Grid-based Workflow System for Chronic Disease Study

Fawaz Al-Hazemi

Electrical Engineering Department, KAIST, SOUTH KOREA
fawaz@kaist.ac.kr, fawazhazemi@ieee.org

Abstract: Chronic disease is linked to patient's' lifestyle. Therefore, doctor has to monitor his/her patient over time. This may involve reviewing many reports, finding any changes, and modifying several treatments. One solution to optimize the burden is using a visualizing tool over time such as a timeline-based visualization tool. However, there was limitation of studying the diabetes patient's history to find out what was the cause of the current development in patient's condition. In this paper, we propose a workflow system based on the Grid-based Interactive Diabetes System (GIDS) to support bioinformatics analysis application for diabetes diseases. GIDS used an agglomerative clustering algorithm as clustering correlation algorithm.
[Fawaz Al-Hazemi. **Grid-based Workflow System for Chronic Disease Study.** *Life Sci J* 2014;11(7):648-650]. (ISSN:1097-8135). http://www.lifesciencesite.com. 91

***Keywords*:** bioinformatics; data mining; data visualization

1 Introduction

One of the crucial chronic diseases is diabetes, which is taking attention of recent development in e-Healthcare system. Prior giving any decision related to a diabetes patient, a medical doctor must screen varies types of the patient's historical data, starting from blood sugar, fat, and cholesterol to HBA1c and Heart ECG; and could be more. Despite the fact that medical doctors have the essential analysis processes of any type of chronic diseases, it is complex to consolidate the analysis of all types of medical reports at once without help from automated machines. Many proposed solutions tickled chronic diseases visualization. For example, Bui et al. [1] proposed a problem centric visualization over the time. The aim of the proposed solution was finding the syndrome and adjusting the treatment plan according to the observation. Unlikely, Toyoda et al. [2] proposed tracking entry system, which was an order-entry technique to keep tracking the order-entry. Apparently, none of these solutions showed concerns on backend computation developments. Diverse historical data about a diabetes patient should be processed all at once by different CAD tools and into the same workspace. Our proposed solution is a Grid-based Interactive Diabetes System (GIDS), which is a system to automate the analysis of diabetes patients' conditions via web-enabled workflow. GIDS provides varieties of optimized timeline views about patient's condition; an agglomerative clustering algorithm controls the variation views [3]. Ultimately, doctors are able to view results of the data analysis into different scales.

2 Workflow Structure

Patient's history varied on both types and frequencies, and GIDS [4] needs a workflow system to process the study of such diversity. Therefore, we constructed a web-enabled workflow initializer that interact with user (doctors or medical researchers) via outlet called output. Figure 1 is illustrating the complete structure.

2.1 Initialization workflow

The initialization workflow is consisting of three parallel tasks namely requesting data, correlation policy, and chronological policy (or viewer setting). The first task (requesting data) is a simple data acquisition from medical database. Then, it followed with basic statistics analysis such as finding the highest distance (or variations) among data. The outcome goes to a similarity analyzer (in this work we used Gower or S15 as in [3]). However, to do this similarity subtask, a correlation policy is required. The correlation policy task is defining which data to be engaged in the similarity calculation. For example, doctor is interested in blood sugar measurements to display but he/she require engaging the variation of blood pressure or cholesterol measurements in the analysis of patient's history. The yield similarity data used by other subtask called the weigh-group average (wgpma) which is a weighing task [3]. Finally, the chronological policy is responsible to control the agglomerative algorithm parameters, which are the tuners for optimized views. The chronological policy is online controller that could be interactively (online) changed by the user and the results will be feedback in short time (depending on the backend performance computing system). All data and controls are moving across the workflow in a XML data files.

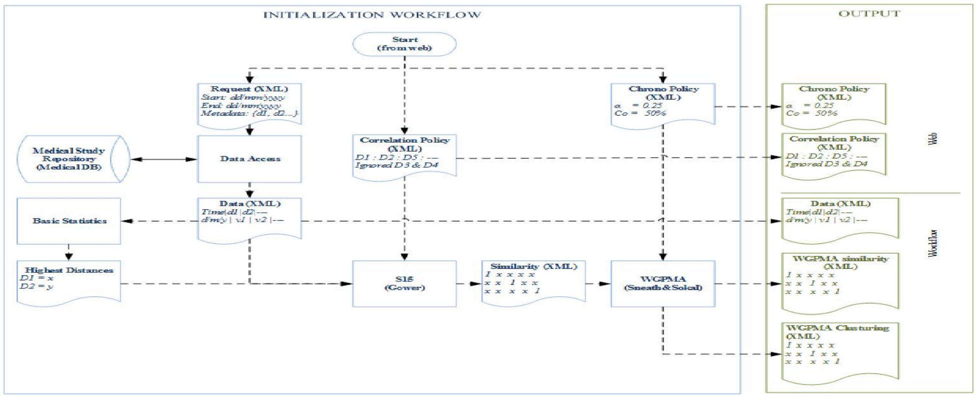
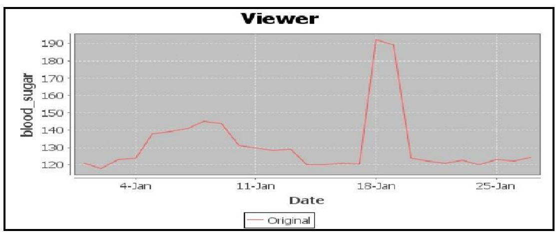
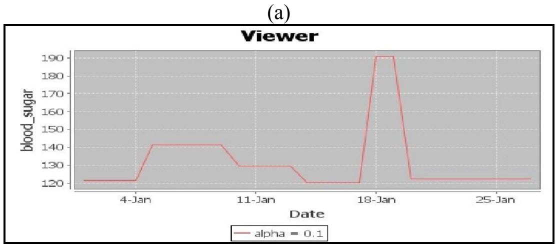


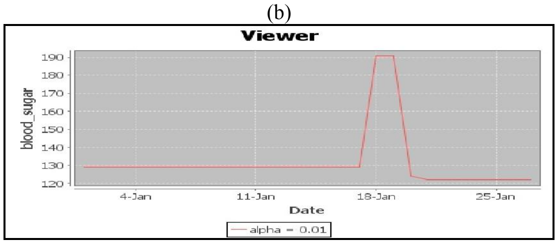
Figure 1 Workflow structure



(a)



(b)



(c)

Figure 2 Different views of patient's blood sugar meausements. (a) Original data. (b) Optimized view for personal computer's displayers. (c) Optimized view for smart devices with limited displayer.

2.2 Output

The web-enabled outlet of our workflow with GIDS is the output. It is the place for user to update the tasks chronological policy and correlation policy, and visualized the data. However, we enable the view of the outcome from the wgpma subtask, so the user of the system will see the hidden computation happened and adjust the correlation policy accordingly.

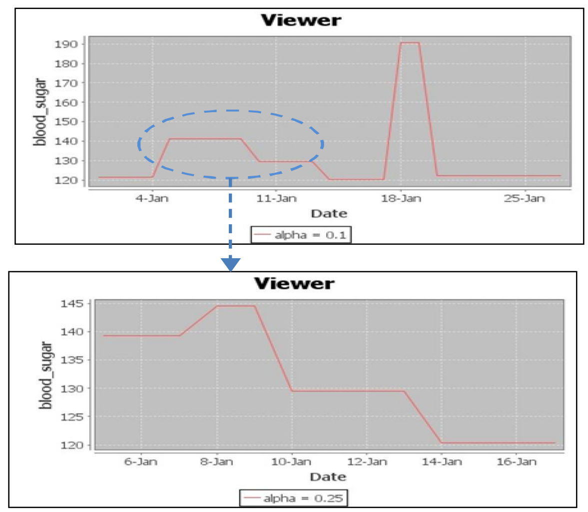


Figure 3 A zooming in example to view interested changes in patient's data between January 5th and 16th.

3 Evaluation

Our main goal of the evaluation is to examine the system capability of producing different views of patient's data. Therefore, we tested GIDS with a diabetic patient's data (less than a month data set). We captured the views generated for different displays, in particular, desktop monitor and smart phone display. The goal of this test was to display the full history of a patient's blood sugar into different displayers, and compare the effects of variation to the available display limitations. Our solution was able to provide optimized summery of patient history without hiding major variations as in Figure 2. Further, we examine the functionality of zooming into portion of displayed data for better understanding in patient's condition. Figure 3 illustrates that the functionality of zooming could show minor hiding data variations. I.e. the initial view shows step-decreases in blood sugar. Apparently, with the help of the zooming functionality, doctors are able to see the hidden rose in blood sugar on days Jan. 8th and 9th.

4 Conclusion

The main goal from the GIDS is assisting medical researchers, who are interested in studying diabetes disease. GIDS is enabling a broader view of diabetes patient's condition and predict the future implication based on current developments in some diabetes factors such as blood sugar, fat, or potassium. GIDS is using the *Chronological Clustering Algorithm* for analysis diabetes records. However, it is expandable to cover any chronic diseases that have similar motivation as the diabetes. Our future work is to migrate the Grid-based computing backend infrastructure to a Cloud-based computing backend and adding a data mining system such as Hadoop to store the chronic disease's data.

5/26/2014

References

[1] Bui, A et al. TimeLine: Visualizing Integerated Patient Records, *IEEE Transactions Information Technology in Biomedicine*, vol. 11, no. 4, 2007, pp. 462-473.

[2] Toyoda, S et al. SAKURA-Viewer: Intelligent Order History Viewer Based on Two-Viewerpoint Architecture, IEEE Transactions Information Technology in Biomedicine, vol. 11, no. 2, 2007, pp. 141-152.

[3] Legendre, P. & Legendre, L.. *Numerical ecology*. Vol. 20. Elsevier, 2012.

[4] Al Hazemi, F. et al. Grid-Based Interactive Diabetes System. First IEEE International Conference on In Healthcare Informatics, Imaging and Systems Biology (HISB), 2011, pp. 258-263.