JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 20, No. 8 : 1460~1465 Aug. 2016

의료 빅데이터를 활용한 CRM 기반 건강예보모형 설계

이상원[1] · 신성윤[2*]

Design of Health Warning Model on the Basis of CRM by use of Health Big Data

Sangwon Lee[1] · Seong-Yoon Shin[2*]

[1]Division of Information & Electronic Commerce, Wonkwang University, Iksan 54538, Korea
[2*]School of Computer Information & Communication Engineering, Kunsan National University, Kunsan 54150, Korea

요 약

오늘날 많은 비용이 국가 의료보장체계의 유지를 위협하고 있다. 국가 질병 통제 및 방지 센터의 감사체계를 동반한 건강관리 역학성에 대한 연구에도 불구하고, 시간 한계, 표본 한계, 대상 질병 한계에 대한 제약이 여전히 존재하고 있다. 이러한 배경에서, 방대한 양의 전수 데이터를 활용하여, 많은 기술들이 건강의 선제적 예측이나 그 대상 질병을 확장하는 분야에 충분하게 적용되고 있다. 우리는 국민건강보험의 구조적 데이터와 소셜네트워크서비스의 비구조적 데이터를 활용하여 질병을 예측하는 모형을 설계하였다. 이 모형은 건강예보서비스를 제공함으로써, 국민건강을 증진시키고 사회적 혜택을 극대화할 수 있다. 또한, 빅데이터 분석에 근거하여, 건강보험비용의 갑작스러운 증가를 감소시키거나 적시적인 질병발생을 예측할 수도 있다. 관련된 의료 예측 사례를 살펴보았고, 제안된 모형의 검증을 위하여 시범과제를 통한 실험을 수행하였다.

ABSTRACT

Lots of costs threaten the sustainability of the national health-guarantee system. Despite research by the national center for disease control and prevention on health care dynamics with its auditing systems, there are still restrictions of time limitation, sample limitation, and, target diseases limitation. Against this backdrop, using huge volume of total data, many technologies could be fully adopted to the preliminary forecasting and its target-disease expanding of health. With structured data from the national health insurance and unstructured data from the social network service, we attempted to design a model to predict disease. The model can enhance national health and maximize social benefit by providing a health warning service. Also, the model can reduce the advent increase of national health cost and predict timely disease occurrence based on Big Data analysis. We researched related medical prediction cases and performed an experiment with a pilot project so as to verify the proposed model.

키워드 : 의료, 예측, 빅데이터, 고객관계관리

Key word : Health, Prediction, Big Data, Customer Relationship Management

Received 19 April 2016, **Revised** 21 April 2016, **Accepted** 14 June 2016

* **Corresponding Author** Seong-Yoon Shin(E-mail: s3397220@kunsan.ac.kr, Tel:+82-63-469-4860)
School of Computer Information & Communication Engineering, Kunsan National University, Kunsan 54150, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2016.20.8.1460 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. INTRODUCTION

Incidences of various infectious diseases such as measles, avian influenza, novel influenza, severe acute respiratory syndrome, Middle East respiratory syndrome, and so on are increasing. This situation causes public unease and social loss. Also, as the income level increases, the national health insurance coverage is expanding. With the constant advances in health technology, medical health costs are increasing dramatically (Fig. 1). Against this backdrop, despite research by the national center for disease control and prevention on health care dynamics with its auditing systems, there are still restrictions of time limitation, sample limitation, target diseases limitation, and so one. The technologies to predict diseases on the basis of Big Data analysis focus on personally customized analysis from the viewpoint of customer relationship management (CRM). Using a huge volume of data, the technologies could be fully adapted for preliminary forecasting and the expansion of target diseases. Our proposed model for disease prediction [1] focuses on the enhancement of national health and the maximization of social benefit through a health warning service. The model could reduce the advent increase of national health cost and predict timely disease occurrence based on Big Data analysis.

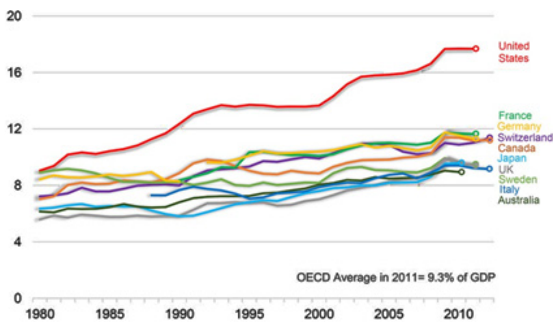

Fig. 1 Health Care Spending as Percentage of GDP (OECD)

Ⅱ. Related Works

There are some health warning models using Big Data; Medisys on its new website, Sickweather on social media such as Twitter or Facebook, and Google influenza trends on key word search.

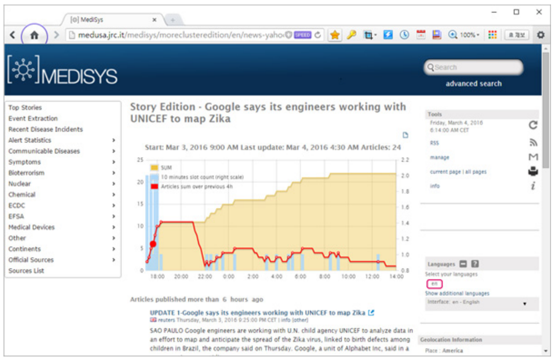

Fig. 2 Medisys

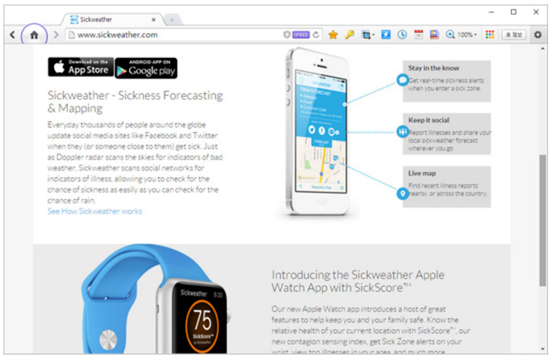

Fig. 3 Sickweather

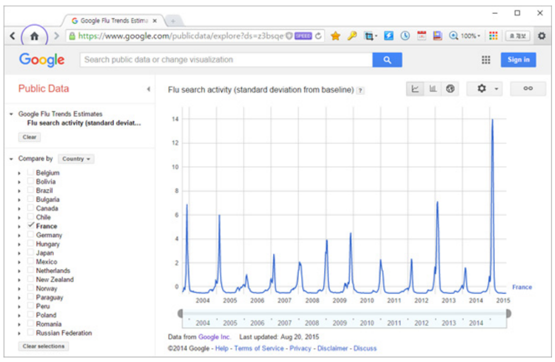

Fig. 4 Google Flu Trends

Medisys' wide range of preventive health services (Fig. 2) supports more than an ounce of prevention. With its Preventive Health Assessments, medical concierge service, pediatric clinics and genetic testing, a person has access to everything he and she needs to be proactive about his or her health and the health of the family. Sickweather (Fig. 3) provides various services (disease live map, alert, widget, and so one). Similar to the concept of scanning the skies for indicators of bad weather, the service scans social networks [2-4] for indicators of illness, allowing a person to check for the chance of sickness as easily as checking for the chance of rain. Google operates a web service, namely Google Flu Trends (Fig. 4) that provides influenza activity estimates for more than 25 countries. By aggregating Google search queries, the service attempts to make accurate predictions about flu activity.

Ⅲ. HWM-CRM Model

We propose a customized disease prediction model, termed the Health Warning Model on Customer Relationship Management (HWM-CRM). We use several technologies to predict diseases on the basis of Big Data analysis, while focusing on personally customized analysis from the viewpoint of customer relationship management (CRM) [5-7]. The model handles two dimensional Big Data; one is the huge volume of citizen data from national health insurance and the other is the huge volume of crawled social network service data. The model supports macroscopic medical trends for the nation and individually customized medical information for each citizen. The service develops a disease prediction model and provides health warnings by monitoring danger signs. For the service, both the database for national health insurance and social media information are integrated. The HWM-CRM model adopts the collection and analysis technologies of verified Big Data in order to guarantee the quality of analysis. To develop a customer-oriented service, an advisory panel of related experts was organized.

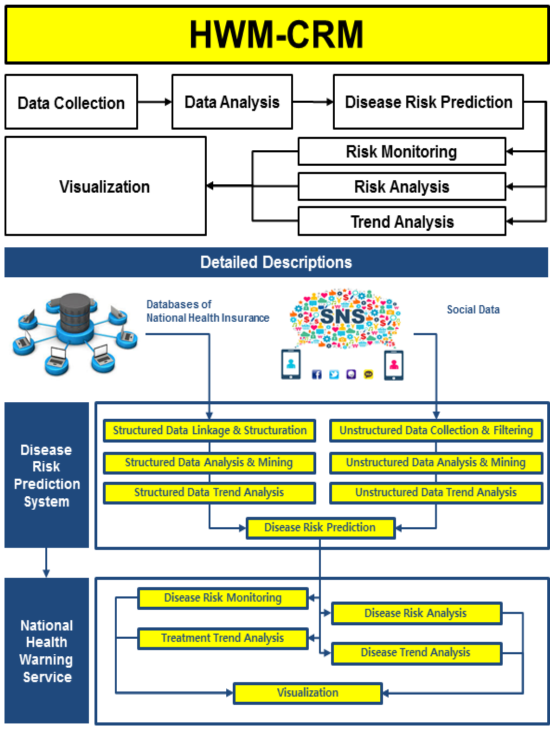

Fig. 5 Architecture of HWM-CRM

3.1. Architecture of HWM-CRM Model

HWM-CRM has an architecture (Fig. 5) with two major parts: a disease risk prediction system and a national health warning service. The disease risk prediction system is composed of two dimensions: structured analysis and unstructured analysis. The structured analysis part uses national health insurance databases and performs several tasks such as structured data linkage & structuration, structured data analysis & mining, and structured data trend analysis. The unstructured analysis part uses social data and performs several tasks such as unstructured data collection & filtering, unstructured data analysis & mining, and unstructured data trend analysis. The two analysis parts result in disease risk prediction and the results are transmitted to the national health warning service.

3.2. Role of HWM-CRM Model

The HWM-CRM model is generated by use of integrated data, i.e., structured data and unstructured data.

3.3. Development of HWM-CRM Model

The HWM-CRM model supports all citizens with health warning information and visualized services. The model is composed of four major processes; target disease selection, knowledge implementation, collection/mining, and prediction analysis; (1) target disease selection, (2) knowledge implementation, (3) collection/mining [8, 9], (4) prediction analysis. The results of the four processes are visualized in various forms (disease treatment trend, disease social trend, and disease risk).

Ⅳ. Experiment of HWM-CRM Model

Even though there were too many disease prediction models such as Sickweather on social media such as Twitter or Facebook, and Google influenza trends on key word search, the models used only social data with low accuracy of prediction. Our proposed HWM-CRM considers both structured data and unstructured social data so as to increase the disease accuracy rather than other existing social-dependent models. This attempt is dependent on structured operational data with referring to unstructured social data with more developed Big Data technology. The model would certainly overcome existing limitations of epidemiological survey method only on the basis of operational National Health Insurance databases, and then improve the accuracy or speed of disease prediction by use of total data on the basis of Big Data analysis technologies. We performed an experiment to verify our HWM-CRM model. The experiment involved four major processes; target disease selection, knowledge implementation, collection/ mining, and prediction analysis.

4.1. Disease Selection Process

There were 22 diseases of the big class-level, 267 diseases of the medium class-level, and 2,093 diseases of the small class-level. We analyzed 1,000 diseases with multiple frequencies and then selected 100 disease candidates. After deducing the degree of fluctuation for the monthly average of the diseases, we confirmed 30 disease candidates. The degree of fluctuation was calculated with the formula ((the spot month - the last month) / the last month) * 100). Secondly, after researching the degree of fluctuation for the monthly average including social network analysis of data buzz trends, we finally confirmed 6 target diseases. The 6 target diseases included influenza, allergic dermatitis, food poisoning, eye infection, hypertrophic rhinitis, and asthma. The target diseases was decided social concerns that were mentioned on social data.

4.2. Knowledge Implementation Process

With 6 target diseases, we established the knowledge classification system. The knowledge classification was performed with the mapping method of vocabulary modeling.

4.3. Collection/Mining Process

We performed data mining with the collected data. The investigation period was from January 1 to December 31 in 2015. We raked the texts related to the 6 target diseases on Twitter and Facebook. Table 1 shows an example of the data mining process on influenza data.

4.4. Prediction Analysis Process

In this process, we conducted a time series prediction and performed risk modeling. In the time series prediction, we designed an auto-regressive moving average model and predicted treatment information. In the risk modeling, we induced a time series model for each risk level (concern, caution, alert, and danger). Fig. 6 shows the gap between the predicted and actual values of allergic dermatitis, and Fig. 7 shows the risk level of allergic dermatitis by the month from 2014 to 2015.

Table. 1 An Example of Influenza Data

Frequency by the Month

Month	Total	Disease1	Disease2	Disease3	Disease4	Disease5	Disease6
January	20,507	35	53	32	7,983	6,767	5,637
February	13,409	23	13	3	4,039	7,599	1,732
March	12,936	44	23	87	7,362	2,583	2,837
April	8,211	33	44	36	3,873	3,923	302
May	8,207	46	303	33	2,736	2,736	2,353
June	2,925	64	36	23	736	1,536	530
July	8,185	36	64	36	3,203	3,023	1,823
August	12,556	265	15	74	1,438	8,731	2,033
September	3,799	75	65	36	2,583	103	937
October	4,571	23	23	33	3,230	918	344
November	2,897	153	44	90	1,303	308	999
December	15,077	9	33	130	9,206	5,510	189
Sum	113,280	806	716	613	47,692	43,737	19,716

Frequency by the Gender

Gender	Total	Disease1	Disease2	Disease3	Disease4	Disease5	Disease6
Male	90,364	353	495	280	44,449	40,394	4,393
Female	22,916	453	221	333	3,243	3,343	15,323
Sum	113,280	806	716	613	47,692	43,737	19,716

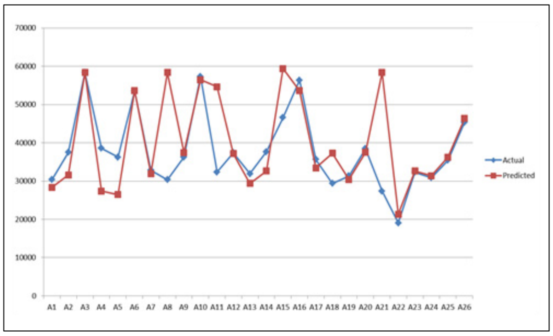

Fig. 6 An Example of the Gap between Predicted and Actual Values of Allergic Dermatitis

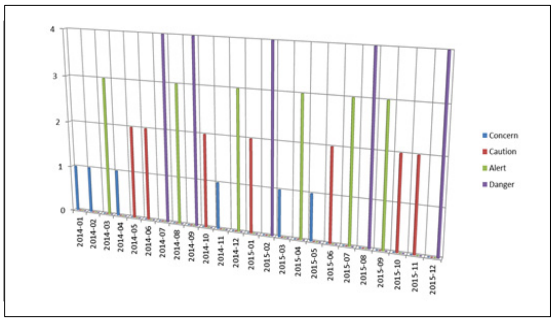

Fig. 7 An Example of the Risk Level of Allergic Dermatitis

V. CONCLUSIONS

We proposed a health warning model for customer relationship management (HWM-CRM) using Big Data. After researching related cases of health prediction involving Big Data analysis, we performed a pilot experiment and verified the results. In order to accelerate this model, we will develop a service for nation-wide health promotion. Since we proposed and developed the model, HWM-CRM, we plan to customize it for each citizen. Other service platforms such as mobile applications should be considered in the near future.

REFERENCES

[1] F. M. Rafiei, S. M. Manzari, and S. Bostanian, "Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10210-1021, Aug. 2011.

[2] S. P. Borgatti, and X. Li. "On social network analysis in a supply chain context," *Journal of Supply Chain Management*, vol. 45, no. 2, pp. 5-22, Mar. 2009.

[3] A. Marin, and B. Wellman, "Social network analysis: An introduction," *The SAGE Handbook of Social Network Analysis*, pp. 11-25, 2011.

[4] O. Serrat, *Social Network Analysis*, Washington, DC: Asian Development Bank, 2009.

[5] S. Knox, A. Payne, L. Ryals, S. Maklan, and J. Peppard, *Customer Relationship Management*, Routledge, 2007.

[6] V. Kumar, *Customer Relationship Management*, Online, John Wiley & Sons, Ltd., 2010.

[7] A. Payne, and P. Frow, "A strategic framework for customer relationship management," *Journal of Marketing*, vol. 69, no. 4, pp. 167-176, Oct. 2005.

[8] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," in *Proceeding of the International Conference on Very Large Data Bases*, Santiago, Chile, pp. 487-499, 1994.

[9] M. W. Berry, S. T. Dumais, and G. W. O'Brien, "Using linear algebra for intelligent information retrieval," *SIAM Review*, vol. 37, no. 4, pp. 573-595, Dec. 1995.

이상원(Sangwon Lee)

한국과학기술원(KAIST) 경영공학 박사
원광대학교 정보전자상거래학부 부교수
※관심분야 : 데이터설계, 데이터분석, 경영정보시스템

신성윤(Seong-Yoon Shin)

군산대학교 컴퓨터과학과 박사
한국정보통신학회 국문지부회장
한국정보통신학회 국제학술부회장
군산대학교 컴퓨터정보통신공학부 교수
※관심분야 : 멀티미디어 시스템 및 응용, 가상현실, 텔레메틱스