Multilevel Clustering of Induction Rules: Application on Scalable Cognitive Agent

Amine Chemchem, USTHB, LRIA, Algiers, Algeria
Habiba Drias, USTHB, LRIA, Algiers, Algeria
Youcef Djenouri, USTHB, LRIA, Algiers, Algeria

ABSTRACT
The tremendous size of data in nowadays world web invokes many data mining techniques. The recent emergence of some new data mining techniques provide also many interesting induction rules. So, it’s important to process these induction rules in order to extract some new strong patterns called meta-rules. This work explores this concept by proposing a new support for induction rules clustering. Besides, a new clustering approach based on multilevel paradigm called multilevel clustering is developed for the purpose of treating large scale knowledge sets. The approach invokes k-means algorithm to cluster induction rules using new designed similarity measures. The developed module have been implemented in the core of the cognitive agent, in order to speed up its reasoning. This new architecture called Multilevel Miner Intelligent Agent (MMIA) is tested on four public benchmarks that contain 25000 rules, and compared to the classical one. As foreseeable, the multilevel clustering outperforms clearly the basic k-means algorithm on both the execution time and success rate criteria.

Keywords: Genetics Algorithm, Intelligent Agent, Induction Rules Clustering, Meta Knowledge Discovery, Multilevel Paradigm

INTRODUCTION
Recently, the induction rules have become inseparable pattern of the artificial intelligence thanks to their existence as the basis for many disciplines, such as the agent technology, data mining and knowledge discovery... This paper is about how to extend data mining techniques to induction rules in order to extract meta-rules. For this reason, it would be interesting to exploit the knowledge currently present on the web and proceed to its mining (Han, et al., 2011; Mariscal et al., 2010). The purpose of this paper is to propose a knowledge clustering process and show how to adapt some data mining tasks to knowledge, and also to perform the classical clustering process by the application of the multilevel paradigm.

DOI: 10.4018/ijssoe.2014070101
Let us point out that the concept of knowledge mining is different from the one we found in the literature (Kaufman et al., 2005; Ryszard, 2003; Poongothai, and Sathiyabama, 2012).

To be clear, we are interested in this work in mining knowledge instead of elementary data and the result of the desired task is therefore meta-knowledge.

The only similar study we found in the literature deals with frequent sequential patterns (Saneifar et al., 2008). In fact in this paper, the authors focus on clustering sets of items, and more precisely the sequential patterns. In our case we are interested by studying the induction rules because of their closeness to the natural language.

K-means algorithm is in the top ten of data mining algorithms (Wu et al., 2008). In this work, this latter is extended to induction rules by introducing a new version of similarity measure and gravity center computation. The algorithm called Kmeans-IR is developed and then performed with introducing the multilevel paradigm in order to be able to tackle very large scale knowledge set.

The proposed algorithms are implemented through the core of the agent intelligent, and demonstrated on a public large scale benchmark including 25000 induction rules. The whole idea behind this work is to improve the reasoning engine process by integrating the induction rules clustering module (IRC) in today’s intelligent agent.

The reminder of the paper is organized as follows. The next section presents some related work about the concept of knowledge mining. Afterwards, induction rules representation and new mathematical preliminaries are proposed according to the morphological aspect. Then a novel clustering approach based on multilevel paradigm for induction rules is proposed.

After that, the suggested algorithms are described and followed by the definition of a new architecture for intelligent agent. Then, experimental results are shown compared to the previously proposed algorithms, performed on a benchmark including three different knowledge bases and another one containing hard SAT instances. Finally we conclude by making some remarks and talking about future works.

RELATED WORKS

We recall that clustering data mechanism consists to put the homogeneous data into the same group or class in order to dispatch the heterogeneous data into different groups. In the literature it exists different manner to group data, the two principals kinds are: the hierarchical and the partitioning clustering (Berkhin, 2006).

For the hierarchical clustering, the clusters are inside each others. This category of clustering is used when data can be separated in different levels (Steinbach, 2003; Han et al., 2011).

In our case we use the pure partitioning clustering, it consists to cluster data separately.

K-means is one of the simplest pure partitioning learning algorithms that solves the well known clustering problem (Han et al., 2011; Mac Queen et al., 1967). The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed initially.

The main idea is to define k centroids, one for each cluster. The centroids should be placed in a cunning way because the clustering result depends on their location in the clusters. In order to optimize the efficacy of the outcomes, it is judicious to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set pending, the first step is completed and an early grouping is done. At this point we need to re-calculate k news centroids of the clusters resulting from the previous step and iterates the process. The latter stops when no more changes of the clusters are observed, in other words when centroids do not move any more.

In the literature, we can find many definitions of knowledge mining, for example:

Definition 1: Poongothai & Sathiyabama, (2012) define the knowledge mining concept as an evaluation process of knowledge
Test Case Reduction Using Data Mining Technique
[www.igi-global.com/article/test-case-reduction-using-data-mining-technique/166543?camid=4v1a](www.igi-global.com/article/test-case-reduction-using-data-mining-technique/166543?camid=4v1a)

A Game Theoretic Solution for the Optimal Selection of Services
[www.igi-global.com/chapter/game-theoretic-solution-optimal-selection/77767?camid=4v1a](www.igi-global.com/chapter/game-theoretic-solution-optimal-selection/77767?camid=4v1a)