International Journal of Medical Informatics 92 (2016) 35–43

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com

The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system

Eldon Y. Li [a,*], Chen-Yuan Tung [b], Shu-Hsun Chang [c]

[a] Department of Management Information Systems, National Chengchi University, Taipei City 11605, Taiwan, ROC
[b] Graduate Institute of Development Studies, National Chengchi University, Taipei City 11605, Taiwan, ROC
[c] Department of Management Information Systems, National Chengchi University, Taipei City 11605, Taiwan, ROC

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form 8 March 2016
Accepted 26 April 2016

Keywords:
Prediction market system
Logarithmic market scoring rules
Infectious diseases
Epidemic prediction
Real-time update
Web-based system
Wisdom of crowds

ABSTRACT

Background: The quest for an effective system capable of monitoring and predicting the trends of epidemic diseases is a critical issue for communities worldwide. With the prevalence of Internet access, more and more researchers today are using data from both search engines and social media to improve the prediction accuracy. In particular, a prediction market system (PMS) exploits the wisdom of crowds on the Internet to effectively accomplish relatively high accuracy.
Objective: This study presents the architecture of a PMS and demonstrates the matching mechanism of logarithmic market scoring rules. The system was implemented to predict infectious diseases in Taiwan with the wisdom of crowds in order to improve the accuracy of epidemic forecasting.
Methods: The PMS architecture contains three design components: database clusters, market engine, and Web applications. The system accumulated knowledge from 126 health professionals for 31 weeks to predict five disease indicators: the confirmed cases of dengue fever, the confirmed cases of severe and complicated influenza, the rate of enterovirus infections, the rate of influenza-like illnesses, and the confirmed cases of severe and complicated enterovirus infection.
Results: Based on the winning ratio, the PMS predicts the trends of three out of five disease indicators more accurately than does the existing system that uses the five-year average values of historical data for the same weeks. In addition, the PMS with the matching mechanism of logarithmic market scoring rules is easy to understand for health professionals and applicable to predict all the five disease indicators.
Conclusions: The PMS architecture of this study affords organizations and individuals to implement it for various purposes in our society. The system can continuously update the data and improve prediction accuracy in monitoring and forecasting the trends of epidemic diseases. Future researchers could replicate and apply the PMS demonstrated in this study to more infectious diseases and wider geographical areas, especially the under-developed countries across Asia and Africa.

1. Introduction

Today's communication networks, computing power, and intelligent algorithms have enabled us to perform with multiple agents voting and rank aggregation, task and resource allocation, kidney exchanges, auctions and exchanges, charitable giving, and prediction markets over the Web [12]. Early in 1988, researchers in University of Iowa introduced a prediction market system (PMS), Iowa Electronic Markets, to predict the outcome of the 1988 U.S. presidential election. Since then PMSs have drawn much attention in betting political events [20,55,70,18,58]. Outside the political arena, a host of PMSs have been introduced such as HedgeStreet Exchange (www.nadex.com), Hollywood Stock Exchange (www.hsx.com), iPredict (www.ipredict.co.nz), Tradesports (www.tradesports.com), among others. Several commercial and open source software tools are available [6]. The former includes ConsensusPoint (www.consensuspoint.com) and InklingMarkets (www.inklingmarkets.com). The latter includes IdeaFutures and Zocalo provided by Sourceforge.net.

Regarding disease prediction, it is heartbreaking to know that epidemic diseases cause thousands of people to lose their lives each year. According to a recently World Health Organization report [68], cholera is one of the most common epidemic diseases. There are 1.4–4.3 million reported cases annually, causing 28–142 thousand deaths each year. This number is surprisingly high given the

* Corresponding author.
E-mail addresses: eli@nccu.edu.tw (E.Y. Li), ctung@nccu.edu.tw (C.-Y. Tung), tommy1426@gmail.com (S.-H. Chang).

http://dx.doi.org/10.1016/j.ijmedinf.2016.04.014
1386-5056/

well-established and effectively-executed cholera prevention and treatment programs in many regions of the world. The quest for an effective system capable of monitoring and predicting the trends of epidemic diseases is a critical issue for communities worldwide.

Many quantitative models for epidemic prediction have been reported in the literature. One could classify them into three categories: mathematical models, computational models, and surveillance models [73,39,33,7,45,24]. While a mathematical model describes the behavior of a system using mathematical properties and language, a computational model uses computing resources to study the behavior of a system for which analytical solutions are not readily available. The former model includes operations research models [73,30,1,34,43,66,33], trend extrapolation models [37,54,38], compartment models [25,72,59,41], incubation-based models [13,22,65], among others. The latter model encompasses simulation models [17,19,3,40,46], agent-based models [60,9,47], neural network models [16,26,2,64], among others.

As [33,p. 411] puts it, "(a)ny mathematical model of the disease and its control is bound to be incorrect." Instead of using the aforementioned models which offer periodic predictions based on historical data, healthcare researchers of the [67] use surveillance systems, called early warning systems (EWSs), to predict trends of infectious diseases by including such environmental variables as climatic factors and vegetation patterns. However, an EWS faces difficulties in acquiring sufficient suitable epidemiological data and in updating the data continuously. Later in 2006, Boston Children's Hospital initiated another surveillance system, the HealthMap project [21], to bring together multiple heterogeneous data sources for a unified and comprehensive view of the current global state of infectious diseases and their local effects on human and animal health [7]. The project established a Website "Healthmap.org" and a mobile app "Outbreaks New Me" both are freely available to the public. More recently in 2008, Google introduced its "Flu Trends" (GFT) application to monitor the severity of flu epidemics weekly in most developed countries. The application uses aggregated Google search data to estimate current flu activities around the world in near real-time [23]. Instead of Google, Polgreen et al. [51] used a similar tool, the Yahoo! search engine, to produce flu trends that are strikingly similar to the actual influenza occurrence pattern. However, web searches for real-time influenza data can indicate only what has happened, but not what will happen; i.e., they are not predictions. In addition, GFT does not provide surveillance signals for most regions in Asia and Africa and can be used only to monitor the flu epidemic; it is not applicable to other diseases.

Soon after the debut of GFT [38], demonstrated with time-series analysis that GFT and emergency department data were good indicators for the early detection and monitoring of the 2009H1N1 pandemic waves in Manitoba 1–2 weeks in advance. Recently Pervaiz et al. [50], developed an EWS for flu epidemics, called FluBreaks (www.dritte.org/flubreaks), based on GFT data. An extensive review of 32 relevant studies recommends the continuance of using search queries and social media for disease surveillance, early detection, and epidemic effect mitigation [5]. Yom-Tov et al. [71] collected Internet data from Twitter and Bing search engine queries and demonstrated the feasibility of creating a public health surveillance system for mass gatherings based on Internet data. Contrary to using the data from search engines and social media [62], developed a Web-based survey for daily health symptomatic surveillance. During a seasonal influenza epidemic, the survey detected a rapid increase in the number of participants with fever through the early aberration reporting system. Finally, despite the popularity of search engine data, Google ceased to update and publish the GFT data on Thursday, August 20, 2015 (see http://www.google.org/flutrends/). Meanwhile, HealthMap began to combine search, social media, and traditional data sources to improve influenza surveillance [56].

In summary, in order to predict the trends of epidemic diseases dynamically and reliably, we need to overcome the aforesaid three difficulties: having real-time and updated data, incorporating experts' experience and knowledge, and offering continuous and long-term predictions. A plausible way to overcome these difficulties is to use a PMS. The PMS can accumulate the wisdom of crowds, continuously renew the relevant data, and make up-to-date predictions. It is useful for forecasting the spread of infectious diseases [52,53].

2. Research purpose

The purpose of this study is to demonstrate a real-life PMS and the matching mechanism of logarithmic market scoring rules (LMSR) for predicting the trends of epidemic diseases. The effectiveness of the system is briefly discussed. We first constructed a PMS specifically for epidemic disease prediction, following the requirements from Taiwan's Center for Disease Control (CDC). Encouraged by the CDC, thousands of health professionals in 543 hospitals and health institutions all over Taiwan were invited to use the system. Among them, 630 registered and were certified to use the system.

In order to prepare the participants for using the PMS, we delivered two lectures on the PMS hosted by the CDC and produced introduction brochures of the PMS in print and online to invited health professionals. Furthermore, we made ourselves available online or by phone to help health professionals understand, participate, and trade in the PMS. The system was open for 31 weeks, and the results confirm that the accuracy of epidemic disease prediction provided by the PMS is better than the historical-average approach being used by the CDC. Both academicians and practitioners could adapt the PMS of this study and create similar systems for epidemic disease prediction in their own regions, allowing the global community at large to benefit from the results of this study.

The remaining sections of this paper are organized as follows. Section 3 discusses the definition and applications of PMS, the architecture for PMSs, and the matching mechanism of logarithmic market scoring rules [28]. Section 4 describes the interface designs of the epidemic PMS implemented in this study as well as the procedure of experiment. Section 5 analyzes the data collected and interprets the findings from the results. Finally, Section 6 draws several conclusions and recommendations and discusses practical and social implications of EPMS.

3. Prediction market systems

A prediction market system (PMS), operating like a futures market [44], can be used as a mechanism to integrate information from different sources to predict the outcomes of future events [12]. PMSs have two major features: providing appropriate reward mechanisms and performing continuous prediction corrections. Traders in a PMS buy or sell the predictions of a future event based on public or private information. The actual results of the future event determine the rewards (or penalties) for the trader, whereas the price of the future event contract represents the entire market's prediction for the outcome of the future event.

In the last two decades, prediction markets have been empirically proven to be remarkably accurate in forecasting future events with a lower prediction error than conventional forecasting methods [8,35,61,69]. For instance [4], analyzed the performance of different predictions concerning five US presidential elections in the period of 1988–2004. They compared the predictions of the Iowa Electronic Markets with those of 964 polls conducted by two distinguished polling agencies, the Gallup Poll and the Harris Poll,