Chapter 14
Burst Moment Estimation for Information Propagation

Tomas Kuzar and Pavol Navrat

Abstract. In the article we concentrate on timing aspect of information propagation on social web. The aim of the information producer is to transfer the information to the broad audience via social web. Producer needs to identify interesting content and publish it to social web in the right time. Right timing of information publishing can increase the potential of spreading it. In the article we describe the process of interesting content identification and we present a model for right moment estimation of information publishing. Our estimation is based on producer’s web usage mining, web topic tracking and event identification.

Keywords: Social Web, Web Content Mining, Web Usage Mining.

14.1 Introduction

Rash development of the web technologies, especially of Web 2.0 and social web applications brought new means of user-web interaction: Wikis, Blogs, or Social Networks Sites. New tools for the user collaboration and information sharing flooded the web space by huge amount of unstructured or semi structured information. That is why we need to cope with the question of effective organizing and making web content easily accessible.

The amount of the published content increased enormously and the web user became flooded by information. That is why it is necessary to develop methods for effective publishing and organizing of the web content. Effective marketing should reflect the relationship between the company and the customer. We suppose that adding social media into marketing process will bring the added value to the company and to the brand. Web marketing costs money and energy that is why it needs to be done effectively.

Tomas Kuzar · Pavol Navrat
Faculty of Informatics and Information Technologies, Slovak University of Technology
e-mail: {kuzar, navrat}@fiit.stuba.sk
Company plays the role of producer of the information. The portals and sites of producers usually have no significant user traffic. The idea is to propagate interesting web content which has small traffic. Such situation occurs in social web space frequently – interesting and important content stays unexplored. Different publications discuss the problem of information propagation, viral propagation and propagation based on influencers. We consider the aspect of timing for very important in web information propagation. Our timing model is based on web usage of the producer portal and monitoring of the situation in social web space. The mediators can to be used in order of information diffusion. Conditions for information diffusion are changing over time. We designed a model for optimal Burst Moment Estimation (BME).

Technology brought many possibilities for users to generate the content. Tools like crawlers and RSS can be easily used for automatic web information gathering. We use RRS to collect the text streams. Principally, we are able to process various streams: new paper articles, forum contributions, blogs or statuses of social network users. We started our explorations on online newspaper stream.

### 14.2 Related Work

Our research was motivated mainly by the problem of social web marketing and its viral effects. Authors in \[4, 5, 6, 7\] analyzed viral aspect of information propagation from multiple points of view: pathways identification, influencers, latency, prediction etc.

The studies of viral effects are many times closely related with the topic identification and topic mining. Problem of topic detection and tracking (TDT) is not a new phenomenon. With the social media technologies many models need to be reevaluated and rescaled. Related work on TDT focuses mainly on historical tracking. We want to concentrate on the problem of prediction. Therefore TDT needs to be extended to be able to process data from various sources and also to cope with rapid changes in published content. Other important part of our research is based on the problem of topic mining presented in \[8\].

Internet and web space is flooded by documents and information. The article discusses the problem of automatic uncovering how ideas spread through the collection over time. Most of the publications presented explorations when documents were connected via hyperlinks. Authors in \[1\] proved that content based connection can be used instead of hyperlink connections.

Web space can be understood as huge source of text streams. Authors \[2\] analyzed the problem of text stream synchronization. We need to discuss problem of topic extraction and topic mining.

Model in \[3\] consists of two parts: topic identification and topic intensity identification. The first part addresses the problem of topic identification, which was already discusses in many previous publications. Second part discusses the importance of the topic – topic intensity.