

# Abstract

## Incremental Mining Techniques

The increasing necessity of organizational data exploration and analysis, seeking new knowledge that may be implicit in their operational systems, has made the study of data mining techniques gain a huge impulse. This impulse can be clearly noticed in the e-commerce domain, where the analysis of client's past behaviours is extremely valuable and may, eventually, bring up important working instruments for determining his future behaviour. Therefore, it is possible to predict what a Web site visitor might be looking for, and thus restructuring the Web site to meet his needs. Thereby, the visitor keeps longer navigating in the Web site, what increases his probability of getting attracted by some product, leading to its purchase. To achieve this goal, Web site adaptation has to be fast enough to change while the visitor navigates, and has also to ensure that this adaptation is made according to the most recent visitors' navigation behaviour patterns, which requires a mining algorithm with a sufficiently good response time for frequently update the patterns.

Typical databases are continuously changing over the time, what can invalidate some patterns or introduce new ones. Thus, conventional data mining techniques were proved to be inefficient, as they needed to re-execute to update the mining results with the ones derived from the last database changes. Incremental mining techniques emerged to avoid algorithm re-execution and to update mining results when incremental data are added or old data are removed, ensuring a better performance in the data mining processes. In this work, we analyze some existing incremental mining strategies and models, giving a particular emphasis in their application on Web sites, in

order to develop models to discover Web user behaviour patterns and automatically generate some recommendations to restructure sites in useful time.

For accomplishing this task, we designed and implemented Spottrigger, a system responsible for the whole data life cycle in a Web site restructuring work. This life cycle includes tasks specially oriented to extract the raw data stored in Web servers, pass these data by intermediate phases of cleansing and preparation, perform an incremental data mining technique to extract users' navigation patterns and finally suggesting new locations of spots on the Web site according to the patterns found and the profile of the visitor.

We applied Spottrigger in our case study, which was based on data gathered from a real online newspaper. Our main goal was to collect, in a useful time, information about users that at a given moment are consulting the site and thus restructuring the Web site in a short term, delivering the scheduled advertisements, activated according to the user's profile. Basically, our idea is to have advertisements classified in levels and restructure the Web site to have the higher level advertisements in pages the visitor will most probably access. In order to do that, we construct a page ranking for the visitor, based on results obtained through the incremental mining technique. Since visitors' navigation behaviour may change during time, the incremental mining algorithm will be responsible for catching this behaviour changes and fast update the patterns. Using Spottrigger as a decision support system for advertisement, a newspaper company may significantly improve the merchandising of its publicity spots guaranteeing that a given advertisement will reach to a higher number of visitors, even if they change their behaviour when visiting pages that were usually not visited.

Keywords: Clickstreams; Algorithms and Strategies for Incremental Data Mining; Web Site Restructuring.