io-port 05978846

Haugland, Vegard; Kjølleberg, Marius; Larsen, Svein-Erik; Granmo, Ole-Christoffer

A two-armed bandit collective for examplar based mining of frequent itemsets with applications to intrusion detection.

Jędrzejowicz, Piotr (ed.) et al., Computational collective intelligence. Technologies and applications. Third international conference, ICCCI 2011, Gdynia, Poland, September 21–23, 2011. Proceedings, Part I. Berlin: Springer (ISBN 978-3-642-23934-2/pbk). Lecture Notes in Computer Science 6922. Lecture Notes in Artificial Intelligence, 72-81 (2011).

Summary: Over the last decades, frequent itemset mining has become a major area of research, with applications including indexing and similarity search, as well as mining of data streams, web, and software bugs. Although several efficient techniques for generating frequent itemsets with a minimum support (frequency) have been proposed, the number of itemsets produced is in many cases too large for effective usage in real-life applications. Indeed, the problem of deriving frequent itemsets that are both compact and of high quality, remains to a large degree open. In this paper we address the above problem by posing frequent itemset mining as a collection of interrelated two-armed bandit problems. In brief, we seek to find itemsets that frequently appear as subsets in a stream of itemsets, with the frequency being constrained to support granularity requirements. Starting from a randomly or manually selected examplar itemset, a collective of Tsetlin automata based two-armed bandit players aims to learn which items should be included in the frequent itemset. A novel reinforcement scheme allows the bandit players to learn this in a decentralized and on-line manner by observing one itemset at a time. Since each bandit player learns simply by updating the state of a finite automaton, and since the reinforcement feedback is calculated purely from the present itemset and the corresponding decisions of the bandit players, the resulting memory footprint is minimal. Furthermore, computational complexity grows merely linearly with the cardinality of the examplar itemset. The proposed scheme is extensively evaluated using both artificial data as well as data from a real-world network intrusion detection application. The results are conclusive, demonstrating an excellent ability to find frequent itemsets at various level of support. Furthermore, the sets of frequent itemsets produced for network instrusion detection are compact, yet accurately describe the different types of network traffic present.

doi:10.1007/978-3-642-23935-9_7