IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 4, Ver. IV (Jul.-Aug. 2016), PP 46-52
www.iosrjournals.org

SGEDSS: Semantic Gene Expression Model for Communication Decision Supportsystem

Eman K. Elsayed[1],Fatma M. Ghanam[2]
[1, 2,]*Mathematical and computer science Dep., Faculty of Science, Alazhar University, Cairo, Egypt*

Abstract: *An effective decision support system needs to deal with several heterogeneous data sources. Sometimes, the knowledge was explored without integration data sources are wrong. Also, Decision support system is weak in the reusability. To address these challenges, this paper proposes semantic enhancement on the gene expression model (genotype/phenotype system) as an Evolutionary Algorithm to be suitable for "communication decision support system" called Semantic Gene Expression Decision Support System (SGEDSS). The proposed method can solve several mining tasks as association rules and classification through a large volume of heterogeneous data. The proposed method SGEDSS adapts the main components of the genotype/phenotype system by ontology-based to ameliorate the decision support system framework. In thispaper, we use certainly ODSS (Ontology Decision Support System framework), which proposed to enhance DSS. The proposed method simulates somatic mutation on the communication application certainly. That is for accurate simple decision also.*

Keywords: *Data Mining, Decision Support System, Evolutionary Algorithm, genotype/phenotype system, Ontology-based.*

I. Introduction

Some AI techniques are biology inspired computing as a neural network with a brain, Artificial Immune system with Immune system, Swarm with bees, fish or particle behavior and genetic algorithms with evolution soon. Generally, the simple example for that is the natural evolution by simple rules as crossover selection or mutation produces complex organisms.

Data integration is defined as the technique to integrate or collect data from different sources and merge them at one place and finally gives a virtual view to the users [2]. It involves combining data residing in different sources and providing users with a unified view of these data. This process becomes significant in a variety of situations. Data integration is a problem facing many organizations that wish to utilize Web data. The basic principle of data integration is to combine (integrate) selected information sources from a specific domain. We need to integrate data in order to facilitate complex queries and data exchange across multiple heterogeneous data sources. A problem of data integration is the treatment of conflicts caused by different modeling of real world entities, different data models or simply by different representations of one and the same object. *Wrapper* architecture [3] was used in providing data services which accomplish data integration tasks across heterogeneous data sources. But *Wrapper* implementation must be optimized to deal with relational database and xml documents. And it used certainly a simple form of queries without using any decision support system processes as data mining andOLAP.

There are different attempts for integration data by ontology-based as in reference [4] which was integrated only different databases. Also reference [5] had an attempt to integrated heterogeneous data without merging. But `that produces conflictresults.We will enhance the gene expression model as an Evolutionary Algorithm to be suitable for communication decision support system. The proposed method in this paper based on two main concepts Gene expression model (Genotype/ Phenotype system) and ontology-based. So we can classifySGEDSS as a new evolutionary algorithm. Gene expression programming is a genetic algorithm because it uses populations of individuals, selects them according to fitness, and introduces genetic variation using one or more genetic operators. Geneexpression model is much more accurate and stable than the ones based on genetic programming (GP) and linear regression (LR). The components of GE are Chromosome, Expression tree, Fitness function for classification and selection strategy.GE is a powerful method of prediction has been recently increased in many fields [6].

The organization of this paper is as follow: section 2, represents reviews on **Gene expression model (Genotype/ Phenotype system)**. In section 2.1, we represent the review on **Chromosome (genome or genotype)**. In section 2.2, we represent the review on **Expression tree (phenotype)**. In section 2.3, we represent the review on **Fitness function for classification (maximum likelihood)**. In section 2.4, we represent the review on the **selection strategy**. In section 2.5, we represent the review on **Somatic mutations**. In section 3, we represent the review on **ontology**. In section 4, we represent the related reviews on current practices of ontology to supporting decision support system. In section 5, we represent **The Proposed Methodology**

Semantic Gene Expression Decision Support System (SGEDSS). In section 6, we represent the **Implementation and analysis of results**. Finally, we present the **Conclusion of thispaper.**

II. Gene Expression Model (Genotype/ Phenotypesystem)

Gene Expression programming (GEP) is the learn algorithm which can determine the relationships between variables in datasets to build models explain these relations [7]. Genotype is the information about chromosomes structure in cells level. But phenotype is the actual observation properties. The relation between the genotype set and phenotype set called Genotype-phenotype map, where each genotype may have many phenotypes. The benefit of this model generally is the mapping between a different type of data which one is simple as genome and other is complex type as phenotype. Then the main two players in GEP system is genotype with fixed length and phenotype is trees with different sizes andshapes.

2.1Chromosome (genome orgenotype)

It is a packaged and organized structure contains most of the DNA of a living organism. It has all the same size [8].

2.2 Expression tree (phenotype)

Phenotype is expression as tree with different size and shape. In our proposed model, we will substitute the tree with ontology-based [8].

2.3 Fitness function for classification (maximumlikelihood)

We have multinomial categories with crisp classification for discrete data. So we select the maximum likelihood function. Suppose there is a sample (a1, a2…, an) of an independent distributed observations coming fromadistributionwithanunknownprobabilitydensityfunction$f_0(.)$,where$f_0$belongstoacertainfamily of distributions $\{f(.|\theta), \theta\epsilon\vartheta\}$ (where $\theta$is a vector of parameters), so that $\theta= f(.|\theta_0)$. The value θ_0 is unknown and is referred to as the true value of the parameter vector. To use the method of maximum likelihood, one first specifies the joint density function for all observations. The joint density function for the sample (a)is:

$$\mathbf{f(a1, a2, an|\theta) = (a1|\theta) xf(a2|\theta) x…xf(an|\theta)} \quad \mathbf{(1)}$$

When the observed values (a1, a2, …, an) are fixed parameters, this function is called the likelihood and become:

$$\mathbf{l(\theta:a1,a2,…,an) = f(a1,a2,…,an|\theta) = \pi_{i=1}^{n} f(ai|\theta)} \quad \mathbf{(2)}$$

2.4 Selection strategy

As in biology certainly in cell level the selection based on the presence of a catalytic activity that provides a growth advantage to micro-organisms having that specific activity [9].

2.5 Somaticmutations

Somatic mutations are mutations that are not inherited from the parents. Assuming that there is a fewer somatic mutations occur in normal cells. Somatic mutation data as well as other type of mutation data are sparse in character. We focus on reducing the effect of heterogeneity in the identification of similar certain data**.** In DSS we need somatic mutation to apply certain query in certain time withoutinheritance.

III. Ontology Based

Ontology is a term borrowed from philosophy that refers to describe the type of entities in the world and how they relate to each other. The concept of ontology was introduced in the AI field to support the sharing and reuse of formally represented knowledge among different AI systems. The usage of ontology is consistent with the definition since it is broken into simpler sets of such concept definitions and relations when being processed. Others presents ontology as a formal description that describes a concept in a particular domain (classes, called concepts), properties of each concept that describes the various features and attributes of the concepts (slots, called role or properties) and restriction on slots (facets, called role restriction). In other hand, ontology is a computer process able model of a specific domain [10]. It is a formal representation that consists of a set of concepts within a domain. The Hierarchical thinking is important in biological system so using ontology based as object oriented knowledge representation is suitable for using. To solve heterogeneity problem, we use ontology merging which is recognized as an important step in ontology engineering. Ontology extraction tools support the automatic extraction of concepts and/or their relations by applying some techniques such as natural language processing or machine learning. Ontology extraction tools extract for example, text, DB and spread sheet usingText2Onto, DB2OWL and mapping master respectively. That is in protégé editor[11].

IV. Related work

In recent years, the ontology-based was used in different aims. Here, we discuss the use of ontology-based and gene expression in the decision support system. Generally, there is no standard method to use ontology for enhancing decision support system, so we present some works related in this field.

Authors in reference [12] proposed a compact representation for genome mutation. They used gene ontology (GO) with gene expression as it is, that is to apply mining tasks. Authors in reference [13] applied gene expression on a large number of genes (huge data) to classify the data. But reference [14] applied gene expression for clustering data by using K-means algorithm. Authors in reference [15] used ontology based as an input to bring more flexibility to decision support system. Authors in reference [16] proposed how to create ontology-based by protégé 3.4.2. They used the obtained ontology for decision making process. Also authors in reference [17] focused on using ontology to build knowledge- driven decision support system by using OWL rules and reasoning process of "Onto Diabetic system". The authors in reference [18] described a prototype called" security decision support system", where they used ontology as an input for a trust and security DSS to assist in the security decision making process but the used data is homogeneous. Also, we presented in reference [1] a general framework for different using of ontology based in DSS. In reference [19] one study proposed using ontology in different way by suggest rapid miner as ontology-based for data mining optimization ontology. In another study in reference [20] used data mining as DSS process in accessing to data set, where the data came from one source. Finally, references [21] [22] [23] used ontology re-engineering for merging ontologies, for decision support system processes, references [24] [25] used the analytic hierarchy process as DSS process references [26] [27] [28] used expert system as DSS process, and references [29] [30] [31] used OLAP as DSSprocess.

The Proposed Methodology Semantic Gene Expression Decision Support System(SGEDSS)

This section presents SGEDSS phases as a semantic enhancement on the gene expression model (genotype/phenotype system) certainly for communication decision support system. That's to solve the problem of heterogeneous huge data sources. The application of the proposed methodology is restricted with ODSS framework [1] as shown in figure 1. ODSS consists of three phases: extraction knowledge from data sources, merging ontologies in universal ontology to create data warehouse ODWH and finally, compatible DSS processes as data mining or OLAP.

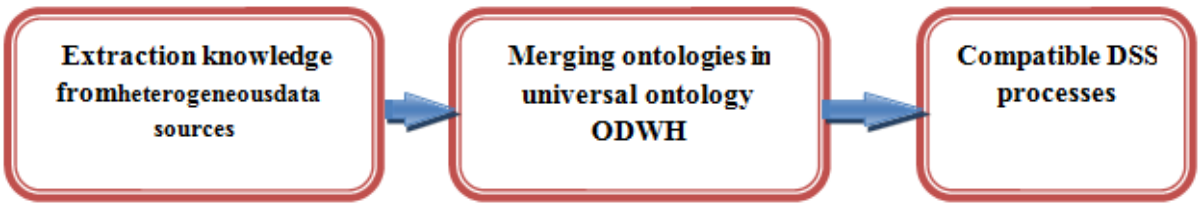

Fig 1: ODSS framework

Then the phases of SGEDSS are:

The first phase is determination the two players Genotype (Chromosome) and Phenotype (Expression tree) where the phenotype has more than one genotype.

The second phase is mapping Genotype and Phenotype that's by merging technique to create Universal Ontology Data warehouse UODW. The benefit of this model generally is the mapping between heterogeneous data which one is simple Chromosomes (customer profile) and other is complex type as phenotype (Communication Ontology). We solve the redundancy which resulted from extraction knowledge from data sources; we use ontology merging which is recognized as an important step in ontology engineering. We merge ontologies resulting from extraction knowledge from data sources. Then filtering the redundancy from the merging of ontologies in the first phase and integrate the result to universal ontology data warehouse. Then we determine the data mart from UODW by classification mining technique.

The third phase is determining fitness function for classification to insert new discrete data in crisp classification that is by likelihood technique as presented in subsection 2.3.

The fourth phase is selection phase that is to select data mart by Selection strategy SQWRL.**The fifth phase is mutation phase in this phase the suitable chosen is a Somatic mutation, where, w**e focus on reducing the effect of heterogeneity in the identification of similar certain data**.** A somatic mutation data as another type of mutation data are sparse in character. We can apply a query to obtain some certain data without relations with each other.

The sixth phase is Compatible DSS Techniques as AHP, Expert system, OLAP or Data mining. In this phase, utilize the analytic hierarchy process (AHP) which is a structured technique for dealing with complex decisions, to make an optimal decision for satisfying the requirement of an individual consumer. The new generation of DSS techniques compatible with ontology as expert system tools that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. OLAP is used to prepare data for analysis, so OLAP is a powerful tool to analyze and make a decision. fig2 presents the SGEDSS framework.

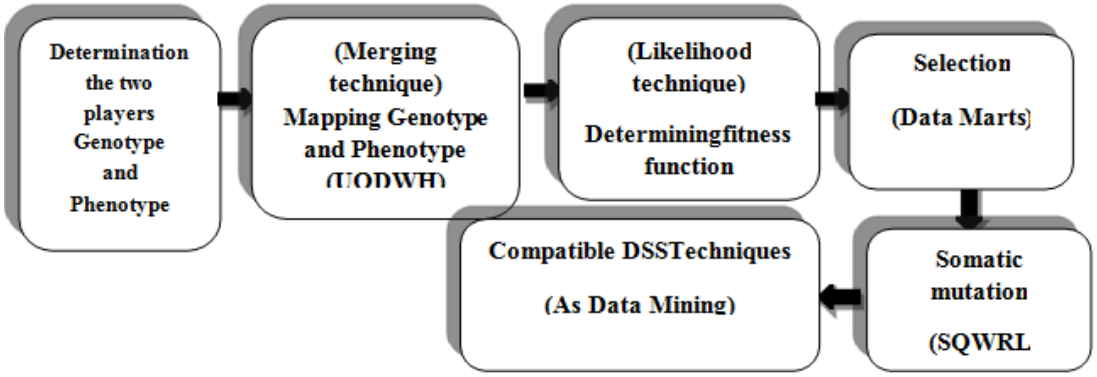

Fig 2: SGEDSS framework

V. Implementation and Analysis ofResults

This section presents the implementation of SGEDSS phases as a semantic enhancement on the gene expression model (genotype/phenotype system) certainly for communication decision support system. The gene expression model GEM has two inputs (genotype, phenotype) to obtain one model. But in our proposed system on SGEDSS contains two inputs (genotype, phenotype) but each could be merge different homogeneous sources. The output is list of models based on which decision we need to take. The same data solves many problems because our data is more dynamics. The data sources are divided into two main parts: structured data and unstructured data. That's to solve the problem of heterogeneous huge data sources. The above framework is applied on UCODW. For example, three different networks, each network has a huge data of customers and stream data from callings between different networks.

The first phase is determination the two players

The two players are Genotype (Chromosome) and Phenotype (Expression tree), where the Genotype is the customer profile and Phenotype is the calling numbers as shown in fig 3.

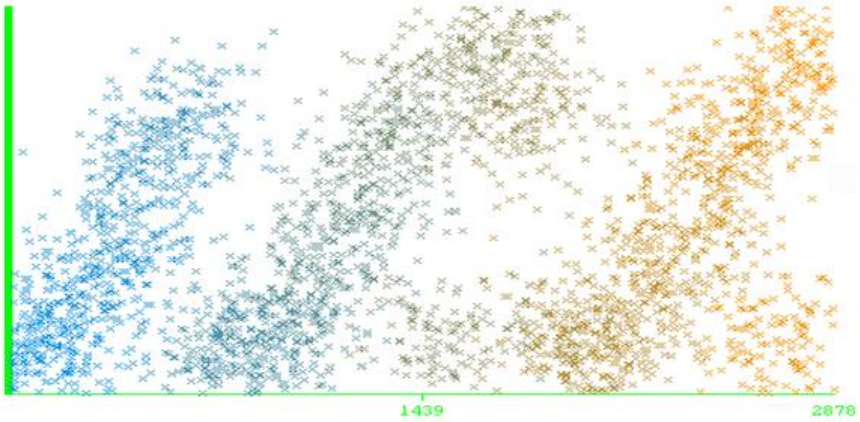

Fig 3: The sample of phenotype data (calling numbers)

The second phase is mapping Genotype and Phenotype

The data of Genotype and Phenotype are inserted directly using mapping DB2OWL and mapping master in protégé editor respectively, where the data is compressed in each column, so object oriented classification is equivalent to smart storage (speed optimized to query mode for data warehouse). To implement

this phase, we use merging technique to create Universal Ontology Data warehouse UODW. That solves the redundancy which resulted from extraction knowledge from data sources; it uses ontology merging which is recognized as an important step in ontology engineering. The benefit of this model generally is the mapping between heterogeneous data which one is simple phenotype (Sheets of telephone numbers) and other is complex type as genotype (Communication Ontology). This paper solves the redundancy which resulted from extraction knowledge from data sources; it uses ontology merging which is recognized as an important step in ontology engineering. In the merging process, there is permission for using overlapping in order not to stop this process. Then merge ontologies resulting from extraction knowledge from data sources. Then we create the semantic data warehouse. The Figure 4 displays the ontology star schema for these data sources by using owl protégé editor.

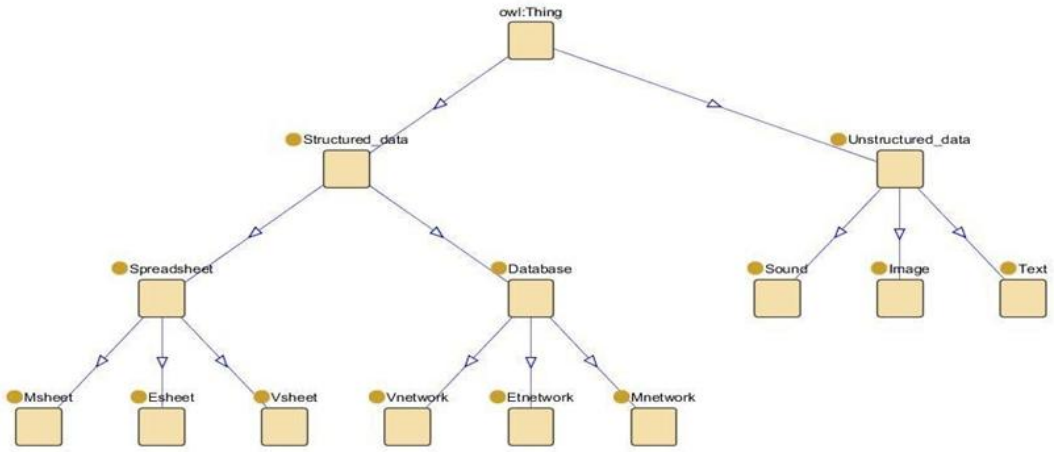

Fig 4: The ontology star schema for heterogeneous data sources

For example, we support x is the first network, f is the second network and r is the third network. y, g and s are the time of calling in the first, second and third networks respectively. z, h and t are the caller in the first, second and third networks respectively. a, b, c are the called number in the first, second and third networks respectively according to this queries:

$$(?x)\wedge(?x,?y)\wedge(?x,'7{:}33{:}45')\wedge(?x,?z)\wedge(?x,?a) \rightarrow sqwrl{:}select(?z,?a).$$
$$(?f)\wedge(?f,?g)\wedge(?f,'10{:}12{:}10')\wedge(?f,?h)\wedge(?f,?b) \rightarrow sqwrl{:}select(?h,?b).$$
$$(?r)\wedge(?r,?s)\wedge(?r,'2{:}35{:}33')\wedge(?r,?t)\wedge(?r,?c) \rightarrow sqwrl{:}select(?t,?c).$$

The third phase is determining fitness function for classification

To insert new discrete data in crisp classification that is by likelihood technique as represented in the section 2.3.

We applied different queries with and without classification phase. When we computed the run time which is resulting from SQWRL queries in each process as shown in fig 5, we found that without classification the runtime will be infinitetime.

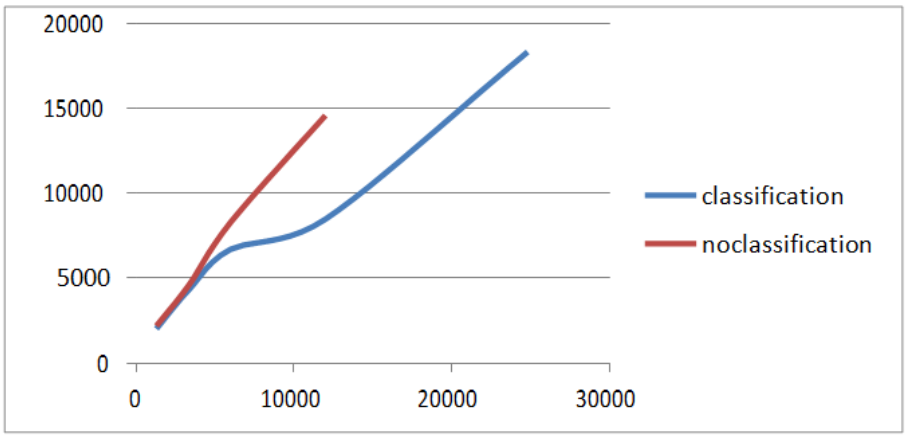

Fig 5: TheComparing between processes with and without classification

The fourth phase is selection phase

That is to select data mart by Selection strategy SQWRL. The data mart is the class where there is the output of the query. When we converting data sources to ontology.

The fifth phase is mutation phase

In this phase the suitable chosen is Somatic mutation. Where, we focus on reducing the effect of heterogeneity in the identification of similar certain data**.** Somatic mutation data as well as other type of mutation data are sparse in character, we can apply query to obtain some certain data without relations with each other. Finally, for decision making we can apply the association rules for SWRL queries in the same time on any classes for universal communication ontology data warehouse (UCODWH), and this is an example for SQWRL queries of associationrules:

$(?\boldsymbol{x})\wedge(?\boldsymbol{x},?\boldsymbol{y})\wedge(?\boldsymbol{x},'\ \mathbf{7{:}33{:}45}')\wedge(?\boldsymbol{x},?\boldsymbol{z})\wedge(?\boldsymbol{x},?\boldsymbol{a}) \rightarrow \boldsymbol{sqwrl{:}select}(?\boldsymbol{z},?\boldsymbol{a})$.

$(?\boldsymbol{f})\wedge(?\boldsymbol{f},?\boldsymbol{g})\wedge(?\boldsymbol{f},'\ \mathbf{10{:}12{:}10}')\wedge(?\boldsymbol{f},?\boldsymbol{h})\wedge(?\boldsymbol{f},?\boldsymbol{b}) \rightarrow \boldsymbol{sqwrl{:}select}(?\boldsymbol{h},?\boldsymbol{b})$.

$(?\boldsymbol{r})\wedge(?\boldsymbol{r},?\boldsymbol{s})\wedge(?\boldsymbol{r},'\ \mathbf{2{:}35{:}33}')\wedge(?\boldsymbol{r},?\boldsymbol{t})\wedge(?\boldsymbol{r},?\boldsymbol{c}) \rightarrow \boldsymbol{sqwrl{:}select}(?\boldsymbol{t},?\boldsymbol{c})$.

Figure 6 shows the association rules results in the same time on any classes for universal communication ontology data warehouse (UCODWH) which are used for determining the relation among networks in decision making.

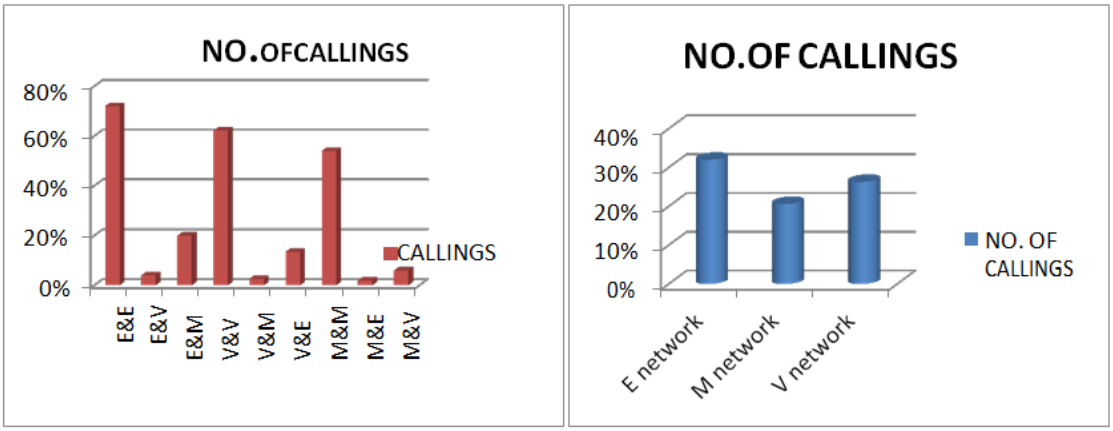

Fig 6: The association rules results for all networks in the same time

VI. Conclusion And Furtherwork

This paper has proposed SGEDSS phases as a semantic enhancement on the gene expression model (genotype/phenotype system) certainly for communication decision support system based on ODSS. That's to solve the problem of heterogeneous huge data sources and reusability. This system consists of six phases: The first phase is determination the two players Genotype and Phenotype in a certain application. The second phase is mapping Genotype and Phenotype to produce universal ontology data warehouse UODW. This phase solves conflict problems and redundancy. The third phase is determining fitness function for classification as maximum likelihood. Without this phase, the runtime may infinite. The fourth phase is a selection in GES which means data marts in DSS. The fifth phase is somatic mutation phase which suitable for application without inheritance information. The final phase is selection DSS process as data mining. The proposed method SGEDSS can solve several mining tasks as association rules through huge data. It integrates the main components of the genotype/phenotype system with ontology- based to ameliorate the decision support system. Using ontology with gene expression on ODSS framework is more general and reusabilitysystem.

In the future, we will go to more generalization and reusability tool for SGEDSS and adapted ontology to solve big data problems. Also, we will apply our proposed method in different data types which need more analysis such as image, text and multimedia. Finally, we will enhance the system to suitable for continuous data with fuzzy not only crisp.

References

[1] E. Eman K., M. Elnahas, and M. GhanamFatma. Framework for using Ontology Base to Enhance Decision Support System.Framework 2.02 (2013).

[2] Srivastava, Kingshuk, P. S. V. S. Sridhar, and AnkitDehwal. Data Integration Challenges and Solutions: A Study. International Journal of Advanced Research in Computer Science and Software Engineering 2.7 (2012).

[3] Zaman, ErMajid, S. M. K. Quadri, and ErMuheet Ahmed Butt. Information Integration for Heterogeneous Data Sources. IOSR Journal of Engineering2.4 (2012): 640-643.
[4] Bizid, Imen, et al. Integration of heterogeneous spatial databases for disaster management. Advances in ConceptualModeling.Springer International Publishing, 2013.77-86.
[5] Buitelaar, Paul, et al. Ontology-based information extraction and integration from heterogeneous data sources. International Journal of Human-Computer Studies 66.11 (2008): 759-788.
[6] Shekarian, Ehsan, and AlirezaFallahpour. Predicting house price via gene expression programming.International Journal of Housing Markets and Analysis 6.3 (2013): 250-268
[7] Ferreira, Cândida. Gene expression programming: mathematical modeling by an artificial intelligence. Vol. 21.Springer, 2006.
[8] Palsson, Bernhard Ø. Systems biology: constraint-based reconstruction and analysis. Cambridge University Press, 2015.
[9] Boersma, Ykelien L., Melloney J. Dröge, and Wim J. Quax. Selection strategies for improved biocatalysts. Febs Journal 274.9 (2007):2181-2195.
[10] Eschenbach, Carola, and Michael Gruninger, eds. Formal Ontology in Information Systems: Proceedings of the Fifth International Conference (FOIS 2008). Vol. 183.IOS Press, 2008.
[11] Protégé and its Libraryhttp://protege.stanford.edu/plugins/owl/owl-library/koala.owl
[12] Kim, Sungchul, Lee Sael, and Hwanjo Yu. A mutation profile for top-k patient search exploiting Gene- Ontology and orthogonal non-negative matrix factorization. Bioinformatics 31.22 (2015): 3653-3659.
[13] Vanitha, C. Devi Arockia, D. Devaraj, and M. Venkatesulu. Gene Expression Data Classification Using Support Vector Machine and Mutual Information-based Gene Selection.Procedia Computer Science 47 (2015): 13-21.
[14] Chandrasekhar, T., K. Thangavel, and E. Elayaraja. Effective clustering algorithms for gene expression data.arXiv preprint arXiv:1201.4914 (2012).
[15] Galopin, A., et al. An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders. MEDINFO 2015: EHealth-enabled Health: Proceedings of the 15th World Congress on Health and Biomedical Informatics. Vol. 216.IOS Press, 2015.
[16] Donfack, Guefack V., et al. Ontology driven decision support systems for medical diagnosis-an interactive form for consultation in patients with plasma cell disease. Studies in health technology and informatics 180 (2011): 108-112.
[17] Sherimon, P. C., and Reshmy Krishnan. OntoDiabetic: An Ontology-Based Clinical Decision Support System for Diabetic Patients. Arabian Journal for Science and Engineering (2015): 1-16.
[18] Bernabe, Jorge Bernal, Gregorio Martinez Perez, and Antonio F. Skarmeta Gomez. Intercloud Trust and Security Decision Support System: an Ontology-based Approach. Journal of Grid Computing 13.3 (2015): 425- 456.
[19] Keet, C. Maria, et al. The data mining OPtimization ontology. Web Semantics: Science, Services and Agents on the World Wide Web 32 (2015):43-53.
[20] Sharma, Arjun, et al. Focused Decision Support: a Data Mining Tool to Query the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Dataset and Guide Screening Management for the Individual Patient. Journal of digital imaging (2015): 1-5.
[21] Hitzler, Pascal, et al. What is ontology merging?.American Association for Artificial Intelligence.2005.
[22] Guzmán-Arenas, Adolfo, and Alma-Delia Cuevas. Knowledge accumulation through automatic merging of ontologies. Expert Systems with Applications37.3 (2010): 1991-2005.
[23] Fahad, Muhammad, NejibMoalla, and AbdelazizBouras. Towards ensuring satisfiability of merged ontology.Procedia Computer Science 4 (2011): 2216-2225.
[24] Chen, Ming-Kuen, and Shih-Ching Wang. The critical factors of success for information service industry in developing international market: Using analytic hierarchy process (AHP) approach. Expert Systems with Applications 37.1 (2010):694-704.
[25] Ishizaka, Alessio, and Ashraf Labib. Selection of new production facilities with the group analytic hierarchy process ordering method. Expert Systems with Applications 38.6 (2011):7317-7325.
[26] Karmakar, S., et al. Development of expert system modeling based decision support system for swine manure management. Computers and electronics in agriculture 71.1 (2010): 88-95.
[27] Shue, Li-Yen, Ching-Wen Chen, and WeissorShiue. The development of an ontology-based expert system for corporate financial rating. Expert Systems with Applications 36.2 (2009): 2130-2142.
[28] Karmakar, S., et al. Development of expert system modeling based decision support system for swine manure management. Computers and electronics in agriculture 71.1 (2010): 88-95.
[29] Prat, Nicolas, Isabelle Comyn-Wattiau, and Jacky Akoka. Combining objects with rules to represent aggregation knowledge in data warehouse and OLAP systems. Data & Knowledge Engineering 70.8 (2011): 732-752.
[30] Abdullah, Ahsan. Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool. Computers and Electronics in Agriculture 69.1 (2009): 59-72.
[31] Pardillo, Jesús, Jose-Norberto Mazón, and Juan Trujillo. Extending OCL for OLAP querying on conceptual multidimensional models of data warehouses. Information Sciences 180.5 (2010): 584-601.