International Journal of Medical Informatics 107 (2017) 18–29

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Subject-enabled analytics model on measurement statistics in health risk expert system for public health informatics

Chi-Jung Chung[a,b], Yu-Chen Kuo[c], Yun-Yu Hsieh[a], Tsai-Chung Li[d,e], Cheng-Chieh Lin[f,g], Wen-Miin Liang[d], Li-Na Liao[g], Chia-Ing Li[b], Hsueh-Chun Lin[a,h,*]

[a] Department of Health Risk Management, College of Public Health, China Medical University, Taichung, Taiwan
[b] Department of Medical Research, China Medical University Hospital, Taichung, Taiwan
[c] Institute of Information System and Applications, National Tsing Hua University, Hsinchu, Taiwan
[d] Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan
[e] Department of Healthcare Administration, College of Medical and Health Science, Asia University, Taichung, Taiwan
[f] Department of Family Medicine, China Medical University Hospital, Taichung, Taiwan
[g] School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan
[h] Department of Health Services Administration, College of Public Health, China Medical University, 91 Hsueh-Shih Rd., Taichung 40402, Taiwan

ARTICLE INFO

Keywords:
Open source
Health risk information
R software
Health measurement data warehouse
Health risk expert system
Process automation

ABSTRACT

Purpose: This study applied open source technology to establish a subject-enabled analytics model that can enhance measurement statistics of case studies with the public health data in cloud computing.

Methods: The infrastructure of the proposed model comprises three domains: 1) the health measurement data warehouse (HMDW) for the case study repository, 2) the self-developed modules of online health risk information statistics (HRIStat) for cloud computing, and 3) the prototype of a Web-based process automation system in statistics (PASIS) for the health risk assessment of case studies with subject-enabled evaluation. The system design employed freeware including Java applications, MySQL, and R packages to drive a health risk expert system (HRES). In the design, the HRIStat modules enforce the typical analytics methods for biomedical statistics, and the PASIS interfaces enable process automation of the HRES for cloud computing. The Web-based model supports both modes, step-by-step analysis and auto-computing process, respectively for preliminary evaluation and real time computation.

Results: The proposed model was evaluated by computing prior researches in relation to the epidemiological measurement of diseases that were caused by either heavy metal exposures in the environment or clinical complications in hospital. The simulation validity was approved by the commercial statistics software. The model was installed in a stand-alone computer and in a cloud-server workstation to verify computing performance for a data amount of more than 230 K sets. Both setups reached efficiency of about 10^5 sets per second.

Conclusions: The Web-based PASIS interface can be used for cloud computing, and the HRIStat module can be flexibly expanded with advanced subjects for measurement statistics. The analytics procedure of the HRES prototype is capable of providing assessment criteria prior to estimating the potential risk to public health.

1. Introduction

Many prior studies of clinical medicine and public health have explored the risk factors of diseases by reviewing medical records in clinics and assessing the health data of patients. These researches employed biostatistics algorithms to analyze biomarkers and environmental factors in correlation to diseases to control and prevention [1]. The analytics used to drive regression methods for public health studies, which may include chronical disease trace, epidemiology surveys, and environmental exposure monitoring, have been used to formulate empirical equations corresponding to the indicators of potential risks [2]. In an Expert System (ES), the feedback pattern according to computer-aided technology can be formulated to support the health risk assessment for regular healthcare interventions including pathological habits, unhealthy diets, and unsafe exposures [3].

In the past couple decades, electronic health (e-Health) services have been widely promoted in the modern hospitals. The current Hospital Information System (HIS) integrates the medical records with

* Corresponding author at: Department of Health Risk Management and Department of Health Services Administration, College of Public Health, China Medical University, 91 Hsueh-Shih Rd., Taichung 40402, Taiwan.
E-mail address: snowlin@mail.cmu.edu.tw (H.-C. Lin).

http://dx.doi.org/10.1016/j.ijmedinf.2017.08.011
Received 15 April 2017; Received in revised form 15 August 2017; Accepted 31 August 2017

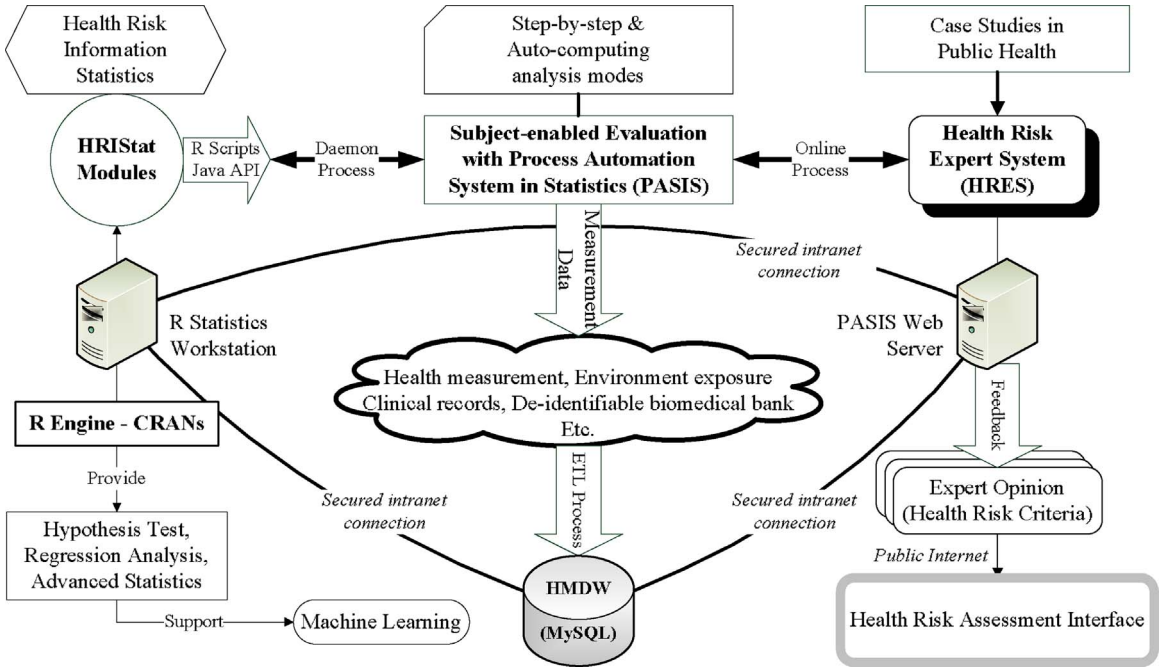

Fig. 1. Infrastructure of the health risk expert system within a secured network.

patient intervention to generate a data warehouse, and it delivers the statistical outcomes by evaluating multiple variables correlated to laboratory experiments, healthcare management, environmental monitoring, etc. [4,5]. Thus the HIS can cooperate with an ES, which collaborates with intelligent analysis and expert opinions, to enable the capability to support clinical decisions in cloud computing [6–8]. In practice, most of these systems have been solely developed by teams in hospitals, and their data analyses are conducted by authorized persons only [9]. In addition, the statistic subjects in similar categories are usually calculated by repeating a routine procedure with the same methodology. In this big data era, data mining models have been further installed inside the core of the ES to formulate patterns that can be categorized for the risk criteria of health management in case studies of medical care [10]. Therefore, the algorithms for discovering rapidly-growing complex data on various subjects require an autonomous mechanism in cloud computing in order to produce the criteria for efficiently assessing the health risks [11,12].

From the perspective of medical informatics, the mature algorithms of statistics are developed for subject-enabled analytics. According to general statistical procedure, collective data can be classified by categorical and continuous types of variables. Then the function of an analytical subject can be derived by a variety of statistics algorithms. For instance, a subject with categorical variables needs Chi-square test; a subject with continuous variables requires Pearson's or Spearman's correlation test; and a subject with both categorical and continuous variables can employ independent T-test, Wilcoxon rank-sum test, ANOVA test, or Kruskal-Wallis test. Subsequently, the significant variables can be substituted into either linear or Logistic regression to explorer their correlations [13]. Additionally, advanced analysis can be performed on the specific models for discovering risk factors in multivariable formulations (e.g. Cox proportional hazard model [14]), or correlative components in a data matrix (e.g. principal component analysis [15]), etc. The commercial statistical tools such as SAS™, STATA™, or SPSS™ have been commonly employed in health services research [16]. Such software offers complete statistics modules to help computing above subjects, but they come with extra cost as researchers need to purchase their licenses. With growing internet technology, freeware can provide helpful services for research and development. Open source technique does not follow profit motivations and shares hierarchical co-ordination experiences by experts based on universal enterprise standards [17,18]. The popular open-source utilities (i.e. freeware) such as Java, R, MySQL database can be installed on a cross-platform operating system [19]. The freeware offers application program interface (API) and functional packages, which allow users to develop extensive modules for their academic research requirements. Thus, instead of purchasing fully-loaded software, a subject-enabled ES that can be flexibly used to integrate analytics APIs into an intelligent system is adequate for the scope of general research scopes [20–22].

In this study, we established a platform for a Health Risk Expert System (HRES) with a subject-enabled analytics model, which integrates self-developed health risk information statistics (HRIStat) modules into a Web-based process automation system in statistics (PASIS). Our model enables the proper statistical subjects for measurement statistics and risk assessment of general case studies with public health data in cloud computing. Two prior case studies that acquired epidemiologic data from local areas in Taiwan were used to evaluate the efficiency and validity of our system. The design and development of our proposed model is detailed in the Methods Section below. The installation and evaluation of our system is found in the Results Section. Finally, the approaches taken in this study are commented upon in the Discussion and Conclusion Sections.

2. Methods

The architecture of the proposed HRES consists of three important domains: 1) a Health Measurement Data Warehouse (HMDW) for case studies, 2) subject-enabled modules of Health Risk Information Statistics ("HRIStat") for cloud computing, and 3) an evaluation interface of a Web-based process automation system in statistics ("PASIS") for case studies with health risk assessment. The self-developed PASIS interface and HRIStat modules were constructed using open-source Java technology and R statistics freeware, respectively, while the HMDW was constructed using MySQL database. The infrastructure of the HRES is illustrated in Fig. 1. In the figure, data transportation among the PASIS, HRIStat, and HMDW is conducted via safe intranet that is secured inside the firewall of an academic network. The online and daemon data transformation in the computing process is in compliance with web service (WS) standards to control the assessment for