} \text{ and } B = \begin{Bmatrix} i_{1,1} & i_{1,2} & \dots & i_{1,k} \\ i_{2,1} & i_{2,2} & \vdots & i_{2,k} \\ \vdots & \vdots & \dots & \vdots \\ i_{j,1} & i_{j,2} & \dots & i_{j,k} \end{Bmatrix} \quad (1)$$

Taking into account that the class for $A(c_a)$ and class for $B(c_b)$ have a meta-class, and data are combined according to the meta-class, the transformed datasets are:

$$A' = \begin{Bmatrix} i_{a,1,1} & i_{a,1,2} & \dots & i_{a,1,m} & i_{b,1,1} & i_{b,1,2} & \dots & i_{b,1,k} & i_{ca} \\ i_{a,2,1} & i_{a,2,2} & \dots & i_{a,2,m} & i_{b,2,1} & i_{b,2,2} & \dots & i_{b,2,k} & i_{ca} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ i_{a,n,1} & i_{a,n,2} & \dots & i_{a,n,m} & i_{b,n,1} & i_{b,n,2} & \dots & i_{b,j,k} & i_{ca} \end{Bmatrix} \quad (2a)$$

$$B' = \begin{Bmatrix} i_{a,1,1} & i_{a,1,2} & \dots & i_{a,1,m} & i_{b,1,1} & i_{b,1,2} & \dots & i_{b,1,k} & i_{cb} \\ i_{a,2,1} & i_{a,2,2} & \dots & i_{a,2,m} & i_{b,2,1} & i_{b,2,2} & \dots & i_{b,2,k} & i_{cb} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ i_{a,n,1} & i_{a,n,2} & \dots & i_{a,n,m} & i_{b,n,1} & i_{b,n,2} & \dots & i_{b,j,k} & i_{cb} \end{Bmatrix} \quad (2b)$$

2.2. Classification Model Induction

The rationale behind the use of FURIA is the fact that the instances after combining the two datasets are unordered and there may have existed some extent of uncertainty in persevering the non-linear relations between the attribute values and the meta-class. By now the attribute values should have been normalized into numeric z-scores for consistency. The FURIA is based on RIPPER [12] (Pruning to Produce Error Reduction) by William Cohen of AT&T Laboratories which is used here because of its high accuracy in predicate-logics rules generation, the information about the information gain for each attribute would be used for inferring a collection of links each represents the predictive power (in term of information gain) towards the prediction class. A selector constraining a numerical attribute A_i (with domain $D_i = R$) in a RIPPER rule can obviously be expressed in the form ($A_i \in I$), where $I \subseteq R$ is an interval: $I = (-\infty, v]$ if the rule contains a selector ($A_i \leq v$), $I = [u, \infty)$ if it contains a selector ($A_i \geq u$), and $I = [u, v]$ if it contains both (in the last case, two selectors are combined). Essentially, a fuzzy rule is obtained through replacing intervals by fuzzy intervals, namely fuzzy sets with trapezoidal membership function. A fuzzy interval of that kind is specified by four parameters and will be written $I^F = (\phi^{s,L}, \phi^{c,L}, \phi^{c,U}, \phi^{s,U})$:

$$I^F(v) \stackrel{\text{def}}{=} \begin{cases} 1 & \phi^{c,L} \leq v \leq \phi^{c,U} \\ \frac{v - \phi^{s,L}}{\phi^{c,L} - \phi^{s,L}} \phi^{s,L} & \phi^{s,L} < v < \phi^{c,L} \\ \frac{\phi^{s,U} - v}{\phi^{s,U} - \phi^{c,U}} \phi^{c,U} & \phi^{c,U} < v < \phi^{s,U} \\ 0 & \text{else} \end{cases} \quad (3)$$

$\phi^{c,L}$ and $\phi^{c,U}$ are the lower and upper bound of the core (elements with membership 1) of the fuzzy set respectively, likewise, $\phi^{s,U}$ and $\phi^{s,L}$ represent the lower and upper bound of the support. To obtain fuzzy rules, the idea is to “fuzzify” (to make something fuzzy) the final rules from our modified RIPPER algorithm. The purpose is to search for the best fuzzy extension of each rule, where a fuzzy extension is understood as a rule of the same structure, everything is the same but the intervals is replaced by fuzzy intervals. For the fuzzification of the antecedent ($A_i \in I_i$) it is important to consider only the relevant data D_T^i , i.e., to ignore those instances that are excluded by any other antecedent ($A_j \in I_j^F$), $j \neq i$:

$$D_T^i = \{x | (x_1 \dots x_k) \in D_T | I_j^F(x_j) > 0 \text{ for all } j \neq i\} \subseteq D_T \quad (4)$$

And split the instances D_T^i into two subsets, the positive instances D_{T+}^i and the negative instances D_{T-}^i . Then to measure the quality of the fuzzication, the rule purity will be used: $Pur = p_i / (p_i + n_i)$, P_i means positive instances, and the N_i means negative instances.

$$p_i \stackrel{\text{def}}{=} \sum_{x \in D_{T+}^i} \mu_{A_i}(x), \text{ and } n_i \stackrel{\text{def}}{=} \sum_{x \in D_{T-}^i} \mu_{A_i}(x) \quad (5)$$

How to make the classifier output, we need a factor to judge, the factor is called certainty factor. Suppose the fuzzy rules $r_1^{(j)} \dots r_k^{(j)}$ have been learned for the class λ_j . For a new query instance x , the support of this class is define by

$$S_j(x) \stackrel{\text{def}}{=} \sum_{i=1 \dots k} \mu_{r_i^{(j)}}(x). CF(r_i^{(j)}) \quad (6)$$

where $CF(r_i^{(j)})$ is the certainty factor of the rule $r_i^{(j)}$. It is defined as:

$$CF(r_i^{(j)}) = \frac{2 \frac{|D_T^{(j)}|}{|D_T|} + \sum_{x \in D_T^{(j)}} \mu_{r_i^{(j)}}(x)}{2 + \sum_{x \in D_T} \mu_{r_i^{(j)}}(x)} \quad (7)$$

where $D_T^{(j)}$ denotes the subsets of training instances with label λ_j . The class predicted by FURIA is the one with maximal support. In the case where x is not covered by any rule, which means that $s_j(x)=0$ for all classes λ_j , a classification decision is derived in a separate way.

2.3. Rule Analysis

During the process of rule induction by FURIA, the entropy for each class of instances is calculated.

$$Entropy(class) = -p_+ \log_2 p_+ - p_- \log_2 p_- \quad (8)$$

In the combined dataset where the meta-class is binary, there are r positive class and s negative class, the equation will be:

$$Entropy(r_+, s_-) = -r/n \log_2 r/n - s/n \log_2 s/n \quad (9a)$$

$$Entropy(u_+, w_-) = -u/n \log_2 u/n - w/n \log_2 w/n \quad (9b)$$

The gain information is therefore:

$$Gain(class, x) = Entropy(class) - \sum_{v \in Values(x)} \frac{|Class_v|}{|class|} Entropy(Class_v) \quad (10)$$

Assume there is x number of overlapped attributes among the two datasets in comparison, such that $\{1..x\}$ attributes from which we can infer the predictive strengths with respective to the meta-class.

$$S = [r_+, s_-], \text{ and } r+s=n.$$

$$Gain(class, x) = Entropy(class) - \sum_{v \in [Positive, Negative]} \frac{|Class_v|}{|class|} Entropy(Class_v) \quad (11)$$

We can index the occurrence of the respective attributes in the rules, and produce a counting list as follow:

$$D\text{-class } A = \begin{pmatrix} Index_{a1} & Class_{a1} \\ Index_{a2} & Class_{a2} \\ \vdots & \vdots \\ Index_{an} & Class_{an} \end{pmatrix} \quad (12a)$$

$$D\text{-class } B = \begin{pmatrix} Index_{b1} & Class_{b1} \\ Index_{b2} & Class_{b2} \\ \vdots & \vdots \\ Index_{bn} & Class_{bn} \end{pmatrix} \quad (12b)$$

where $Index_{a/bn}$ means the index number for the instances that originally come from dataset A and dataset B respectively, and for the result

$$Index_{an} = Index_{bn} \wedge Class_{an} = Class_{bn} = 1 \text{ (Assume we only concern about } c=1, \text{ sick)} \quad (13)$$

The common instances with the same index and classification result K will be:

$$K = D\text{-class } A \cap D\text{-class } B = \begin{pmatrix} Index_1 & Class = 1 \\ Index_2 & Class = 1 \\ \vdots & \vdots \\ Index_k & Class = 1 \end{pmatrix} \quad (14)$$

where k means the number of how many instance are there the two datasets overlapped in classification result. And the similarity of the two datasets is therefore:

$$Similarity(A, B) = Ratio_K = k/n \quad (15)$$

In general, the result of SPC would be presented as a tuple, as follow

$$Similarity(A, B) = \{S | CF | Acc | Conf\} \quad (16)$$

where S is computed from Eqn. 15, CF is confidence factors averaged out from all the overlapped attributes involved as by Eqn. 7, Acc is the accuracy of the FURIA induced model in terms of the number of correctly classified instances over the total number of instances, and $Conf$ is confidence indicator derived from the amount of overlapped attributes over the total number of attributes of the dataset.

$$Conf = overlap(i_a, i_b) = \frac{|i_a \cap i_b|}{\max(|i_a|, |i_b|)} \quad (17)$$

3. Experiment

In order to demonstrate the efficacy of the SPC method proposed here, two empirical datasets, available for download from UCI repository [14] are used. Lung cancer is released in May 1992 by Hong, Z.Q. and Yang, J.Y. This data was used by Hong and Young to illustrate the power of the optimal discriminate plane. The data described 3 types of pathological lung cancers. This date is published in July 1988 in UCI. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1, 2, 3, 4) from absence (value 0). The characteristics of the two datasets in comparison are shown in Table 1.

Table 1. Number of attributes and instances for the two datasets

NAME	INSTANCE#	ATTRIBUTE#	CLASS#
Lung Cancer	800	42	7
Heart Disease	200	48	2

In order to demonstrate the efficacy of the SPC method proposed here, two empirical datasets, available for download from UCI repository [14] are used. Lung cancer is released in May 1992 by Hong, Z.Q. and Yang, J.Y. This data was used by Hong and Young to illustrate the power of the optimal discriminate plane. The data described 3 types of pathological lung cancers. This date is published in July 1988 in UCI. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1, 2, 3, 4) from absence (value 0). The characteristics of the two datasets in comparison are shown in Table 1.

The two datasets are subject to SPC measure by following the steps in Section 2. For the sake of objective comparison with existing methods, the following two methods are used – Pearson correlation as reported in [10] by using Dependency Network, and Mutual Information Score [15] which is a measure of mutual dependence of two sets of variables. The formula for Mutual Information as reported in [16] is defined as follow:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right)$$

where $p(x,y)$ is the joint probability distribution function of X and Y , and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. The results of the different similarity measures by the three methods, SPC, Correlation and Mutual Information are presented in Table 2. The results have been normalized in [0..1] for easy comparison. In SPC, the meta-class is ‘illness’ that abstracts all the non-zero class values.

Table 2. Results of various similarity measures by SPC, Correlation and Mutual Information

Similarity Measure	SPC	Correlation	Mutual Information
Lung Cancer & Heart Disease	0.3902	0.1037	0.1992

Interestingly it can be seen that SPC has the highest value among the three similarity measure methods. It indicates a phenomenon that by taking account of the predicated class we can obtain more a precise similarity measurement than going through all the attributes in pairwise fashion (as in correlation and mutual information). The dependence graphs that show all the attributes of the two diseases and only the significant overlapped attributes are depicted in Figure 1 and Figure 2 respectively.

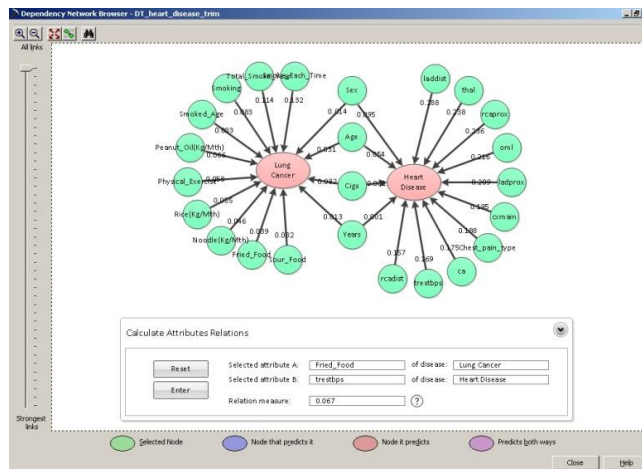


Figure 1. Dependency graph that shows all the attributes

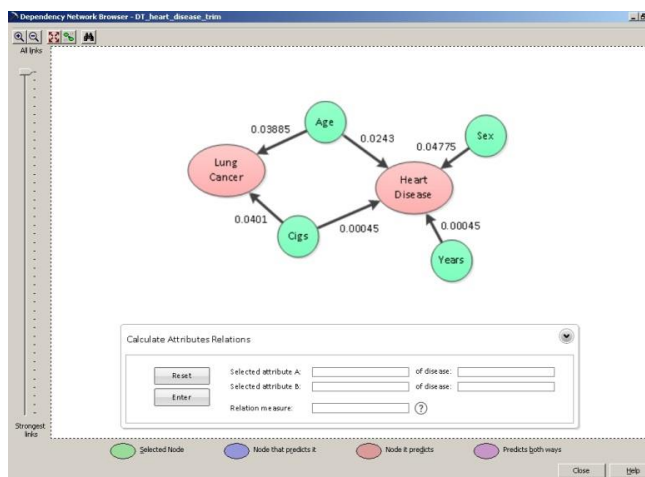


Figure 2. Dependency graph that shows only the significant overlapped attributes

4. Conclusions

Similarity measure is an important data exploratory tool in life. In the past many researchers attempted to provide computational methods in measuring how close/different two biomedical data are. Existing methods include statistical means like correlation, frequency counting, clustering and other sophisticated data mining techniques. In this paper we present an alternative similarity measure method called Similarity by Predicted Class (SPC) which can more precisely infer a quantitative measure between two different datasets by referring to their common predicated class (*e.g.*, illness, fatality, maliciousness, *etc.*). SPC works completely different from pairwise computing methods as SPC depends on the predictive power of the overlapped attributes pertaining to a common predicted class of the two datasets. In particular, fuzzy unordered rule induction algorithm is used in the inference process that offers fuzziness of the relevance of the attributes to the predicted class, and it is extended from a known classifier called RIPPER which offers relatively high accuracy. Our new method is demonstrated by a case of comparing lung cancer dataset and heart disease dataset. Experiments are conducted using UCI medical datasets and the results demonstrate that the methodology and the technical tasks within are possible. This methodology is believed to be useful for medical experts who want to investigate dependencies of attributes among a wide variety of medical data and it has positive implication to medical applications such as information retrieval, medical document matching and categorizing biomedical data, *etc.*

Acknowledgements

The authors would like to express gratitude to the funding support of the project with code MYRG073(Y1-L2)-FST12-FCC by the University of Macau. Thanks also go to the MSc graduate student Mr. Luke Lu who contributed to the programming of software and conducting the experiment.

References

- [1] M. Strickert, F. -M. Schleif, U. Seiffert and T. Villmann, "Derivatives of Pearson Correlation for Gradient-based Analysis of Biomedical Data", *Inteligencia Artificial*, vol. 12, no. 37, (2008), pp. 37-44.
- [2] M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, "Cluster analysis and display of genome-wide expression patterns", *PNAS*, vol. 95, no. 25, (1998), pp. 14863-14868.
- [3] H. Bacelar-Nicolau, F. Nicolau, A. Sousa and L. Bacelar-Nicolau, "Measuring Similarity of Complex and Heterogeneous Data in Clustering of Large Data Sets", *Biocybernetics and Biomedical Engineering*, vol. 29, no. 2, (2009), pp. 9-18.
- [4] H. Bacelar-Nicolau, "On the generalized affinity coefficient for complex data", *Biocybernetics and Biomedical Engineering*, vol. 22, no. 1, (2002), pp. 31-42.
- [5] C. D. Kaddi, R. M. Parry and M. D. Wang, "Multivariate Hypergeometric Similarity Measure", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 99, (2013) March 28, pp. 1.
- [6] I. Spasic and S. Ananiadou, "A Flexible Measure of Contextual Similarity for Biomedical Terms", *Pacific Symposium on Biocomputing*, vol. 10, (2005), pp. 197-208.
- [7] L. G. Shapiro, I. Atmosukarto, H. Cho, H. J. Lin, S. Ruiz-Correa and J. Yuen, "Similarity-Based Retrieval for Biomedical Applications", *Case-Based Reasoning on Images and Signals, Studies in Computational Intelligence*, vol. 73, (2008), pp. 355-387.
- [8] B. T. McInnes and Y. Liu, "UMLS: Similarity: Measuring the Relatedness and Similarity of Biomedical Concepts", *Proceedings of the NAACL HLT, (2013)*, Demonstration Session, pp. 28-31.
- [9] T. Mabotuwana, M. C. Lee and E. V. Cohen-Solal, "An ontology-based similarity measure for biomedical data-application to radiology reports", *J Biomed Inform.*, vol. 46, no. 5, (2013) October, pp. 857-868.
- [10] S. Fong, L. Lu, J. Fiaidhi and S. Mohammed, "A Visualization Methodology For Studying Relations of Medical Data Via Extended Dependency Networks", *The Ninth IASTED International Conference on Biomedical Engineering (Biomed 2012)*, (2012) February 15-17, Innsbruck, Austria, pp. 283-290.
- [11] J. Kang and J. F. Naughton, "Schema Matching Using Interattribute Dependencies", *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 10, (2008) October, pp. 1393-1407

- [12] R. Abraham, J. B. Simha and S. Iyengar, "Medical datamining with probabilistic classifiers", Proc. ICIT IEEE, (2007).
- [13] Y. Hang and S. Fong, "Aerial Root Classifiers for Predicting Missing Values in Data Stream Decision Tree Classification", Proc. 2011 SIAM International Conference on Data Mining (SDM 2011), WS03, (2011) April, Mesa, Arizona, USA, pp. 1-10.
- [14] C. J. Merz and P. Murthy, "UCI repository of machine learning database", <ftp://ftp.ics.uci.edu/pub/machinelearning-databases/>.
- [15] W. T. Li, "Mutual information functions versus correlation functions", J. Stat. Phys., vol. 60, no. 5-6, (1990), pp. 823-837, doi:10.1007/BF01025996.
- [16] Y. Y. Yao, "Information-theoretic measures for knowledge discovery and data mining", In Karmeshu. Entropy Measures, Maximum Entropy Principle and Emerging Applications, Springer, (2003), pp. 115-136.