

Blog Snippets Based Drug Effects Extraction System Using Lexical and Grammatical Restrictions

Shiho Kitajima, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan

Rafal Rzepka, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan

Kenji Araki, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan

ABSTRACT

Obtaining medical information has a beneficial influence on patients' treatment and QOL (quality of life). The authors aim to make a system that helps patients to collect narrative information. Extracting information from data written by patients will allow the acquisition of information which is easy to understand and provides encouragement. Additionally, by using large-scale data, the system can be utilized for discovering unknown effects or patterns. As the first step, the purpose of this paper is to extract descriptions of the effects caused by taking drugs as a triplet of expressions from illness survival blogs' snippets. This paper proposes a method to extract the triplets using specific clue words and parsing the results in order to extract from blogs written in free natural language. Moreover, recall was improved by combining the their proposed method and a baseline system, and precision was improved by filtering using dictionaries we created from existing medical documents.

Keywords: Blog Snippets, Drug Effects, Information Extraction, Japanese, Medical Information, Opinion Mining, Text Mining, Triplet Extraction

1. INTRODUCTION

Currently, the total number of patients with serious illnesses in Japan has reached about 18,026,000 (Ministry of Health, Labour and Welfare, 2011). When someone is diagnosed with a disease or feels that they might be ill, medical information is required to solve various

questions in such situations. If patients obtain the required information and they are satisfied with it, this can motivate patients in decision-making and proactive treatment, and have beneficial effects for their QOL (quality of life) and health (Setoyama & Nakayama, 2011). Most major sources of medical information for patients are medical personnel, who are

DOI: 10.4018/ijmdem.2014040101

recognized as a reliable source (Taniguchi, 2004; Hesse et al., 2005). However, in order to make better decisions, both “evidence information”, which shows the effects for the population stochastically, and “narrative information” which consists of individual stories from personal experience are necessary (O’Connor, 2002). Although obtaining stories of the experiences of other patients with the same illness can reduce a patient’s anxiety and encourage them (Matsumoto et al., 2005; Maeda et al., 2009), it is difficult for medical staff to provide patients with enough narrative information. Therefore, patients are using various sources such as the Internet in addition to doctors, as has been clarified by research (Setoyama & Nakayama, 2011; Walsh et al., 2010). However, it is not easy to extract and collect answers required by patients appropriately, since there is a huge amount of information on the Web, and it is written in free natural language. In addition, we can see from survey results that it is hard for patients to ask doctors or nurses certain questions (Tsuchiya & Horn, 2009; Kokubu 2008), and we can also see that patients feel that there is a lack of information. Accordingly, we aim to make a system that helps patients to obtain information about other patients with the same illness.

Currently, pharmaceuticals used in Japan consist of about 18,000 prescription drugs, which are prescribed based on a diagnosis of a doctor, and about 12,000 OTC (Over-the-Counter) drugs, which people can buy at pharmacies and drugstores without a prescription by a doctor. Although these drugs are necessary for medical treatment, side effects that occur with use are a problem. Early detection of side effects and taking countermeasures are important issues in the medical field (Mikami et al., 2013). Thus, a great deal of research to estimate the relationship between drugs and side effects has been carried out, and there are many medicine search services (NPO Narrative of Disease and Health DIPEX-Japan, n.d.; Pharmaceuticals and Medical Devices Agency, n.d.). These are based on data created by manufacturers and distributors of pharmaceuticals and information reported by doctors

or pharmacists. As yet, medical institutions cannot deal with unknown side effects because additional risks of using medication, such as those not reported, often become known after a product is used by a larger number of patients for a longer period of time. Therefore, we need to collect as much information as possible in order to ascertain effects and side effects of medicines clearly. Thus, we focused on large-scale data on the Web which we can obtain in real time. We extract medical information from blogs written by patients and their families in order to acquire narrative information. Although patients’ understanding of diseases and medical care and the reliability of Internet sources are still far from satisfactory, information on the Web is extensive and instantaneous, and thus can be used to understand people’s changes and trends. Additionally, Japanese people usually disclose personal information anonymously on social media and blogs. It is considered that anonymous narratives are unreliable due to lacking a sense of responsibility. However, anonymized information on the Web gives us information that may be inaccessible in the real world, since anonymity allows users to express opinions and ideas that they may not want someone close to know, without the need to worry about appearances (Orita et al., 2007).

We aim to develop a system that searches for medical information from patients’ blogs. For example, in our proposed system, when the user inputs a name of a disease, the system collects blogs written by patients of that disease, extracts information related to treatment from them, and outputs statistical data such as a distribution of treatments. Moreover, when the user inputs a drug name, we aim to output data on how many patients felt what kind of effects, side effects, and unknown effects. We believe that a system that searches for drug effects will be very useful for patients’ mental support and decision-making regarding their treatment, and would be useful for investigating unknown effects.

As a first step, we intend to extract descriptions of both positive and negative effects caused by taking drugs. The core of an opinion

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/blog-snippets-based-drug-effects-extraction-system-using-lexical-and-grammatical-restrictions/113304?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Communications and Social Science, InfoSci-Communications, Online Engagement, and Media eJournal Collection, InfoSci-Knowledge Discovery, Information Management, and Storage eJournal Collection, InfoSci-Networking, Mobile Applications, and Web Technologies eJournal Collection, InfoSci-Surveillance, Security, and Defense eJournal Collection, InfoSci-Journal Disciplines Engineering, Natural, and Physical Science. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Policy Decision Support Through Social Simulation

Luis Antunes, Ana Respicio, João Balsa and Helder Coelho (2011). *Gaming and Simulations: Concepts, Methodologies, Tools and Applications* (pp. 1530-1538).

www.igi-global.com/chapter/policy-decision-support-through-social/49465?camid=4v1a

Communication and Relation Building in Social Systems

Thomas Hansson (2008). *Handbook of Research on Digital Information Technologies: Innovations, Methods, and Ethical Issues* (pp. 372-391).

www.igi-global.com/chapter/communication-relation-building-social-systems/19854?camid=4v1a

HaMA: A Handicap-based Architecture for Multimedia Document Adaptation

Asma Saighi, Roose Philippe, Nassira Ghoualmi, Sébastien Laborie and Zakaria Laboudi (2017). *International Journal of Multimedia Data Engineering and Management* (pp. 55-96).

www.igi-global.com/article/hama/182651?camid=4v1a

Spatio-Temporal Analysis for Human Action Detection and Recognition in Uncontrolled Environments

Dianting Liu, Yilin Yan, Mei-Ling Shyu, Guiru Zhao and Min Chen (2015). *International Journal of Multimedia Data Engineering and Management* (pp. 1-18).

www.igi-global.com/article/spatio-temporal-analysis-for-human-action-detection-and-recognition-in-uncontrolled-environments/124242?camid=4v1a