

Diverses approches permettant l'introduction du temps dans la fouille de données d'usage du Web

Alzenny Da Silva

Projet AxIS, INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, B.P. 105
78153 Le Chesnay cedex, France
Alzenny.Da_Silva@inria.fr
<http://www-rocq.inria.fr/axis>

Résumé. Le volume des transactions commerciales réalisées en ligne sur l'Internet ne cesse de croître. Le Web est donc devenu l'une des plates-formes commerciales les plus importantes et beaucoup d'opérateurs de sites Web sont incités à analyser l'utilisation de leurs sites afin d'améliorer leur rentabilité. Cependant, la manière dont un site Web est visité peut changer en raison de modifications liées à la structure et au contenu du site lui-même, ou bien en raison de l'évolution du comportement de certains groupes d'utilisateurs. Ainsi, les modèles d'usage doivent être mis à jour continuellement afin de refléter le comportement réel des visiteurs. Dans ce contexte, nous présentons une brève vue de l'ensemble des techniques appliquées à la fouille des données temporelles. Une attention particulière est apportée aux méthodes consacrées à la découverte de modèles sur les données d'usage du Web. Nous évoquons également certaines questions en suspens dans ce contexte.

1 Introduction

De nombreuses sociétés souhaitent exploiter les grandes quantités de données accumulées quotidiennement durant leurs opérations commerciales pour mieux comprendre leur activité globale et mieux appréhender le contexte commercial, afin d'adopter de meilleures décisions stratégiques. Il en résulte une utilisation croissante de la fouille de données (Data Mining), qui est définie comme l'application de l'analyse de données et des algorithmes de découverte sur les grandes bases de données ayant comme but la découverte de modèles non triviaux (Fayyad et al., 1996). Plusieurs algorithmes ont été proposés afin de formaliser les nouveaux modèles découverts, de construire des modèles plus efficaces, de traiter de nouveaux types de données et de mesurer les différences entre les ensembles de données. Cependant, la plupart des algorithmes de fouille de données supposent que les modèles sont statiques et ne tiennent pas compte de l'éventuelle évolution de ces modèles au cours du temps.

Concernant les données volumineuses et dynamiques, le Web est devenu l'exemple le plus pertinent grâce à l'augmentation colossale du nombre de documents mis en ligne et des nouvelles informations ajoutées chaque jour. Les modèles d'accès au Web ont une nature dynamique, dû au contenu et à la structure d'un site Web ou bien dû au changement d'intérêt de