Spatial-Temporal Data Representation in Ontology System for Personalized Decision Support

Muhaini Othman[12], Nikola Kasabov[1], Raphael Hu[1]

[1]Knowledge Engineering and Discovery Research Institute
Auckland University of Technology, New Zealand
{muhaini.othman, nikola.kasabov, raphael.hu}@aut.ac.nz

[2]Universiti Tun Hussein Onn Malaysia, Malaysia
{muhaini@uthm.edu.my}

Abstract. This research is focus on designing a novel framework for personalized decision support and subsequently develops an application that employs discrete and spatial-temporal data for personalized knowledge representation and prognosis. The framework aims to unify machine learning approach and ontology knowledge representation approach for a better decision support. The main elements are the ontology system and personalized modeling engine where the relation between these two elements performs as an evolving system that interacts actively and harnesses the knowledge from each other hence enriching the knowledge base by discovering new knowledge. The personalized modeling engine is a process of model creation for a single input vector in a problem space based on nearest neighbor spatial-temporal data and information available in the ontology system. The first part of this study will focus on developing the personalized modeling engine to process the first case study dataset relate to weather and stroke occurrences. Several methods for personalized modeling will be investigated include classical and new methods that could be utilized to process spatial-temporal dataset.

Keywords: ontology, personalized modeling, spatial-temporal data, decision support

1. Introduction

The increment of spatial-temporal data (STD) collected in many domain areas, including bioinformatics, engineering, medicine, environment, telecommunication, computer vision and many more has leads towards the emergent of information technology initiatives that facilitate the knowledge acquisition, organization and dissemination among research community. The initiatives include but not exhaust to spatial-temporal representation in ontology and databases, spatial-temporal modeling and spatial-temporal reasoning that fall under data mining research. Both the temporal and spatial dimensions add substantial complexity to data mining tasks [1]. Spatio-temporal data mining here refers to the extraction of implicit knowledge, spatial and temporal relationships or other patterns not explicitly stored in spatio-temporal databases [2].

2. Rationale and Significance of the Study

In the domain areas of bioinformatics, the concerns of manipulating STD to represent knowledge is crucial because it could leads to the notion of improving and saving lives either for human, animal or environment. Currently, there a lot tools that is used knowledge representation in bioinformatics areas such as ontology, formal logic and representation method. However, most ontology are developed heterogeneously and designed for diverse purposes. Ontologies used in bioinformatics differ with respect to their scope, size and granularity [3]. Thus there is an emerging demand for the integration and exploitation of heterogeneous biomedical information for improved clinical practice, medical research and personalized healthcare [4]. Nevertheless, the main issue here is about how ontology can be utilized to represent STD since a spatio-temporal ontology can have many rules of qualitative reasoning on spatial and temporal data which provide a valuable source of domain independent knowledge that should be taken into account when generating patterns [1]. Towards developing personalized decision support specifically for healthcare, the spatio-temporal ontology system is insufficient because it just a representation of spatio-temporal knowledge and how it relates to each other through rules that need a machine learning approach where the relation between these two elements performs as an evolving system that interacts actively and harnesses the knowledge from each other.

The unification of spatio-temporal ontology system and a machine learning approach could be the tool for a better understanding of health related problems like chronic disease including cardiovascular disease, stroke, cancer and many more unsolved medical problems. For instance, health related problem like chronic diseases are the major cause of death in almost all countries and it is projected that 41 million people will die of a chronic disease by 2015 unless urgent action is taken [5]. Various initiatives has been taken to control the increment of chronic disease patients such as clinical prevention using combination drug therapy and calculation of person's risk by referring to risk chart taking account several risk factors. Additional initiatives involve the use of statistic method to generate survival model and investigate several risk factor associate with chronic disease such as Cox Proportional Hazards Model [6][7][8]. There are also several machine learning applications that used global models for prediction of person's risk or the outcome of a certain diseases [9][10][11][12]. Hence, using global models for prediction of a person's risk is inadequate based on the assumption that every person or individual have their own unique characteristics, since, personal human health is defined by many factors such as the food they eat, their lifestyle, life stage, ethnic origin, previous growth and development, gender, environment influences, genetic differences, allergies, diseases and many other important factors [13]. Consequently, the emerging approach that can be utilized to solve the problem is personalized modeling, where a model is created for every single new input vector of the problem space based on its nearest neighbors using transductive reasoning approach [14]. Nevertheless, the personal health can also be influenced by space (such region and distance) and temporal constraint (before and after) and relations between them. In the case of personalized modeling this valuable information could improve the outcome of result immensely. However, there is lack of efficient methods for the analysis of such complex data that is not only discrete but also spatio-temporal in nature and for the discovery of complex spatio-temporal patterns in it, especially for on-line and real time applications. So this research it will focus on developing an application that employs discrete and spatio-temporal data for personalized knowledge representation and prognosis (etc. stroke). This research will further improved the framework proposed by Kasabov et al. [15], the framework refers to integrating local and personalized modeling methods with a global ontology knowledge and data repository for a better personalized decision support, for new knowledge discovery and for a better understanding. Accordingly, the personalized modeling method to be integrated in the system, Integrated Method for Personalized Modeling (IMPM) introduced by Kasabov et al. [16].

3. Objectives

i. To review the literature concerning how integration of ontology and personalized modeling can best predict possible outcomes for a new person using spatial-temporal historical data.
ii. To design a framework that incorporates spatial-temporal ontology system with personalized modeling engine which facilitate new knowledge discovery and to provide better decision support.
iii. To develop a practical application that implements proposed spatial-temporal ontology system with personalized modeling engine framework.
iv. To verify the proposed spatial-temporal ontology for personalized decision support framework by testing it on a case study of chronic disease such as stroke.

4. Research Question

The main research question here is: Can spatial-temporal ontology system for personalized decision support be developed to produce a better personalized knowledge representation and risk prognosis for a person?

More specifically this research aims to answer the following questions:

i. How to develop a novel spatial-temporal ontology system for personalized decision support using incrementally spatial-temporal data from various sources?
ii. How to select optimal set of features, neighbourhoods, model and its parameters for personalized modeling engine using spatial-temporal data?
iii. How to accurately estimate the best time window for predicting the health status of a person?
iv. How to visualise these complex spatial-temporal data in an ontology system?
v. How to extract the spatial-temporal data from ontology for spatio-temporal pattern recognition process?

5. Literature Review

5.1. Ontology and Personalized Modeling

Ontology is an explicit specification of a conceptualization [17]. Thus ontology definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean and formal axioms that constrain the interpretation and well-formed use of these terms [18], 1995). In biomedical areas, the use of ontologies as knowledge base will enables bioinformatics researchers to keep pace with processing and analyzing ever-increasing amounts of new biological data, and to draw new insights from them [4].

Personalized modelling differ to global modelling in the sense of each individual data vector (e.g. a patient in the medical area) may need an individual, local model that best fits the new data, rather than a global model, where new data are matched without taking into account any specific information about these data [15].

Personalized modeling framework for gene data analysis and biomedical applications was proposed by Kasabov et al. [16]. The framework is called Integrated Method for Personalized Modeling (IMPM) (see Figure 1) and summarized below:

P1 Data collection, data filtering, storage and update.

P2 Compiling the input vector *x* for a new person.

P3 Selecting a subset of relevant to the new sample *x* variables (features) *V*x from a global variable set *V*.

P4 Selecting a number *Kx* of samples from the global data set *D* and forming a neighbourhood *Dx* of similar samples to *x* using the variables from Vx to define the similarity.

P5 Ranking the *Vx* variables within the local neighbourhood *Dx* in order of importance to the outcome, obtaining a weight vector *Wx*.

P6 Training and optimising a local prognostic/ classification model *Mx*, that has a set of model parameters *Px*, a set of variables *Vx* and local training/testing data set *Dx*.

P7 Generating a functional profile *Fx* for the person *x* using the selected set *Vx* of variables, along with the average profiles of the samples from *Dx* that belong to different outcome classes, e.g., *Fi* and *Fj*. Performing a comparative analysis between *Fx*, *Fi* and *Fj* to define what variables from *Vx* are the most important for the person *x* that make him or her very differential from the desired class. These variables may be used to define a personalized course of health improvement.

At subsequent iterations of the method (P3-P6), the parameters *Vx* and *Kx* along with other parameters will be optimised via an optimization procedure. This is done in order to acquire an optimal or near optimal personalized model. A classification or prediction procedure is also to be used as fitness function to evaluate the performance of the candidate personalized model during the learning process. There are several classification procedures that can be utilized, such as KNN, WKNN, WWKNN, and Spiking Neural Network.

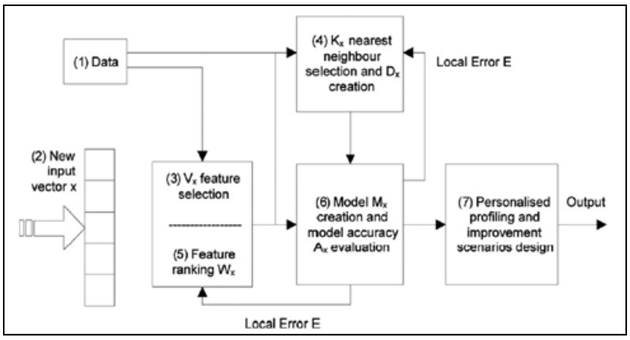


Figure.1: Functional Block Diagram of IMPM (Kasabov et al., 2010).

Kasabov et al. [15] (see Figure 2) also proposed an integrated framework of ontology knowledge bases with personalized modeling for bioinformatics decision support. That was originally based on previous paper by Gottgtroy et al. [19] (see Figure 3).

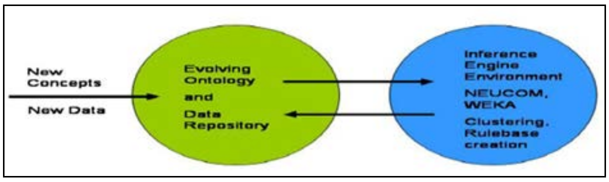


Figure 2: The ontology-based personalized decision support framework consist of two interconnected part : (i) an ontology/data base sub-system (ii) machine learning sub-system [15]

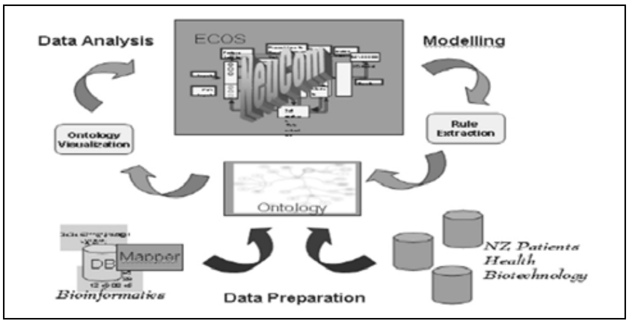


Figure 3: A sample ontology-based decision support system. The inference engine at the top utilized data retrieved from Ontology in Protégé [19]

5.2. Existing Methods for Personalized Modeling

KNN is a supervised learning algorithm that has been successfully used for classifying sets of samples based on nearest training samples in a multi-dimensional feature space, and was originally proposed by Fix and Hodges in 1951. The basic idea behind the KNN algorithm is:

- Firstly, a set of pairs features (e.g. (x1, y1), …, (xn, yn)) are defined to specify each data point, and each of those data points are identified by the class labels C = {c1, …, cn}.

- Secondly, a distance measure is chosen (e.g. Euclidean distance, or Manhattan distance) to measure the similarity of those data points based on all their features.

- Finally, the k-nearest neighbours are found for a target data point by analyzing similarity and using the majority voting rule to determine which class the target data point belongs to.

WKNN is designed based on the transductive reasoning approach, which has been widely used to evaluate the output of a model focusing on solely an individual point of a problem space using information related to this point [20]. In the WKNN algorithm, each single vector requires a local model that is able to best fit each new input vector rather than a global model, thus each those new input vector can be matched to an individual model without taking any specific information about existing vectors into account. In contrast to the KNN algorithm, the output of a new input vector is calculated not only dependent upon its k-nearest neighbor vectors, but also upon the distance between the existing vectors and the new input vector which is represented as a weight vector w, this being the basic idea behind the WKNN algorithm.

WWKNN is a novel personalized modeling algorithm which was proposed by Kasabov [14] in 2007. The basic idea behind this algorithm is: the output of each new input vector is measured not only dependent upon its k-nearest neighbors, but also upon the distance between the existing vectors and the new input vectors, and also the power of each vector which is weighted according to its importance within the sub-space (local space) to which the new input vector belongs.

The remarkable information processing capabilities of the brain have inspired numerous mathematical abstractions of biological neurons. "Neural Network" (NN) is known as artificial neural network (ANN), is defined as a hardware or software computational model that is inspired by the biological nervous systems, such as the brain, process information. Many ANN models have been successfully developed and applied across many disciplines, including classification, time series prediction, and pattern recognition, and so on [21]. However, the current ANN models do not provide good performance when applied to complex stochastic and dynamic processes such as modelling brain diseases. For that reason, new ANN models should be developed in order to become more accurate and efficient in knowledge discovery and information processing.

Wolfgang Maass [22] describes past and current neural network models into three generations. “Spiking neural networks” (SNN) is the third generation of neural network models, which is a complex and biologically plausible connectionist model. In biological neurons are connected at synapses and signals transfer information from one neuron to another. Quite a few models of SNN have been developed so far, which are: Hodgkin_Huxley's model [23]; Spike Response Models [24][25][26]; Integrate-and-Fire Models [26][27]; Izhikevich models [28][29][30], etc.

In the biological neuron, Membrane or Soma is the processing unit with the Dendrites receives the input spikes and Axon release the output from Membrane. Synapse is the point where Dendrites and Axon are met and integrated as shown below with the computational representation of a neuron:

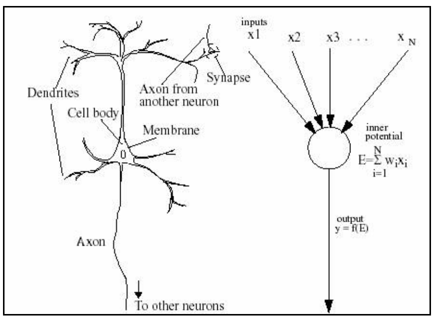


Figure 4: Biological and Computational Neuron Model [39].

SNN simulates how the biological neuron processes the information with two major characteristics. First, the weights in SNN have different strength and second characteristic is a spike is released when it reaches the exceeded threshold.

SNN has been increasingly applied in the field of science and engineering as in other disciplines to solve complicated prediction and classification problems, in recent years, SNN is becoming to be a powerful computational tool that has been widely adopted for diagnosing and monitoring the prognosis of a disease, as evidenced by over 500 published papers each year featuring neural network applications in medicine [31]. Since the introduction of spiking neuron, there have been several enhancements and variants of the Spiking Neural Network (SNN) such as Evolving Spiking Neural Network (ESNN) [32] and Extended Evolving Spiking Neural Network (EESNN) [33].

6. Proposed Framework

The framework (refer to Figure 5) consist of two major elements which is the spatial-temporal ontology system (STOS) and personalised modelling engine (PME). The general idea of this framework is developing ontology system for STD representation. The STOS will access the PME for knowledge discoveries. The outcome of the learning will be updated back into the ontology system and evolve the ontology.

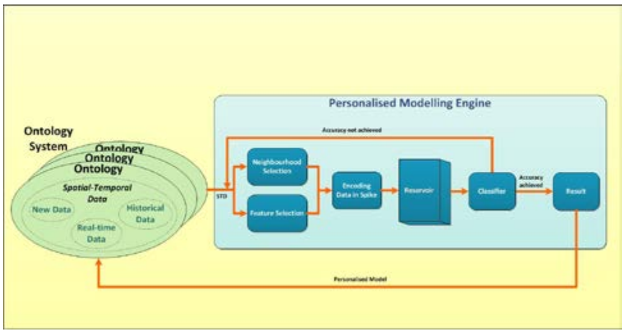


Figure 5: Spatial-Temporal Ontology System for Personalized Decision Support (STOS-PDS) Framework.

This framework contains two modules where the first module is the personalized modeling engine. This module plays important part in processing the spatial-temporal data and discovers new knowledge from the data. Based on the first case study, weather and stroke occurrences, the aims is to discover any relation between the weather features and a person that triggered a stroke apart from other contributed features such as hypertension, diabetic, cardio vascular disease, and smoking. If the engine manage to prove that there is a relation between weather and stroke occurrences than the next step is how can the engine produce a better model that could predict the best time window before the event of stroke to prevent it from happening. If this could be achieved the system could become a tool for health warning for a person. It is called personalized because the engine should be able to produce model that best fit an individual. The first classifier that is chosen for the case study is KNN and then Extended Evolving Spiking Neural Network (EESNN) method adapted from Hamed et al. [33].

The decision of choosing EESNN as the classifier lies in the attribute of the case study data. Space and time can be viewed as the important aspects of all real world phenomena. Spatio-temporal data contains information relating space and time. According to Fayyad and Grinstein [34], nowadays, approximate 80% of the available datasets have spatial components and are often related to some temporal aspects. Such data is for example environmental and audio/visual, medical, and brain signals, etc. But a great challenge so far is how to process these complex spatio-temporal data (STD).

In general, classical statistical and computational techniques are insufficient when they are applied to spatio-temporal datasets due to:

i. spatio-temporal datasets are embedded into continuous space, whereas the classical datasets are often discrete.

ii. the patterns in these datasets are often local, but classical techniques normally focus on global patterns.

iii. they model either space or time separately, or mix both components in a simple way, missing to capture essential relations between variables in the STD.

In addition, as mentioned in previous, one of my goal is to estimate the personalized risk for each individual patient. So far, existing models for estimating risk remain grounded in traditional statistical methods and in problem statements that have not evolved significantly over the years [35].

As summarised above, there is growing need for utilising computational models to estimate personalized risk in order to enhance more precise medical decision making. Recently, the concept of SNN is considered as an emerging computational technique for the analysis of spatio-temporal datasets. SNN models have been successfully utilized in several tasks, but they process input data streams as a sequence of static data vectors, ignoring the potential of SNN to simultaneously consider space and time dimensions in the input patterns. It can be viewed that SNN would be more potential and suitable for STD pattern recognition. Thus, EESNN might be an ideal novel method to produce more accurate risk prognosis for facilitating doctors to make more precise decision making to ensure patients receive optimal treatment.

The second vital module is ontology system which is the global knowledge-based to represent spatial-temporal data. This ontology system must be able to organize objectively existing data, real time data and new data. The system will be built using Protégé ontology development environment and should be integrated with existing chronic disease ontology by Verma et al. [36]. This ontology system should be used as reference engine to be queried by PME and the ontology must be able to evaluate and acknowledge new findings or relationship that the PME have discovered. The ontology will evolve further to integrate several more ontologies such as Gene Ontology and Nutritional Ontology. These ontologies will be map to each other and be linked based on the spatial-temporal data that been captured.

Nevertheless there are several challenges that need to be addressed in this research. Kasabov et al. [15] addressed several challenging issues in integrating these two elements.

i. How to use the new knowledge to further evolve the existing ontology?
    - The rank aggregation technique proposed by Domshlak et al. [37].
    - The knowledge pattern technique proposed by Clark et al. [38].

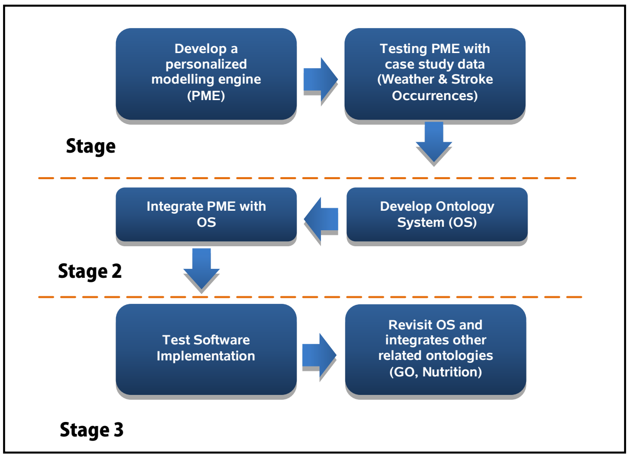
ii. How to link the ontology with machine learning engine?
    - Is specific interface needed (web based or offline)?
    - How to make sure PME obtain the right data from STOS?

iii. How can changes in a concept be detected?

iv. How should an inference engine obtain sufficient information to act in a context-aware manner?

7. Research Methodology

The research is organized in three important stages and adopt top down approach. The steps are concluded below:



8. Case Study

8.1. Dataset

The first case study comprises Weather and Stroke Occurrences dataset. The whole dataset consists of 11,453 samples (all with first-ever occurrence of stroke) from six population regions: Auckland (NZ), Perth and Melbourne (Australia), Oxfordshire (UK), Dijon (France), and Norrbotten and Vasterbotten counties (Northern Sweden). These study areas are grouped in the Southern Hemisphere region (Auckland, Perth, and Melbourne) and Northern Hemisphere region (Oxfordshire, Dijon, Norrotten and Vasterbotten counties).

Table 1: Weather and Stroke Occurrences Dataset

Region	**Frequency**	**Percent**
Auckland	2805	24.5
Dijon (France)	1756	15.3
Melbourne	1316	11.5
Oxfordshire (UK)	543	4.7
Perth	766	6.7
Sweden	4267	37.3
Total	**11453**	100.0

At the first stage of data analysis, the variables which consist of 9 features can be divided into three types of data, continuous data, categorical data and spatial-temporal data (4 patient variables and 5 weather variables) and all these weather parameters are only measured for the day of stroke occurrence:

Patient variables

- age (continuous data), gender, history of hypertension, smoke status (categorical data)

Weather variables

- temperature, humidity, wind speed, wind chill, atmospheric pressure (temporal data)

8.2. Design of the Study

Case-crossover design is applied for the data pre-processing in this dataset, due to there is no "non-stroke" patients in the original dataset.

The day of stroke occurrence until 30 days before stroke occurrence is consider as the "stroke" group and 60 days before stroke occurrence for the same participant as the "normal/control" group, due to hypothesis that weather parameters 60 days before the index stroke had no influence on the stroke occurrence 60 days later and 30 days before stroke being the critical time windows that could contribute to stroke occurrence 30 days later. The data samples are created by counting down 60 days from the day of stroke occurrence.

Patient	Temperature (Day of stroke occurrence)	temp_lag1 (1 day before)	temp_lag2 (2 days before)	temp_lag3 (3 days before)	...	temp_lag30 (30 days before)
1	22.09999084					
2	22.09999084	22.09999084				
3	21.3999939	22.09999084	22.09999084			
4	21.3999939	21.3999939	22.09999084	22.09999084		
5	21	21.3999939	21.3999939	22.09999084		
6	19.79998779	21	21.3999939	21.3999939		
7	19.59999084	19.79998779	21	21.3999939		
8	19.59999084	19.59999084	19.79998779	21		
9	20.29998779	19.59999084	19.59999084	19.79998779		
10	22.29998779	20.29998779	19.59999084	19.59999084		
11	23.19999695	22.29998779	20.29998779	19.59999084		
12	21.69999695	23.19999695	22.29998779	20.29998779		
13	21.69999695	21.69999695	23.19999695	22.29998779		
14	19.29998779	21.69999695	21.69999695	23.19999695		
15	19.29998779	19.29998779	21.69999695	21.69999695		
16	18.8999939	19.29998779	19.29998779	21.69999695		
17	20.09999084	18.8999939	19.29998779	19.29998779		
18	20.09999084	20.09999084	18.8999939	19.29998779		
19	19.69999695	20.09999084	20.09999084	18.8999939		
...	...	...	...	...	...	...

Figure 6: Auckland patient dataset (stroke and normal groups)

8.3. Preliminary analysis

Signal-to-Noise ratio analysis is used to evaluate how important a variable is to discriminate sample belonging to different classes. The sample is taken from only the Auckland region and only the weather features are measured.

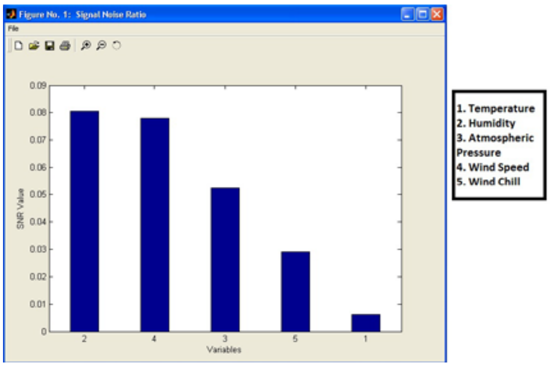


Figure 7: All features are ranked from higher priority to lower priority by using Signal-Noise-Ratio (SNR).

Another analysis method that has been used is correlation coefficient method to visualize the relationship between variables (Figure 8).

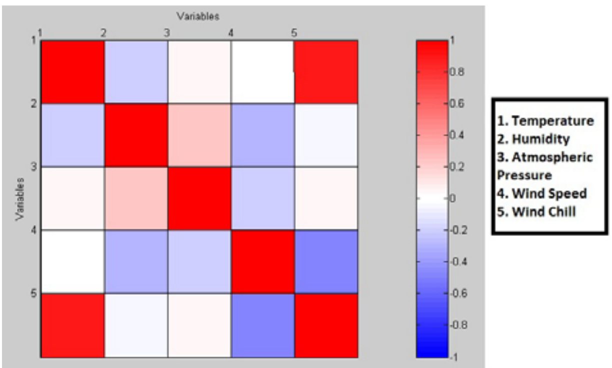


Figure 8: all features are either having a positive relation (red) or negative relation (blue).

From the SNR analysis (Figure 7) shows that humidity and wind speed seem to become much important variables than temperature and it shows that these two variables may contribute to the occurrence of stroke. As for the second analysis using correlation and coefficient method shows that humidity and wind speed have a negative relationship where it shows when the humidity reading are high, the wind speed are low. Whereas the temperature and wind chill has a positive and strong relationship because when the temperature drops, the wind chill also drops.

8.4. Experiment

The different data type intensify the complexity to analyse the data, therefore the data is clustered manually. The experiment will used KNN classifier to calculate the accuracy by clustering the patient into nearest neighbour group which has almost similar age group, same hypertension status and smoking status. Figure 9 show the hierarchical view of the data being clustered. The patients are selected only from Auckland region and autumn season consist of male subject between age of 60-69 which has the history of hypertension and smoking and result to a total of 12 patients (see Figure 9).

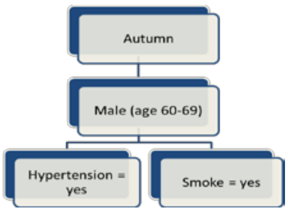


Figure 9: Selected group of patients.

The result of classification accuracy is stated in the Table 2 and the algorithm is applied on Leave-One-Out Cross-Validation (LOOCV) method. The value of *k* used is 3 and 5 nearest neighbours and this shows how KNN algorithm is not suitable for spatio-temporal data. Even a high value of *k* did not show high classification accuracy.

Table 2: Classification accuracy comparison between different values of *k* in KNN.

Model		Personalised	
		KNN (k=3)	KNN (k=5)
# Selected Features		5	5
Accuracy of Each Class (%)	Class1	50.00	66.67
	Class2	66.67	66.67
Overall Accuracy (%)		58.33	66.67

Next experiment will used EESNN method that implement the concept of SNN and reservoir.

9. Time frame of the study

No	Task/Activities	2012				2013				2014			
1.	Preliminary Study and Analysis												
2.	Design Framework for Spatial-Temporal Ontology for PDS												
3.	Design and Develop Personalized Modelling Engine (PME)												
4.	Testing and validating the PME for stroke and weather data.												
5.	Writing and Submitting D9												
6	Developing Ontology System and Integrates with PME												
7	Testing and validating the software implementation												
8	Integrating other ontology and implement it on other domain specific case studies												
9.	Conclusions / Thesis Writing												

10. Conclusion

This research is still in the preliminary stage where all possibilities are to be considered. Optimistically the following research benefits can be achieved:

i. A contribution to the bio-informatics field through the development of a theoretical foundations and practical implementations of a novel spatial-temporal ontology for personalised decision support framework.

ii. The framework will be used to develop a system that is capable of automated reasoning about dietary involvement and directed health recommendations at the level of individual.

iii. The developed web-based software will be tested and offered to end-users from different countries.

References

[1] Andrienko, G., Malerba, D., May, M. and Teisseire, M. Mining spatio-temporal data. *Journal of Intelligent Information Systems*, Volume 27, (2006), pp. 187-190.

[2] Koperski, K., Adhikary, J., and Han, J. (1996). Spatial Data Mining: Progress and challenges. *Proceedings of the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*. Montreal, Canada: 55-70.

[3] Hartmann, S., Kohler, H., Wang, J. (2010). Ontology Consolidation in Bioinformatics, *Proc. 7th Asia-Pacific Conference on Conceptual Modelling*, Brisbane, Australia.

[4] Anjum, A., Bloodsworth, P., Branson, A., Hauer, T., McClatchey, R., Munir, K., Dmitry Rogulin, Jetendr Shamdasani. (2007). The Requirements for Ontologies in Medical Data Integration: A Case Study, *11th International Database Engineering and Applications Symposium*.

[5] Abegunde, D., Beaglehole, R., Durivage, S., Epping-jordan, J., Mathers, C., Shengelia, B., Strong, K., Tukuitonga, C., Unwin, N. (2005). Preventing chronic diseases: a vital investment: *WHO global*

report. Switzerland, WHO: world Health Organization.

[6] Lumley T, Kronmal RA, Cushman M, Manolio TA, Goldstein S. (2002). A stroke prediction score in the elderly: validation and Web-based application. *Journal of Clinical Epidemiology*, 55(2):129-36.

[7] Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. (1991). Probability of stroke: a risk profile from the Framingham Study. 22(3):312-8.

[8] Yusuf HR, Giles WH, Croft JB, Anda RF, Casper ML. (1998). Impact of multiple risk factor profiles on determining cardiovascular disease risk. *Prev Med* 27(1):1-9

[9] Khosla, A., Cao, Y., Lin, C.C., Chiu, H., Hu, Y., and Lee, O. (2010). An integrated machine learning approach to stroke prediction. *In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10)*. ACM, New York, NY, USA, 183-192.

[10] Das, A., Ben-Menachem, T., Cooper, G.S., Chak, A., Sivak Jr, M.V., Gonet, J.A., Wong, R.CK., (2003). Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model, *The Lancet*, Volume 362, Pages 1261-1266.

[11] Anderson, J., Hansen, L. L., Mooren, F. C., Post, M., Hug, H., Zuse, A. et al. (2006). Methods and biomarkers for the diagnosis and prognosis of cancer and other diseases: Towards Personalised medicine. *Drug Resistance Updates*, 9 (4-5).

[12] Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N. & Roth, D. (1999). A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. modification of diet in renal disease study group. *Annals of Internal Medicine*, 130, 461-470. 4, 44

[13] Lange, M.C., Lemay, D.G., German, J.B., (2007). A multi-ontology framework to guide agriculture and food towards diet and health, *Journal of the Science of Food and Agriculture*, Volume 87, Issue 8, pages 1427–1434.

[14] Kasabov, N. (2007). Evolving connectionist systems: The knowledge engineering approach, *Springer*.

[15] Kasabov, N., Song, Q., Benuskova, L., Gottgtroy, P., Jain, V., Verma, A., Havukkala, I.,Rush, E., Pears, R., Tjahjana, A., Hu, Y., MacDonel, S. (2008). Integrating Local and Personalised Modelling with Global Ontology Knowledge Bases for Biomedical and Bioinformatics Decision Support. *Smolin, et al. (eds.) Computational Intelligence in Bioinformatics*, ch. 4, Springer, Heidelberg.

[16] Kasabov, N. and Hu, Y. (2010). Integrated Optimization method for Personalised Modeling and Case Studies for Medical Decision Support. *International Journal on Functional Informatics and Personalised Medicine*, Vol 3, No. 3.

[17] Gruber T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220.

[18] Gruber, T. R. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal Human-Computer Studies Vol. 43, Issues 5-6, p.907-928.

[19] Gottgtroy, P., Kasabov, N., Macdonell, S. (2006). Evolving Ontologies for Intelligent Decision Support. *Fuzzy Logic and The Semantic Web*, ch. 21, pp. 415–439. Elsevier, Amsterdam.

[20] Vapnik, V. (1998). Statistical learning theory. NY, Wiley-Interscience.

[21] Kasabov, N. (2010). To spike or not to spike: A probabilistic spiking neuron model. *Neural Networks*, 16-19.

[22] Maass, W. (1997) Networks of Spiking Neurons: The third generation of neural network models. *Neural Networks*, 10(9), 1659–1671.

[23] Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117, 500-544.

[24] Gerstner, W. (1995). Time structure of the activity of neural network models. *Physical Review*, 51, 738-758.

[25] Kistler, W., Gerstner, W., et al. (1997). Reduction of Hodking_Huxley equations to a single variable threshold model. *Neural Computation*, 9, 1015-1045.

[26] Gerstner, W., & Kistler, W. M. (2002). Spiking neuron models:Single neurons, populations, plasticity. *Cambridge, MA: Cambridge Univ. Press*.

[27] Maass, W., & Bishop, C. M. (Eds.). (1999). Pulsed Neural Networks. *Cambridge, MA: The MIT Press*.

[28] Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5), 1063-1070.

[29] Izhikevich, E. (2007). Dynamical systems in neuroscience. *The MIT Press*.

[30] Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 3593-3598.

[31] Gant, V., Rodway, S., & Wyatt, J. (2001). Artificial neural networks: practical considerations for clinical applications' in Dybowski R and Gant V 'Clinical applications of neural networks'. *Cambridge University Press*.

[32] Wysoski, S., & Benuskova, L. et al. (2009). Evolving spiking neural networks for audio-visual information processing. *Neural Networks*.

[33] Hamed, H.N.A., Kasabov, N. Shamsuddin, S. M,. Widiputra, H. & Dhoble, K, (2011). An Extended Evolving Spiking Neural Network Model for Spatio-Temporal Pattern Classification. *Proceedings of 2011 International Joint Conference on Neural Networks*, San Jose, CA, 2653-2656, IEEE Press.

[34] Fayyad, U. M., & Grinstein, G. G. (2001). Information visualization in data mining and knowledge discovery. *Los altos*, CA: Morgan Kaufmann.

[35] Syed, Z & Rubinfeld, I., (2010). Scaling Unsupervised Risk Stratification to Massive Clinical

Datasets. I*nternational Journal of Knowledge Discovery in Bioinformatics (IJKDB)*. Volume 2, Issue 1.

[36] Verma, A., Kasabov, N., Rush, E. & Song, Q. (2009). Ontology Based Personalized Modeling for Chronic Disease Risk Analysis: An Integrated Approach. *Advances in Neuro-Information Processing*. Lecture Notes in Computer Science. Volume 5506/2009, 1204-1210.

[37] Domshlak, C., Gal, A., Roitman, H.: Rank Aggregation for Automatic Schema Matching. *IEEE Transactions on Knowledge and Data Engineering.* 19(4), 538–553 (2007)

[38] Clark, P., Thompson, J., Porter, B.: Knowledge Patterns. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, pp. 191–208. Springer, Berlin (2004)

[39] Hemming, C. (2003). Using neural networks in linguistic resources. From:http://www.hemming.se/gslt/LingRes/NeuralNetworks.htm