Knowl Inf Syst (2008) 14:143–160
DOI 10.1007/s10115-007-0078-2

REGULAR PAPER

Learning to extract and summarize hot item features from multiple auction web sites

Tak-Lam Wong · Wai Lam

Received: 12 April 2006 / Revised: 14 November 2006 / Accepted: 26 January 2007 /
Published online: 31 March 2007

Abstract It is difficult to digest the poorly organized and vast amount of information contained in auction Web sites which are fast changing and highly dynamic. We develop a *unified* framework which can automatically extract product features and summarize hot item features from multiple auction sites. To deal with the irregularity in the layout format of Web pages and harness the uncertainty involved, we formulate the tasks of product feature extraction and hot item feature summarization as a single graph labeling problem using conditional random fields. One characteristic of this graphical model is that it can model the inter-dependence between neighbouring tokens in a Web page, tokens in different Web pages, as well as various information such as hot item features across different auction sites. We have conducted extensive experiments on several real-world auction Web sites to demonstrate the effectiveness of our framework.

Keywords Information extraction · Web mining · Conditional random fields

The work described in this paper is substantially supported by grants from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project Nos: CUHK 4179/03E and CUHK4193/04E) and the Direct Grant of the Faculty of Engineering, CUHK (Project Codes: 2050363 and 2050391). This work is also affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing and Interface Technologies.

T.-L. Wong (✉)
Department of Computer Science, City University of Hong Kong,
83 Tat Chee Avenue, Kowloon, Hong Kong
e-mail: wongtl@cityu.edu.hk

W. Lam
Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Shatin, Hong Kong

1 Introduction

The fast development of the Internet technology can effectively vanish the geographical barrier of different communities. Especially, the easily accessible World Wide Web creates a profit-generating market place and convenient shopping environment for sellers and customers respectively. One example is the online auction business such as the Web site *ebay.com*. In auction sites, sellers can place their items such as a brand new digital camera, or a second hand MP3 player for bidding. Potential buyers can then browse auction sites and bid their favorite items by asking a price that they are willing to pay. Once the bidding period ends, the potential buyer who has asked for the highest bidding price can eventually purchase the item. In the past decade, online auction Web sites have been becoming increasingly popular. According to the press release from *ebay.com*, they currently have 147 million community members and approximately 50 million items for sale at any given time.[1] There are several reasons accounting for the popularity of the online auction business. One reason is that sellers can reduce their cost by making use of the online auction environment. Since a huge number of visitors browse online auction sites per day, sellers do not need to set up their own Web sites and promote their products. On the other hand, potential buyers can ask the price of items depending on their budgets and have a chance to successfully purchase an item at a lower price if they can bid the item with the right asking price and at the right time.

In every minute, a huge number of sellers and potential buyers are attracted to participate in online auction Web sites. These auction sites contain a tremendous amount of items from different categories listed for bidding with continuously changing price. Moreover, there exists mutual influence among different participants and items in auction sites. For example, the number of bids for a particular item can be seriously affected if another similar product is listed for bidding with a lower bidding price. As a result, online auction sites become fast changing, highly dynamic, and complex systems. For instance, a digital camera may receive a large number of bids ranging from few US dollars to few hundred US dollars in just 1 or 2 days. Therefore, acquiring the up-to-date and accurate information in auction Web sites offers many benefits to both sellers and potential buyers.

Though online auction Web sites bring much benefit and convenience to sellers and potential buyers, the massive amount of continuously changing information contained in auction sites poses a serious difficulty for sellers and potential buyers to digest and analyze. For example, when a seller intends to place an item for bidding, he/she is required to set a start bidding price. Some sellers may set the start bidding price with their subjective expectations. This can easily result in either that the start bidding price is set too high and hence the chance of the item being sold may be very slim, or that the start bidding price is set too low and hence the return may decrease. Some other sellers may manually analyze the items currently listed for bidding and their price before setting the start bidding price. However, this manual process for analyzing the vast amount of information is time-consuming and tedious. Besides sellers, it is also beneficial for potential buyers to obtain up-to-date, detailed, and accurate information to assist the decision. For example, before bidding for a particular item, the potential buyer may study the description of the item, and other similar items

[1] The article was posted on May 25th, 2005 and was accessible at *http://biz.yahoo.com/bw/050525/255399.html?.v=1*.