

Une J-mesure orientée pour élaguer des modèles de chroniques

Nabil Benayadi* and Marc Le Goc*

*LSIS, UMR CNRS 6168, Université Paul Cézanne
Domaine Universitaire St Jérôme
{nabil.benayadi,marc.legoc}@lsis.org

1 Introduction

Les systèmes de supervision de la plupart des applications industrielles génèrent une très grande quantité d'informations et les collectent dans des bases de données. Ce papier concerne la découverte de modèles de chroniques à partir de séquences d'événements. Chaque événement appartient à une certaine classe. Selon l'approche stochastique (Le Goc et al. (2005)), un ensemble de séquences est représenté sous la forme d'une chaîne de Markov afin de l'utiliser par la suite pour générer un modèle de chroniques (Le Goc et al. (2005)) sous forme de relations binaires entre classes d'événements $C^i \mapsto C^o$. Le nombre des relations binaires peut être très grand, par conséquent une réduction de ce nombre est nécessaire. Pour cela, nous proposons une adaptation de la J-Mesure de la théorie de l'information aux chaînes de Markov, la BJ-Mesure, pour formuler des heuristiques d'élimination d'hypothèses.

2 Élagage d'un modèle de chroniques

Considérant la propriété d'absence de mémoire de la chaîne de Markov, la relation $C^i \mapsto C^o$ entre deux classes C^i et C^o peut être considérée comme l'une des quatre relations entre deux variables aléatoires binaires $y = \{C^i, \neg C^i\}$ et $x = \{C^o, \neg C^o\}$, connectées à travers un canal binaire discret sans mémoire (Shannon (1948)), avec $\neg C^i \equiv C_\omega - \{C^i\}$ et $\neg C^o \equiv C_\omega - \{C^o\}$. Les occurrences de la classe d'événement C^i portent de l'information sur les occurrences de C^o dans la séquence ω si et seulement si $p(C^o|C^i) > p(C^o)$. La relation binaire entre C^i et C^o dépend de l'écart entre $p(C^o)$ et $p(C^o|C^i)$. Nous mesurons cet écart par la formule suivante :

$$BJM(C^i \mapsto C^o) = p(C^o|C^i) \cdot \log_2\left(\frac{p(C^o|C^i)}{p(C^o)}\right) + \frac{1}{\|\neg C^o\|} \cdot p(\neg C^o|C^i) \cdot \log_2\left(\frac{p(\neg C^o|C^i)}{p(\neg C^o)}\right) \quad (1)$$

Soit $S = \{C^i \mapsto C^o\}$ un ensemble de relations binaires construites à partir de la séquence ω . Selon la propriété d'absence de mémoire de la chaîne de Markov, les relations binaires contenues dans S sont indépendantes. L'ensemble S est vu comme une succession de plusieurs canaux binaires de transmissions sans mémoire. La BJ-Mesure d'un chemin $M = \{C^i \mapsto$