Developing an Automatic Diagnostic System Driven by Consumer-Generated Health Care Data

H Takata, H Nogawa, H Tanaka

Citation

Abstract
In order to develop an accurate automatic differential diagnostic system with statistical methods, a large amount of data from patient interviews is required. On the other hand, many Internet users often supply their own information to the public. Our hypothesis is that the medical data collected from general Internet users is useful to develop an automatic statistical diagnostic system. The purpose of this study is to assess this hypothesis. We installed a simple health information website on the Internet. On this website, we hosted a Web application that can infer diseases compatible with the information entered by the users. On this site, if its users had visited a medical institution after using this Web application, we requested them to inform us about doctors’ diagnoses. We developed an automatic diagnostic system based on the Naïve-Bayes algorithm with only these data. For evaluation, we used as test data the cases from clinical case questions in a Japanese National Medical Practitioners Qualifying Examination, which medical students must pass to become medical doctors in Japan. We evaluated the correct-answer rate to prove whether this new application makes the minimum satisfactory judgment demanded of doctors. We collected 8812 cases from our website and developed an automatic diagnostic system based on the Naïve-Bayes algorithm with these data. The correct-answer rate of this new application was 69.4% (95% CI, 58.8-80.1%). This rate is higher than the borderline (66%) of the examination. This result suggests that medical data collected from general Internet users is valuable in developing an automatic differential diagnostic system and that an online automatic diagnostic system can be developed that improves its own accuracy with user log data.

INTRODUCTION
We developed an automatic diagnostic system with a new method and evaluated its precision.

The first step in the medical diagnostic process is listing a number of probable diseases during patient interviews and estimating the probability of each disease. This process is referred to as a differential diagnostic process, in which the doctors list those diseases that are compatible with the results of the patient interviews and are common to people of the same age and sex as the patient. This process is important in primary care, where > 70% of outpatients are diagnosed correctly through interviews only [1-3].

In this process, doctors estimate diseases and their probabilities with simple rules, which are based on Bayes’ theorem [4-6]. However, it is difficult for doctors to list all considerable diseases completely and to estimate their probabilities accurately. Hence, we developed a statistical diagnostic estimation system based on Bayes’ theorem and public data [7]. This will be useful for supporting diagnostic processes for doctors and the education of medical students.

Some computer programs that infer disease from patient interviews, or with that users can search diseases have already been developed. Some of these systems work with a rule-based inference method [8, 9]. Most of these rule-based systems accept only some limited symptoms or work only for limited specialties. For instance, Andrew ME, et al. and Maizels M, et al. developed a computer application for headache diagnosis [10, 11], and Kopec D, et al. developed an expert system for aiding migraine diagnosis [12]. Many researchers are trying to develop rule-based diagnostic systems for the general primary care field [8,9]. However, these systems are still today not in widespread use [13]. Moreover, few systems are tested with real patients [14].

We considered it difficult to develop useful diagnostic systems with rule-based approaches for the general field, where treatment of many kinds of sickness is needed. The reason is that development of rule-based diagnostic systems requires clearly stated rules for diagnosis, such as clinical practice guidelines. However, guidelines are not uniformly successful in improving care and several instances of
implementation failure have been described, some resulting in substantial waste of time and resources [15-17].

Therefore, we believe that statistical methods with a lot of data are more appropriate for developing automatic diagnostic systems for the primary care field than rule-based approaches. However, in order to develop a new diagnostic system with a statistical method, large amounts of data from integrated sources are required. Researchers and developers find it sometimes difficult to get large amounts of data from one integrated source. The reason is that many patients visit only small clinics and their data are stored separately.

On the other hand, many researchers use the Internet for their surveys. Ritter P, et al. and Kongsved SM, et al. reported that Internet-based questionnaires were as good as mailed questionnaires for collection of data to evaluate patient education and other intervention [18,19]. Graham AL, et al. reported that race/ethnicity and income do not affect the psychometric properties of most Internet-administered measures examined in their survey [20]. Moreover, many Internet users often supply their own information to the public.

OBJECTIVES

Internet users are biased sample and the medical data entered by them would be not accurate, but we expected that these matters are not so serious to use for development of automatic statistical diagnostic systems. Put another way, we expected that data collected from general Internet users are accurate enough to use for development of automatic statistical diagnostic systems. The purpose of this study was to inspect this hypothesis. Therefore, we collected the data on our website and developed a new automatic diagnostic system with this data without any data cleansing or correcting process even if the data seemed to be misentered by users.

METHODS

DATA FOR ESTIMATION

We collected user data for estimation on a simple health information website on the Internet (http://www.curebot.jp, in Japanese only.) This site was public from February 2007 to January 2008. We did not advertise this web site. On our site, we installed a simple application that can estimate diseases through a simple Web interview form only. The application used for this website was developed by H. Takata, et al. in 2006 [7]. This Web application accepts information on the user’s age, sex, and symptoms, and lists diseases that are compatible with the information entered by the user, along with knowledge about symptoms and the epidemiological data. In addition, it can interactively specify the possible diseases by displaying questions about the symptoms and receiving the answers. However, the inference accuracy of this program is low when a patient is a neonate, pregnant, or suffering from a psychological illness [7]. Hence, we wrote a message warning these users not to use our site. On this site, if a user had visited our site previously, we questioned the user whether he/she had consulted a physician since his/her last access to the site. If so, we asked the users to tell us the names of any diseases their doctors had told them, and to agree to let us use this information for development of better diagnostic systems. Information was collected only from those users who agreed. Therefore, all study participants provided informed consent with our Web forms. Figure 1 is a translated copy of consent form for this agreement.

Figure 1

Figure 1. Consent form

40345 users accessed this website, and 8812 users agreed this study and entered their diagnostic information. We classified all these cases by sex and age groups (0, 1–9, 10–19, 20–29, 30–39, 40–49, 50–59, and over 60). We calculated the prevalence of each disease the users corresponded to in each sex and age group and the proportion of users suffering from each symptom of a disease. Even when a user reported more than one diagnosed disease, we did not correct these data. However, this clearly tended to negatively influence the probabilities of all symptoms related to only one of these diseases.
IMPLEMENTATION

We used the Naïve-Bayes algorithm for the inference method [6,21]. We assumed the prior probability of each disease equals the prevalence of disease in the age and sex group, and the conditional probabilities for all symptoms equal the proportion of users suffering from each symptom of the disease. Moreover, we assumed that multiplying prior probability by conditional probabilities for all symptoms that users entered gives the probability of the disease [6]. In other words, our system assumed that each symptom was an independent event. This assumption is called the Naïve-Bayes assumption. The Naïve-Bayes assumption is not strictly correct in this case. However, it is often used in the medical field and often leads to an accurate conclusion appropriate for practical use [5,6,21]. We implemented our new system as a Web application using Ruby 1.9 [22]. Our new application requires MySQL 5.0 [23], Apache 2.0 [24] and RedHat Linux 8.0 operating system [25].

Our application involves the following data estimated by our Web survey.

- Prevalence of each disease in each age and sex group.
- Proportion of users suffering from each symptom of a disease.

We entered these data into our application with the data structure used in our previous implementations [7].

Figures 2–4 are screenshots of this system (Our application shows only the Japanese version. We have overwritten the Japanese sentences/words in the original screenshot images with English translations for readability).

Figure 2 is the screenshot of the starting page with the form for the user’s age and sex.

**Figure 2**
Figure 2. Screenshot 1.

Figure 3 is the screenshot of the page displayed after that in Figure 2. This page shows the form for primary complaint. Users select the most pressing symptom from 80 symptoms using the two-pane drill down menu in this page. The upper box is the pane for selecting categories. The lower box is the pane for selecting one complaint from the category selected in the upper box.

**Figure 3**
Figure 3. Screenshot 2.

Figure 4 is the screenshot of the next displayed page. This page shows a form asking whether the user has another serious symptom (upper box) and a list of several considerable diseases sorted by estimated probability (lower box). The symptom in question in the upper box is the symptom with the highest likelihood for the disease with the highest probability in the lower box. After answering the question in the upper box, our application specifies the diseases in the lower box, updates their probabilities, and displays a new question in the upper box. This process is repeated until the user finishes answering a predetermined number of questions regarding the symptoms (default number is 10) or finishes answering all questions related to common diseases expected from previous answers.

**Figure 4**
Figure 4. Screenshot 3.
EVALUATION

In order to evaluate this kind of system, the rate of correct diagnosis is one of the most frequently used indicators. However, the rate of correct diagnosis is not a good indicator for evaluation in primary care because in this field most patients have only a common cold, which is not a severe disease. Failure to diagnose rare, severe diseases is more significant than to diagnose a large number of slight diseases correctly. However, there are not any authoritative evaluation methods for automatic diagnostic systems for primary care.

Therefore, in order to evaluate the precision of this new application, we used a qualifying examination for evaluation of minimum ability to function as a medical doctor. We used as test data the cases from clinical case questions in the Japanese National Medical Practitioners Qualifying Examination. This examination is very authoritative in Japan. Passing this examination is necessary to become a medical doctor (and to take this examination, examinees are required to finish 6 years of medical course work).

Examinees of this test are required to answer some clinical case questions. These questions involve a clinical case and ask the examinees to select the correct diagnosis or appropriate treatment from a number of choices. In the 101st National Medical Practitioners Qualifying Examination (carried out in 2007), 220 clinical case questions were included. We evaluated the correct-answer rate to prove whether this system can make the minimum satisfactory judgment of that demanded of physicians.

We selected questions for evaluation from 220 clinical case questions from the 101st examination with the following criteria.

No diagnoses are included in the question, and diagnostic knowledge is required to answer it correctly.

The case in question is not in patient categories excluded from our website, i.e., pregnant women, neonates, or people suffering from psychological illness.

The question does not require the examinee to make a diagnosis from information (symptoms, history, or laboratory tests) not included in our original application.

Seventy-two questions met these criteria. The clinical cases in these questions were entered in our application, and our system listed a number of diseases accordingly. In this examination, the correct answers are available to the public, but the correct diagnosis is not clear when the question is not asking for the correct diagnosis itself. Therefore, we recognized the answer as correct only when the answer given by this system met the following requirements.

When the original question required the examinee to select the correct diagnosis from various choices, our program estimated the probability of correct diagnosis higher than the probabilities of any other disease candidate.

When the original question required the examinee to select an appropriate treatment, we let two practitioners determine the correct diagnosis, and when their diagnoses did not correspond, we requested another practitioner’s opinion about which diagnosis was more appropriate. Then, our program estimated the probability of the disease the practitioners inferred higher than any other disease.

A person who had not received any special medical education entered the case data from the clinical case questions.

RESULTS

We collected 8812 cases from our website, as shown in Figure 5. This graph shows that our samples are biased toward the 20–39-year-old group, and 46.2% of users are male, and 53.8% are female.

Figure 5

Figure 5. Cases collected on our website.

The diseases of these 8812 cases are shown in Table 1, (the cases are categorized with the International Classification of Diseases 10 (ICD10) [26]). However, we wrote a warning for neonates, pregnant users or users suffering from psychological illnesses; a few of these users are included in our samples. We did not delete these cases. Moreover, the data was not corrected for these samples.
The correct-answer rate of this new application was 69.4% (50/72, 95% CI, 58.8–80.1%). This rate is higher than the minimum passing score (66%) of the examination [27].

Correct-answer rates in each field are shown in Figure 6.

DISCUSSION

PRINCIPAL RESULT

In this study, we developed an automatic diagnostic system based on the data collected on an Internet website, and evaluated its performance in listing a number of diseases based only on patient interviews. One of the criteria to pass the 101st National Medical Practitioners Qualifying Examination is a correct-answer rate of clinical case questions greater than 66% [27]. The correct-answer rate of this application satisfied this criterion. Therefore, we consider the performance of this application to be high enough to pass this examination and to function as a medical doctor in terms of the ability to list considerable diseases from a patient interview. However, there is no statistical significance (p = .05) between the correct-answer rate of this system and the minimum passing score of the examination.

LIMITATIONS

In this study, we evaluated the performance of this
Developing an Automatic Diagnostic System Driven by Consumer-Generated Health Care Data

application to list a number of considerable diseases through patient interviews only. It is important for medical doctors to infer diagnosis before calling for special laboratory tests. However, medical doctors are required to have a lot of other skills and knowledge. Therefore, we do not insist that software applications based on data from the Internet are an alternative to medical doctors. Moreover, in this study, cases for evaluation and training were limited. Therefore, further determinant research is required to confirm the findings of this study.

CONCLUSIONS

Medical data collected from general Internet users is shown to be useful for developing an automatic statistical diagnosis system. The data from the Internet is sometimes inaccurate, and in some cases confirming its accuracy is difficult. Moreover, Internet users are often a biased sample (Figure 5). Even though cases from the Internet have these problems, the results of our study showed that sampling from the Internet is useful for developing an automatic statistical diagnostic system.

The correct-answer rate of our new application is especially high in some fields (Figure 6), such as pulmonology and gastroenterology. In these fields, our website succeeded in acquiring more cases than in other fields (In Table 1, pulmonology corresponds to the 10th category of ICD10, and gastroenterology to the 11th category). This suggests that collection of more data from the Internet can lead to a more accurate system, and that an online automatic diagnostic system can be developed that improves its own accuracy with user log data.

We forecast that increasing user-generated contents on the Internet will make it easier to develop artificial medical intelligence applications, and will give them better performance. Currently, there are many kinds of Web services working with user-generated content. Growth of these services is one of the trends called Web 2.0. This trend in the medical field will greatly increase user-generated medical data on Internet. We predict that these new Web services will be useful not only for customers, but also for future development of artificial medical intelligence.

References

Author Information

H. Takata, MD
Department of Bioinformatics, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 113-8510, Japan

H. Nogawa, MD, PhD
Japan Medical Information Network Association, Toho Hukasawa Building 5F, 2-2-1 Yushima, Bunkyo-ku, Tokyo, 113-8510, Japan

H. Tanaka, PhD
Center of Information in Medicine, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 113-8510, Japan