

Extraction de comportements inattendus dans le cadre du Web Usage Mining

Dong (Haoyuan) Li*, Anne Laurent**, Pascal Poncelet**

*LGI2P, École des Mines d'Alès, Parc scientifique Georges Besse, 30035 Nîmes cedex 1
Haoyuan.Li@ema.fr

**LIRMM, Université Montpellier 2, 161 rue Ada, 34392 Montpellier cedex 5
laurent@lirmm.fr, poncelet@lirmm.fr

Résumé. Au cours de ces dernières années, la fouille de données d'usage du Web s'est de plus en plus concentrée sur l'extraction des comportements des utilisateurs à partir de fichiers logs. Bien que l'extraction de motifs séquentiels permette de trouver des comportements fréquents, les décideurs sont de plus en plus intéressés par des comportements inattendus qui contredisent les croyances sur des connaissances existantes. Dans cet article, nous présentons une nouvelle approche, WebUser, pour découvrir des comportements inattendus par rapport aux croyances du domaine. Les expérimentations menées, avec des bases de croyances générées à partir des comportements connus, montrent que notre approche permet d'extraire des comportements inattendus qui peuvent être utilisés pour, par exemple, améliorer la structure des sites Web ou repérer des usages particuliers.

1 Introduction

L'utilisation de techniques de fouille sur des données d'usage du Web pour extraire des connaissances sur les comportements des utilisateurs, i.e. le Web Usage Mining (WUM), est un domaine de recherche particulièrement étudié ces dernières années. Ainsi, de nombreuses approches ont été proposées pour analyser les logs et offrir de nouvelles connaissances utiles comme, quelles sont les ressources souvent visitées ? quel est le parcours des utilisateurs sur un site ? Ces connaissances peuvent être utilisées pour, par exemple, restructurer un site, personnaliser un site en fonction du comportement des utilisateurs ou encore précharger des pages comme les approches introduites par Büchner et Mulvena (1998), Spiliopoulou et al. (1999), Eirinaki et Vazirgiannis (2003), Srivastava et al. (2000) et Mobasher (2007).

Parmi les techniques de fouille de données utilisées, l'extraction de *motifs séquentiels* présentée par Agrawal et Srikant (1995) est souvent utilisée car particulièrement bien adaptée à ce contexte selon les approches proposées par Mobasher et al. (2002), Masegla et al. (2003), Huang et al. (2006), Missaoui et al. (2007), et Masegla et al. (2007). Par exemple, via les motifs séquentiels, il est possible d'extraire des comportements du type : “sur un site Web de forum, 40% des utilisateurs visitent la page `TopicList`, puis la page `Search`, puis la page `Login`, et enfin la page `PostTopic`”, ou bien dans un site d'e-commerce “80% de clients s'intéressent aux pages sur les sacs de ordinateur portable après avoir ajouté un PC portable