An Effective Analysis Method of Discussions in Bulletin Board Sites

Shigeaki Sakurai
Toshiba Corporation, Japan

INTRODUCTION

As both Internet environments and computer environments dramatically grow, it is easier and easier to discuss many topics with many people. Large amount of discussions are daily performed and stored in bulletin board sites. Most of the discussions are unimportant for most of people. But small proportion of the discussions may be able to have important influence on organizations and the companies referred to these discussions. If the organizations and the companies do not early recognize the discussions and do not timely perform appropriate countermeasures, their brand images may be injured or they may overlook the symptom of the boom leading to the sales. They are asked to early discover the discussions. However, since there are many bulletin board sites and many discussions are performed in each site, it is impracticable for the organizations and the companies to manually check all discussions in detail. On the other hand, they have various interests. Even if some discussions are important for an organization, the discussions are not always important for the other organization. The organizations and the companies need to discover discussions corresponding to their interests. They need a method that can easily analyze the discussions and can help them to discover true important discussions. This article introduces one of methods. The method extracts events related to their interests from discussions. The discussions are ranked by referring to the event. The discussions with high order are extracted as important discussions. The method also extracts keywords representing the discussions. The extracted discussions and their keywords help the organizations and the companies to quickly understand the topics of the discussions. The organizations and the companies can discover true important discussions without heavy burden. In the following sections, the method is introduced.

BACKGROUND

Large amount of information is uploaded to the Web and is regarded as a knowledge database. We believe that the usage of the knowledge database can help our decision making in various situations and improve our daily living. However, the knowledge database is too huge and it is not always well-defined. Most of the knowledge database is described by natural language. We cannot sufficiently use it without appropriate text mining techniques. Many studies based on this theme have been still proposed.

For example, Hu and Liuite (2004) proposed a method that analyzes customer reviews on the Web. The method extracts product features and identifies which sentences include opinions for the products. In addition, it summarizes the identified opinions. Morinaga, Yamanishi, Tateishi, and Fukushima (2002) proposed a method that analyzes reputations for products. The method extracts opinions by using syntactic and linguistic rules described by experts. It identifies whether the reviews include positive opinions or negative opinions. Either positive labels or negative labels are assigned to the reviews. The method also attaches the product name and the degree of the confidence to the opinions. Kobayashi, Iida, Inui, and Matsumoto (2005) proposed a method that extracts attribute-value pairs and judges whether the pairs are an opinion of the author. The method uses machine learning techniques. Here, the pairs are composed of target objects and adjective expressions related to them. They are extracted from textual data on the Web sites. On the other hand, Esuli and Sebastiani (2006) proposed SENTIWORDNET for text mining methods which especially focus on opinions in the textual data. It is a lexical resource associated with WordNet (Miller, Fellbaum, Tengi, Langone, Ernst, & Jose, 2012) where WordNet is a huge lexical database of English. It assigns three scores to the terms in WordNet synset. The scores represent how

objective, positive, and negative the terms are. Even if these methods do not always focus on the analysis of bulletin board sites, we can apply them to the analysis to some extent. This is because the sites include various textual data. However, they focus on individual sentences or texts and do not sufficiently consider text sets sequentially described by many authors. In the following, this article focuses on the textual data collected from bulletin board sites.

ANALYSIS OF BULLETIN BOARD SITES

Data Included in Bulletin Board Sites

This article deals a textual data as shown in Figure 1. The data is called the thread. A thread corresponds to a discussion. The thread is a sequential article set included in a bulletin board site. Each article describes the contents related to a specific topic.

Outline of an Analysis Method

This section introduces an outline of an analysis method of threads (Sakurai & Orihara, 2006a, 2006b). The method is composed of three extraction processes: the event extraction, the important thread extraction, and the relevant incident extraction. Figure 2 shows an outline of the analysis method. The method is input threads stored in the article database. It outputs analysis results to the analysis result database by referring to the analysis knowledge database and the classification model.

Event Extraction

The event extraction process picks up events by analyzing texts in articles. The events are used in the evaluation of important threads. If the process overlooks necessary events, the important thread extraction process fails to pick up the important threads. It is necessary for the event extraction process to avoid overlooking

Figure 1. An example of a thread in a bulletin board site
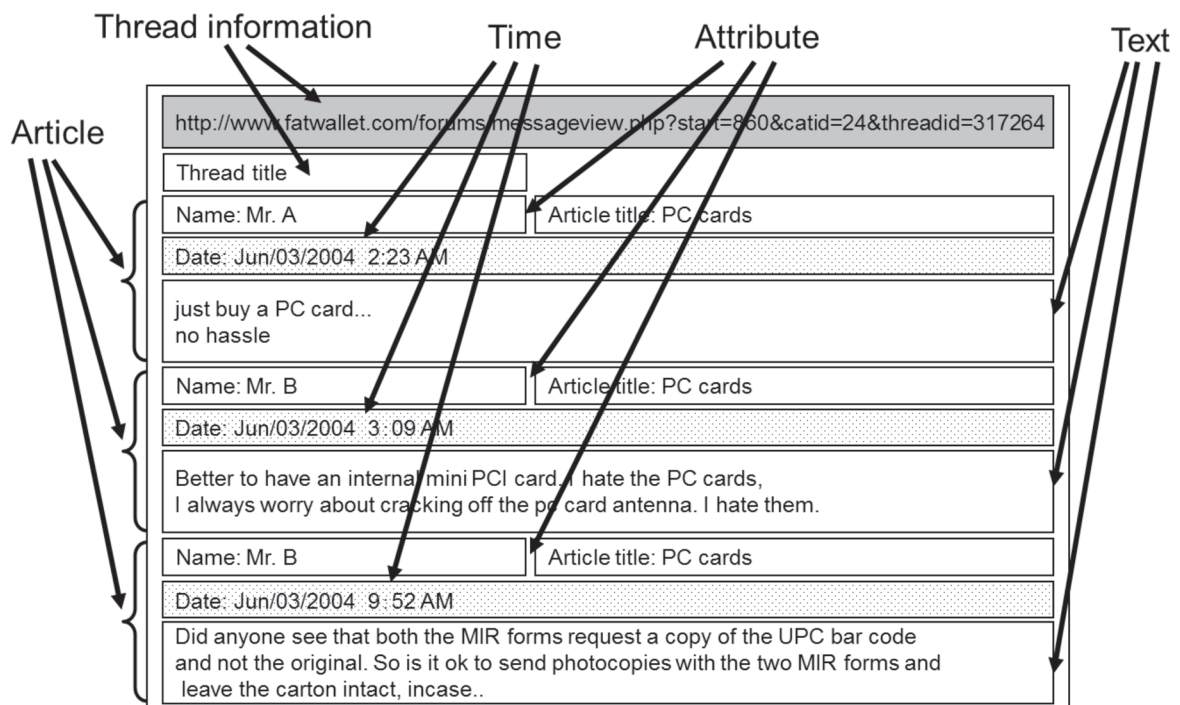
8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/an-effective-analysis-method-of-discussions-in-bulletin-board-sites/112610

Related Content

Workflow Modeling Technologies
Maria N. Koukovini, Eugenia I. Papagiannakopoulou, Georgios V. Lioudakis, Nikolaos L. Dellas, Dimitra I. Kaklamaniand Iakovos S. Venieris (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 5348-5356).*
www.irma-international.org/chapter/workflow-modeling-technologies/112983

Teaching in Visual Programming Environments
Wilfred W. F. Lauand Allan H. K. Yuen (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 2600-2608).*
www.irma-international.org/chapter/teaching-in-visual-programming-environments/112676

IS Design Considerations for an Innovative Service BPO: Insights from a Banking Case Study
Myriam Raymondand Frantz Rowe (2016). *International Journal of Information Technologies and Systems Approach (pp. 39-56).*
www.irma-international.org/article/is-design-considerations-for-an-innovative-service-bpo/152884

Ubiquitous Professional Training for Teachers using the uProf! Model
Sabrina Leoneand Giovanni Biancofiore (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 7410-7418).*
www.irma-international.org/chapter/ubiquitous-professional-training-for-teachers-using-the-uprof-model/112439

Method of Fault Self-Healing in Distribution Network and Deep Learning Under Cloud Edge Architecture
Zhenxing Lin, Liangjun Huang, Boyang Yu, Chenhao Qi, Linbo Pan, Yu Wang, Chengyu Geand Rongrong Shan (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/method-of-fault-self-healing-in-distribution-network-and-deep-learning-under-cloud-edge-architecture/321753