Approximation du score CFOF de détection d'anomalie dans un arbre d'indexation iSAX : Application au contexte SI de la SNCF

Lucas Foulon[*,**], Christophe Rigotti[***]
Serge Fenet[****], Denis Jouvin[**]

[*]Univ Lyon, CNRS, LIRIS, UMR5205, F-69621, Villeurbanne, France
lucas.foulon@sncf.fr
[**]SNCF Mobilité, DSI Production Ferroviaire, F-69393, Lyon, France
denis.jouvin@sncf.fr
[***]Univ Lyon, INSA-Lyon, CNRS, INRIA,
LIRIS, UMR5205, F-69621, Villeurbanne, France
christophe.rigotti@insa-lyon.fr
[****]Univ Lyon, Université Claude Bernard Lyon 1, CNRS,
LIRIS, UMR5205, F-69621, Villeurbanne, France
serge.fenet@liris.cnrs.fr

Résumé. La finalité de notre travail est la détection des anomalies dans les traces de fonctionnement de l'infrastructure de communication du Système d'Information (SI) de la SNCF. Deux techniques récentes et indépendantes semblent particulièrement appropriées dans notre cas. Il s'agit d'une part du stockage et de l'indexation de séries temporelles dans un arbre appelé arbre iSAX, et d'autre part d'un score de détection d'anomalie nommé CFOF dont la robustesse au phénomène de concentration en haute dimension a été établie de façon formelle. Dans cet article nous montrons qu'il est possible d'utiliser la structuration des informations dans l'arbre iSAX pour déterminer rapidement une approximation du score CFOF. La valeur obtenue est proche du score exact sur des données synthétiques et réelles. Les premiers retours d'expertises indiquent que la méthode semble pertinente pour le déclenchement d'alarmes sur les données issues de trace d'activité du SI de la SNCF.

1 Introduction

Une anomalie peut être définie comme une déviation par rapport à ce qui est défini comme normal. La détection d'anomalie constitue une tâche importante dans de nombreux domaines tels que l'analyse de données, la reconnaissance d'image médicale, la détection d'intrusion dans les systèmes informatiques, ou encore la fraude à la carte bancaire. Avec le développement croissant des volumes de données issues des applications métier, la détection d'anomalie représente par ailleurs un enjeu de plus en plus important dans de nombreux domaines industriels.