IEEE VIS 2016
POSTERS
23-28 October, Baltimore, Maryland, USA

A Visualization System for Clustering Dementia Patients based on Automated Similarity Analysis

Hyoji Ha* Hyunwoo Han† Sungyun Bae‡ Jihye Lee§ Sunjoo Bang¶ Sangjoon Son‖
Changhyung Hong** Hyunjung Shin†† Kyungwon Lee‡‡

*†‡§Life Media Interdisciplinary Program, Ajou University
¶††Department of Industrial Engineering, Ajou University
‡‡Department of Digital Media, Ajou University
‖**Department of Psychiatry, Ajou University

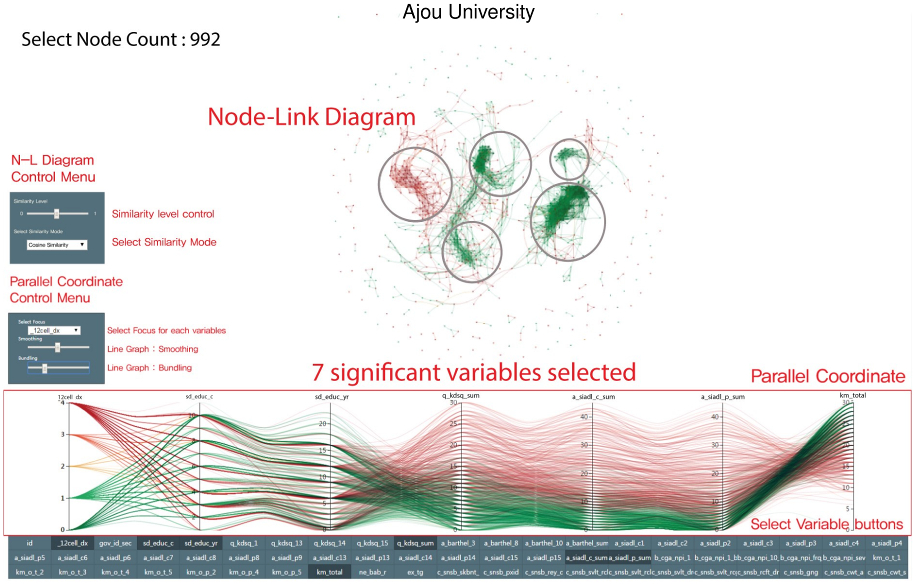

Figure 1: Patient clustering through parallel coordinate and node-link diagram, node-link diagram presented under the 7 selected and useful variables (http://foxtail.cc:2345/reportsystem/viewer?user=hyunwoo&page=9).

Abstract

This study aims to develop a visualization system based on automated similarity analysis in order to propose a method to help facilitate the diagnosis of dementia patients. This cluster visualization consists of a node-link diagram, which involves a clustering of dementia patients as they represent each node, and parallel coordinate visualization, which suggests the test results of patients in the form of line graphs. We configured this visualization after analysing the variables used when examining the patients with dementia and the actual result records obtained from CREDOS (Clinical Research Center for Dementia of South Korea). This system provides a real-time response depending on which variables are selected in the parallel coordinate, and automatically analyses similarity for the node-link diagram while changing the clustering structure and the location of the node according to its dynamic similarity. We thus discovered that it allows users to easily analyse the test progress based upon certain variables, by combining the selected variables via parallel coordinate and node-link diagram analysis. This work is expected to satisfy the growing need for a new system in which the clinician can subdivide and regroup the patients as needed, which eventually helps diagnose various aspects of dementia patients.

Keywords: Biomedical and Medical Visualization, Data Clustering, Visual Analysis Models.

*~‡‡e-mail: {hjha0508,ainatsumi,roah,alice0428,smalsunjoo,sjsonpsy, antiaging,shin,kwlee}@ajou.ac.kr

Index Terms: J.3.3 [Computer Applications]: Life and Medical Sciences—Medical information systems;

1 Introduction

Since early detection and treatment are considered very important in dementia, it is essential to design a system which enables speedy and diversified clinical judgements and systematically analyzes different factors related to the onset of dementia. However, it is difficult to provide details about the patient status using the current dementia diagnosis indicators, since it is rather widely and simply categorized as Alzheimer's disease, vascular dementia and the relative seriousness of illness[1]. Therefore, patients classified into the same progression stage based on the current diagnostic indicators might have noticeable differences in their psychological or neurological test results. As an alternative approach, we aim to develop a clustering method that can be visualized to subdivide various groups of dementia patients according to their progression. It is expected that the system will help clinicians to consider various options for dementia diagnosis by directly interacting with the system. We first refined the data including patient information, psychological testing information and neurologic examination obtained from the CREDOS cohort in oder to visualize the fundamental information. We further identified the diagnosis result patterns of the refined data using parallel coordinate visualization as a method for analysing multi-dimensional data. Next, we developed a module to calculate the similarity using various methods to form a node-link diagram. The node-link diagram can be visualized. By applying this module, when a user selects certain variables, the node-link diagram can be drawn based upon the automatically calculated similarity. Our visualization can indicate multiple clusters from the results of the parallel coordinate and node-link diagram. Finally, we verified its improved effectiveness when this system was applied to real-life clinical diagnoses.

Table 1: Data type included in the CREDOS Cohort (436 variables)

Patient information	Government ID, Cohort ID, Gender, Age, Level of education
Caregiver information	Caregiver personal information
Psychological testing information	KDSQ(Korean Dementia Screening Ques-tionnaire), CGA-NPI(Caregiver-Administered Neuropsychiatric Inventory), S-IADL(Seoul-Instrumental Activities of Daily Living), CDR(Clinical Dementia Rating Scale), K-MMSE(Korean Version of Mini-Mental State Examination), SNSB(Seoul Neuropsychological Screening Battery)
Neurologic examination	MRI data (Brain volume), Neurological examination for dementia

2 DATA ANALYSIS

We used the dementia examination data from Korean elderly people from a cohort called CREDOS. It includes 21094 health records and 436 factors in the dataset. Table 1 shows the types of variables included in CREDOS. This data was collected from 37 hospitals in South Korea. After sorting the data by ruling out the cases with missing factors or duplicate government ID, we selected about 1000 patients via stratified sampling.

3 VISUALIZATION PROPOSAL

It is also important to combine multiple visualization types in order to develop a visual analytics system with multiple coordinated views for multi-dimensional analysis of various variables[2]. We elaborated parallel coordinate visualization and node-link diagram for dementia patients, with reference to the existing research regarding scientific data based visual analytics[3].

3.1 Parallel coordinate visualization

Each dimension in the parallel coordinate indicates the variables from CREDOS data such as personal information and test scores of patients, each line indicating one patient. Also, each line was colored to suggest the variable 12cell-dx, which illustrates dementia progression staging. Depending on the diagnosis result, it is presented as dark green (SMI: favorable), green (MCI: rather favorable), yellow (VCI: neutral), scarlet (SVD: rather dangerous), and red (AD: very dangerous). Dimensions in the parallel coordinate can be set as needed by the user among variables such as personal information, psychological examination or neurological examination.

3.2 Node-link diagram based on automated similarity analysis

In this research, the node-link diagram is constructed such that each test result is represented after the similarity is calculated according to the variables selected in the parallel coordinate. Since the system provides options such as cosine similarity, Spearman correlation similarity and Pearson correlation coefficient similarity as well as similarity adjustments (from 0 to 1), the user can freely balance the values in order to draw a node-link diagram to examine the similarities. One can also subdivide the clusters of nodes within the node-link diagram according to the variables selected in the parallel coordinate. Colors of the node represent the same indications as shown by each color in the parallel coordinate.

4 IMPLEMENTATION: CLUSTERING ANALYSIS

Figure 2 suggests how the node-link diagram will appear after selecting all 23 variables including KDSQ, CGA-NPI, Barthel-IADL, S-IADL, GDS, CDR, K-MMSE, SNSB, KSF-GDS and identity, all of which are related to the psychological examination results. On the other hand, Figure 1 on the other hand shows the node-link diagram represented with 7 variables after selecting K-MMSE, SIADL and level of education, which were the factors considered useful by psychiatrists. The node-link diagram suggested in Figure 2 shows that SMI and MCI gather as one cluster while VCIs are dispersed on the outskirts, and finally, SVD and AD contrast with green tone nodes act as a different cluster.

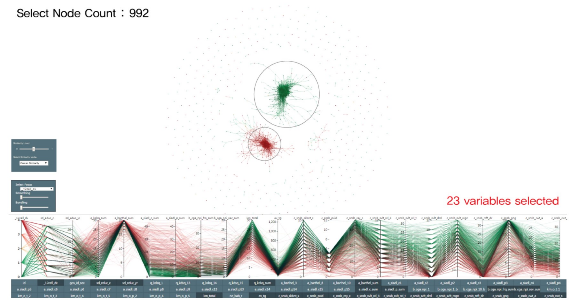

Figure 2: Node-link diagram presented under all the variables related to psychological examination.

However, the node-link diagram shown in Figure 1 demonstrates that a group of SMIs and MCIs is not assembled as one cluster but it is broken down into four groups, as well as SVD and AD which formed three smaller groups. It is also observed that VCI nodes blend with the other nodes, unlike the left figure. We thus can discover that, although the nodes represent the same progression staging, groupings can appear differently if the test results are numerically different. This holds forth the possibility that the existing progression staging which ranges from SMI, MCI, VCI, and SVD to AD can be better organized, and can also prevent patients from being classified incorrectly. It is also expected to help the clinician in personalizing a variety of treatment methods for each patient as the groups they will be classified into are smaller with precise details.

5 DISCUSSION

We next conducted usability evaluation among psychiatrists in order to verify our system. After the psychiatrists had experienced the visualization system for about 45 minutes to one hour, we performed a survey with questions related to 'Usability as a diagnosis tool', 'General understanding of the visualization system', 'Clarity of information delivery' 'Functionality' and 'Overall ease of use', based on a 7-point Likert scale. Since answers to all five questions indicated that satisfaction was above average, we could infer that this visualization system can be appropriately applied as a clinical diagnosis method.

6 CONCLUSION

In this study, we proposed a system that can update the location and clustering of nodes in a node-link diagram when the user selects the desired variable in the parallel coordinate, depending on the variable combination. As a result, our visualization method showed that it is more effective if the essential variables for psychological examination are used rather than all the variables, and that the system is highly likely to be an appropriate alternative tool for clinicians to make an accurate dementia stage diagnosis. We further plan to collect and refine the data on different diseases in addition to dementia, and thereby devise a medical diagnostic tool based upon visual analytics.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea_(NRF) Grant funded by the Korean Government_(MSIP) (No.2015R1A5A7037630) and supported under 2016 BK21 Pro-gram by Ajou University.

REFERENCES

[1] H.Braak and E.Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol (Berl), 82(4)*, 39–59, 1991.

[2] S.Cheng and K.Mueller. Improving the Fidelity of Contextual Data Layouts Using a Generalized Barycentric Coordinates Framework. *IEEE PacificVis, Hangzhou, China*, 295–302, 2015.

[3] H.Ding, C.Wang, K.Huang, R.Machiraju. iGPSe: A visual analytic system for integrative genomic based cancer patient stratification. *BMC bioinformatics, 15(1)*, 203, 2014.