Analyzing the Evolution of Web Usage Data

Alzennyr Da Silva

Projet AxIS, INRIA
Domaine de Voluceau, Rocquencourt, B.P. 105
78153 Le Chesnay cedex, France
Alzennyr.Da_Silva@inria.fr

Abstract. Analyzing Web usage has become a very important strategy for Web site operators as it provides them a better understanding of the users' behavior. This insight can enable the operators to improve their service and thereby attract more visitors. Taking into account the temporal dimension in such analyses has become a necessity since the way a site is visited can indeed evolve due to modifications in the structure and content of the site, or even due to changes in the behavior of certain user groups. Consequently, the models associated with these behaviors must be continuously updated in order to reflect the actual behavior of the users. One solution to this problem, proposed in this article, is to update these models using summaries obtained by means of an evolutionary approach based on clustering methods. To do so, we carry out various clustering strategies that are applied on time sub-periods. We compare the results obtained using this method with the results obtained by a traditional global analysis.

Keywords: Clustering, Evolving Data, Web Usage Mining.

1 Introduction

The Web is one of the most relevant examples of an evolving and dynamic data source. This is due to the fact that new information is constantly being added to existing pages while a huge number of new documents are appearing on-line each day. The access patterns to these pages therefore are of a dynamic nature, due both to the on-going changes in the content and structure of the Web site as well as to changes in the users' interest.

The access patterns can be influenced by certain parameters of a temporal nature such as: the time of the day, the day of the week, recurrent factors (summer/winter vacations, national holidays, the Christmas period) and non-recurrent global events (epidemics, wars, economics crises, the World Cup), etc.

A usage based analysis consists in studying the traces left by users when they visit a Web site. More precisely, Web Usage Mining (WUM) consists in extracting interesting information from Web server's log files. Most methods in this domain take into account the entire period during which usage traces were recorded, the results obtained naturally being those which prevail over the total period. Consequently, certain types of behaviors, which take place during short sub-periods are not detected and thus remain unknown by traditional methods. It is however important to study these behaviors and thus to carry out an analysis relating to significant time sub-periods. This will make it possible to study, for example, possible shifts in the user's interests concerning the Web site's services over time sub-periods. It will then be possible to study the temporal evolution of users' profiles by providing descriptions that are able to integrate the temporal aspect. Furthermore, as the volume of mined data is very great, it is important to define summaries to represent user profiles.

These considerations have give rise to many studies in data analysis, especially concerning the adaptation of traditional static data based methods to the dynamic data framework. The work presented in this article continues in this line of research and proposes to follow changes in users' behavior. We use summaries obtained by an evolutionary clustering approach applied over time sub-periods to carry out a follow-up of the evolution.