World Applied Sciences Journal 29 (Computer Sciences, Engineering and Its Applications): 195-199, 2014
ISSN 1818-4952
© IDOSI Publications, 2014
DOI: 10.5829/idosi.wasj.2014.29.csea.2265

Patient Monitoring in Gene Ontology with Words Computing Using SOM

V. Khanaa, K.P. Thooyamani and R. Udayakumar

School of Computing Science, Bharath University-73, India

Abstract: The computing-with-words *(CW)* is a methodology in which words are used in place of numbers for computing and reasoning. The CW plays a vital role in Fuzzy logic and vice versa. I propose a ontological self organizing map (OSOM) which produces summarized and visualized information about dataset namely ontological data. Initially Kohonen developed SOM which is further extended to OSOM. Gene Ontology (GO) annotations of genes and gene products are used as the specified data. The result is presented on two datasets composed of GO annotations of genes. The results show that the OSOM-based visualization method has the cluster tendency of the genes and gene products and that the summarization provides useful information about the mapped groups of genes and gene products. I also propose a solution to automate the process of patients' vital data collection using "sensors" and "RFID". Sensors are attached to the existing medical equipment and are interconnected to exchange service. Summarized result along with the patient vital data collection are sent to the expert system from where it processed and distributed to the medical staff and the medical staff can access patient vital data using RFID also.

Key words: Bioinformatics · Computing with words (CW) · Fuzzy logic · Ontology · Self-organizing maps (SOMs) · Radio Frequency Identification (RFID)

INTRODUCTION

Computing, in its usual sense, is centered on manipulation of numbers and symbols. In contrast, computing with words, or CW for short, is a methodology in which the objects of computation are words and propositions drawn from a natural language, e.g., small, large, far, heavy, not very likely, the price of gas is low and declining. As a methodology, computing with words provides a foundation for a computational theory of perceptions - a theory which may have an important bearing on how humans make - and machines might make - perception-based rational decisions in an environment of imprecision, uncertainty and partial truth [1].

Approach

Approach Overview: A different view toward applying CW principles by using ontologies, Collections of words organized in hierarchical taxonomies. In this project the system has words as inputs and produces a linguistic output. However, unlike the standard definition of CW, the uncertainty is not in the language itself but in the relationships between the words and, more importantly, how those words are used to describe objects [2]. Words that are connected in the taxonomy are related by phrases such as "is a," "part of," and "adjacent to." Hence, the ontology itself is composed of a constrained (albeit crisp) collection of linguistic perceptions. For example, in the Gene Ontology (GO), *collagen* (GO: 0005581) is a part of *proteinaceous extracellular matrix* (GO: 0005578), which is an *extracellular matrix part* (GO: 0044420).

I apply a novel extension to the SOM that allows us to use the SOM with ontological data. Ontological data are unique in that the data samples are composed of collections of terms or words taken from a predefined corpus. Unlike conventional object data, the samples do not have a numerical location. Examples of ontological data include web sites, medical-record annotations and publications. In this paper, I apply our ontological SOM (OSOM) to produce cluster visualization and functional summarization of annotated gene products in the GO. The relational data of the gene products are produced by GO similarity measures [3].

Corresponding Author: V. Khanaa, School of Computing Science, Bharath University-73, India.

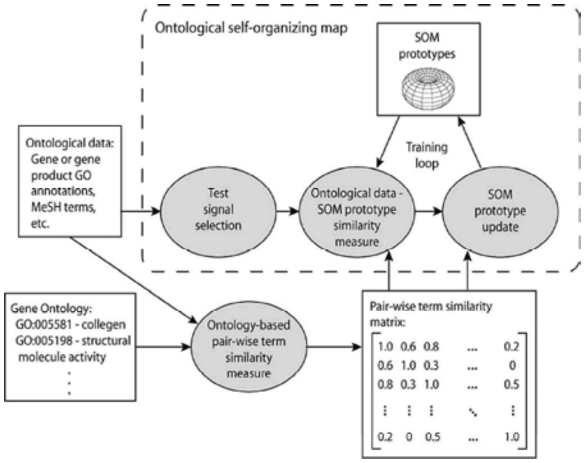

Ontological self-organizing map
SOM prototypes
Training loop
Ontological data: Gene or gene product GO annotations, MeSH terms, etc.
Test signal selection
Ontological data - SOM prototype similarity measure
SOM prototype update
Gene Ontology: GO:005581 - collegen GO:005198 - structural molecule activity
Ontology-based pair-wise term similarity measure
Pair-wise term similarity matrix:
1.0 0.6 0.8 ... 0.2
0.6 1.0 0.3 ... 0
0.8 0.3 1.0 ... 0.5
0.2 0 0.5 ... 1.0

Fig. 1: OSOM training block diagram.

Fig. 1 shows the block diagram of the OSOM training algorithm. The inputs are the ontological data and the pair wise term-similarity matrix [4]. The OSOM itself operates very much like a conventional SOM: 1) A random test signal is chosen; 2) the winning prototype is selected; and 3) all prototypes are moved toward the test signal according to a predefined network topology.

Gene-Ontology Similarity Measures: Information about gene products and how they are similar to one another is of great importance in bioinformatics. Traditional approaches use the DNA sequence as well as the expression values from micro array experiments. However, additional information, which is more symbolic in nature, is available about genes and gene products [5]. This symbolic information comprises the GO terms and index terms in publications about gene products. I use these symbolic data to build visualizations and functional summarizations of the genes or gene products. Previously, I developed methods to compute the similarity of two gene products that are annotated by GO terms. Section IV-C describes how I utilize the OSOM to produce cluster visualization of the ontological data. The visualization method maps the *ontological profiles* (i.e., the OSOM prototypes) of the OSOM network to a 2-D toroidal grid (although any predefined network topology could be chosen).Cluster tendency is shown by the relations between neighboring ontological prototypes on the grid, which are displayed as gray levels—black represents *no relation* and white represents *highly related*.

To summarize, each gene product *Gi* is represented by a collection of terms *Gi = {Ti*1.. . *Tin }*, where *Ti*1 is the first annotation of *Gi* (e.g., GO:0016740—transfers activity). Based on these sets of terms, a similarity between two gene products can be found by performing an aggregation on the pair wise similarities among each set of terms. For example, the average is computed as

$$s(G_i,G_j)\frac{\sum_{k=1}^{n}\sum_{l=1}^{m}R_{kl}}{mn} \tag{1}$$

where *Gi* is annotated by *n* terms, *Gj* is annotated by *m* terms and *Rkl* is the pair wise similarity between the *k*th term in *Gi* and *l*th term in *Gj.* The pair wise term similarity *skl* (or dissimilarity) can be computed in many ways; shortest-path-based similarity and information-theoretic constructs are the most widely used [6].

Ontological Self-Organizing Map: The SOM is a two-layer lateral-feedback neural network that topologically maps itself to the training data. The network structure is often set to a 2-D square or hexagonal grid, where each network node, or prototype, is laterally connected to its neighbors. The network-learning algorithm is as follows.

- Randomly draw a sample from the training data $\vec{x}_d$.
- Find the closest SOM prototype *p* according to a chosen distance metric

$$p = \arg\min_{\mathrm{j}}\left\{\left\|\vec{x}_d - \vec{a}_i^{(old)}\right\|\right\}. \tag{2}$$

- Update SOM prototypes by

$$\vec{a}_i^{(new)} = \vec{a}_i^{(old)} + \in (t)\ .\ h_{ip}\ .\ \left(\vec{a}_d - \vec{a}_i^{(old)}\right) \tag{3}$$

where _(*t*) is the learning rate and *hip* is the neighborhood function, which is defined

$$h_{ip}(t) = \exp\left(-\frac{\left|\vec{a}_i - \vec{a}_p\right|^2}{\sigma^2(t)}\right) \tag{4}$$

where $\vec{a}_i$ is the location of SOM prototype in the predefined neighborhood (e.g., square or hexagonal grid) [7]. This algorithm is repeated until a maximum number of iterations or convergence is reached. Typically, the learning rate and the width of the neighborhood function are reduced during iteration, with the effect that late iterations are only applying small updates to network prototypes that are local to the winning prototype *p*.

Prototype Representation: The algorithm that I propose as the OSOM is an adaption of the standard SOM to ontological data. First, I construct an ontological weight vector for each node in the OSOM grid. This weight vector is a fuzzy-membership representation of all the terms present in the training data [8]. For example, the *GPD*194 dataset contains a total of 64 terms among the entire gene products combined; thus, the OSOM weight vector has a length of 64. Each weight-vector element is associated with one term and the value of the weight is the membership of the associated term in the description of the ontological prototype [9]. I denote the OSOM weight vectors as $\vec{w}_i \in [0,1]^{N_T}$. Second, I replace the distance metric in step 2 of the SOM with a similarity measure. The measures that I use are vector matrix - multiplication-based operations that are simple extensions of the measures described in Section III, What makes these similarity measures different is that they are the similarity between an OSOM prototype and a gene or gene product; the similarity measures in Section III were for two genes or gene products. In practice, one could choose any similarity measure that measures the similarity of two sets of terms; there are many measures that fit this description [10]. However, I recommend using similarity measures that perform some aggregation on the pair wise similarity matrix R.

Prototype Update: The OSOM can use the standard weight-vector update of the SOM by substituting $\vec{g}_d$ for $\vec{x}_d$

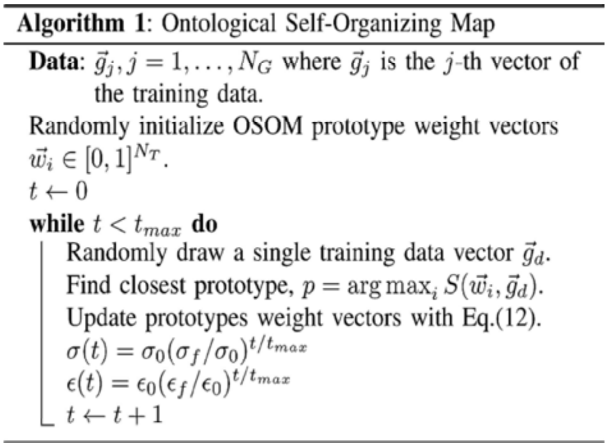
$$\vec{w}_i^{(new)} = \vec{w}_i^{(old)} + \in (t)\ .\ h_{ip}\ .\ \left(\vec{g}_d - \vec{w}_i^{(old)}\right) \tag{5}$$

Let us recall that _g is binary; hence, this update simply moves the prototype toward the corresponding corner of the *NT* - dimensional hypercube. This update, however, ignores the term–term similarities. I can also replace the standard form of the weight-vector update equation with a similarity-based update. In order to create a similarity-based update equation, I defined the following two axioms [11].

- At each iteration, the weight-vector elements that correspond to the terms in the test signal $\vec{g}_d$ must increase [12].
- At each iteration, the weight-vector elements that are similar to the terms in $\vec{g}_d$, as evidenced by R, must also increase. With these axioms in mind, I created the following update equation:

$$\vec{w}_i^{(new)} = \vec{w}_i^{(old)} + \in (t)\ .\ h_{ip}(\mathrm{t})\ .\ \left(F\left(R, \vec{g}_d\right) - \vec{w}_i^{(old)}\right)\forall_i \tag{6}$$

where *p* denotes the closest OSOM prototype to the randomly chosen training vector $\vec{g}_d$ and $\left(F\left(R, \vec{g}_d\right) - \vec{w}_i^{(old)}\right)$ is the update operator. As shown below, the update operator is computed from the columns of the similarity matrix that correspond to nonzero elements of the training vector $\vec{g}_d$, These columns of the similarity matrix represent the similarity between the terms in $\vec{g}_d$ and all other terms (e.g., *Rij* is the similarity of the *i*th and *j*th terms). Hence, the update operator, i.e., omputes a row aggregation on the columns of the similarity matrix R that corresponds to the terms in the training vector $\vec{g}_d$.



```
Algorithm 1: Ontological Self-Organizing Map
Data: $\vec{g}_j, j = 1, \ldots, N_G$ where $\vec{g}_j$ is the $j$-th vector of
      the training data.
Randomly initialize OSOM prototype weight vectors
$\vec{w}_i \in [0,1]^{N_T}$.
$t \leftarrow 0$
while $t < t_{max}$ do
    Randomly draw a single training data vector $\vec{g}_d$.
    Find closest prototype, $p = \arg\max_i S(\vec{w}_i, \vec{g}_d)$.
    Update prototypes weight vectors with Eq.(12).
    $\sigma(t) = \sigma_0(\sigma_f/\sigma_0)^{t/t_{max}}$
    $\epsilon(t) = \epsilon_0(\epsilon_f/\epsilon_0)^{t/t_{max}}$
    $t \leftarrow t + 1$
```

where, $i = \{l \in \square \mid l \leq N_T ;)(\vec{g}_d)l = 1\}, k = 1,\ldots, N_T$ and R_{ki} is the *i*th column of the *k*th ow of the similarity matrix R.

Cluster Visualization: The visualization method that I propose is composed of two distinct steps. First, the objects (e.g., gene products, articles, etc.) are mapped to the trained OSOM network by the nearest prototype rule—for each object $\vec{g}$ find the best-match prototype with $p = \arg\max_i S(\vec{w}_i, \vec{g})$. . The prototype *p* is then annotated with the object information of $\vec{g}$ (e.g., the gene product id) [13]. This groups similar objects (i.e., gene products) into cluster-like arrangements, where each OSOM prototype essentially represents a cluster (sometimes an empty cluster). Second, the similarity between neighboring OSOM prototype nodes is mapped into a gray-scale or color image for this paper, red indicates *very similar*, blue indicates *very dissimilar*. The color map used in this paper is shown in Fig. 2.

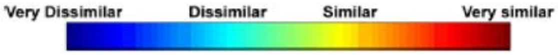


Fig. 2: Color map used in visualizations shows relative similarity between network prototypes.

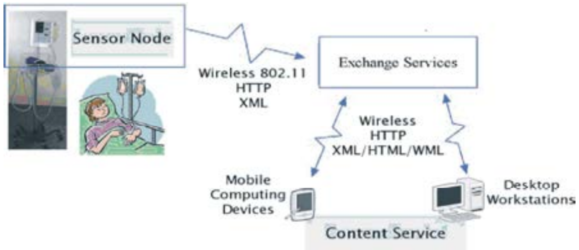


Fig. 3: Proposed Solution of Patient Monitoring.

Cluster Summarization: Cluster summarization, i.e., the potential output of a CW engine, of the ontological prototypes is achieved by examining the OSOM prototype weight vectors. For the case of genes or gene products, this summarization is a functional summarization of each group. The ontological content of each OSOM prototype is represented by a weight vector, as discussed in Section IV. Each element of the weight vector can be viewed as the relative influence of a specific annotation in defining the profile of its associated OSOM prototype. Thus, high values in a weight vector signify a high likelihood that the objects mapped to a location are annotated by the associated term(s). I define the most-representative term (MRT) of an ontological prototype as the term that has the highest associated weight in the OSOM prototype weight vector. If there is more than one maximum weight-vector element, then the MRT is defined as the term with the highest information content. This provides a simple linguistic output for a potentially complex organization of the ontological description of groups of genes.

Patient Monitoring: The proposal is a system to automate the process of collecting patient's vital data *via* a network of sensors connected to legacy medical devices and deliver this information along with the summarized values of gene and gene product to the medical center's for storage, processing and distribution.

Figure 3 depicts the proposed solution.At the patient's bedside, there are *sensor nodes* which are loaded with software to collects, encode and transmit data through wireless communication channels to be stored. The *Exchange Service* acts like a broker between local and remote services. It is responsible to receive collect data from sensors and to dispatch it to appropriate storage service.It also receives requests from content service whose functionality is two folded: (1) it is responsible to provide services to store collected data; and (2) it provides a platform for development, testing and deployment of applications needed by medical staff. Mobile and stationary devices interacts with applications using *Content Service.* This service acts like a “door” where medical staff devices can access all available information. Using the RFID the medical staff can access the patient information directly before entering the ward.

CONCLUSION

The results in show that the OSOM is a powerful tool to visualize the relationships between objects composed of ontological data. Because these data are represented as collections of terms, the standard SOM is ill-equipped for these data.

Thus, the OSOM can do everything the SOM is able to, but it can also analyze ontological data. The OSOM encodes the ontological data directly and computes a visualization of the gene products that shows how they are related to one another. Additionally, the weight values in the OSOM prototypes are the relative strength of each GO term in defining the genes and gene products mapped to that prototype. Each prototype is essentially a sentence that describes the mapped genes and/or gene products and then the summarized information along with the patient vital data is sent to the medical staffs without any human interference and the medical staff can access this information using RFID also.

REFERENCES

1. Zadeh, L., 1973. Outline of a new approach to the analysis of complex system and decision processes, IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 1,Jan. 1973. IEEE Trans. Syst., Man, Cybern.
2. Zadeh, L., 2002. From computing with numbers to computing with words— From manipulation of measurements to manipulation of perceptions, Int. J. Appl. Math. Comput. Sci., 12(3): 307-324.
3. Mendel, J., 2001. The perceptual computer: An architecture for computing with words, in Proc. FUZZ-IEEE, Melbourne, Vic., Australia, pp: 35-38.
4. Mendel, J., 2007. Computing with words and its relationships with fuzzistics, Inf. Sci., 177(4): 988-1006.

5. Mendel, J., 2007. Computing with words: Zadeh, Turing, Popper and Occam, IEEE Comput. Intell. Mag., 2(4): 10-17.
6. The Gene Ontology Consortium, The Gene Ontology (GO) database and informatics resource, Nucl. Acids Res., 32: D258-D261.
7. Xu, D., J. Keller, M. Popescu and R. Bondugula, 2008. Applications of Fuzzy Logic in Bioinformatics. London, U.K.: Imperial College.
8. Peter Varady, Zoltan Benyo and Balazs Benyo, 2002. An Open Architecture Patient Monitoring System Using Standard Technologies IEEE Transactions, 6(1).
9. Udayakumar, R. and Kumaravel A. Rangarajan, 2013. Introducing an Efficient Programming Paradigm for Object-oriented Distributed Systems, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(5S): 4596-4603.
10. Udayakumar, R., V. Khanaa and K.P. Kaliyamurthie, 2013. Performance Analysis of Resilient FTTH Architecture with Protection Mechanism, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(6): 4737-4741.
11. Saravanan, T. and R. Udayakumar, 2013. Comparision of Different Digital Image watemarking techniques, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1684-1690.
12. Saravanan, T. and R. Udayakumar, 2013. Optimization of Machining Hybrid Metal matrix Composites using desirability analysis, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1691-1697.
13. Thooyamani, K.P., V. Khanaa and R. Udayakumar, 2013. Detection of Material hardness using tactile sensor, Middle-East Journal of Scientific Research, ISSN:1990-9233 15(12): 1713-1718.