Knowledge-Based Systems 88 (2015) 165–183

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Case-based maintenance: Structuring and incrementing the case base

Brigitte Chebel-Morello [a,*], Mohamed Karim Haouchine [b], Noureddine Zerhouni [a,b]

[a] FEMTO-ST institute, Besançon, France
[b] ENSMM, Besançon, France

ARTICLE INFO

Article history:
Received 25 July 2014
Received in revised form 1 July 2015
Accepted 26 July 2015
Available online 29 July 2015

Keywords:
Case-based mining
Case-based maintenance
Prototyping
Instance reduction
Auto-increment
Competence
Performance

ABSTRACT

To avoid performance degradation and maintain the quality of results obtained by the case-based reasoning (CBR) systems, maintenance becomes necessary, especially for those systems designed to operate over long periods and which must handle large numbers of cases. CBR systems cannot be preserved without scanning the case base. For this reason, the latter must undergo maintenance operations.

The techniques of case base's dimension optimization is the analog of instance reduction size methodology (in the machine learning community). This study links these techniques by presenting case-based maintenance in the framework of instance based reduction, and provides: first an overview of CBM studies, second, a novel method of structuring and updating the case base and finally an application of industrial case is presented.

The structuring combines a categorization algorithm with a measure of competence CM based on competence and performance criteria. Since the case base must progress over time through the addition of new cases, an auto-increment algorithm is installed in order to dynamically ensure the structuring and the quality of a case base. The proposed method was evaluated through a case base from an industrial plant. In addition, an experimental study of the competence and the performance was undertaken on reference benchmarks. This study showed that the proposed method gives better results than the best methods currently found in the literature.

1. Introduction

Case-Based Reasoning (CBR) is an approach to problem solving and learning through the storing of solutions to similar problems such as cases in a memory called a case base [1]. The case is a body of knowledge representing an experience, defined to a description of the problem resolution event (new case). It is composed of two options: a description of the situation representing a problem and a solution used to remedy this situation (case = (Problem P, Solution S)). A case is placed in a case base and is called a source case which will be used to solve a new case called the target case. A general CBR cycle may be described as having five phases: elaborating, retrieving, reusing, revising and retaining. From a case to be solved, the elaboration phase builds a new problem in a target case by completing or altering the problem description from a possibly incomplete description. According to a similarity metric, cases similar to the target case are first found (the retrieving step) and then adapted by solution construction (reusing). Finally the solution is validated if necessary. The retaining phase consists of storing the new case once validated, provided storage is considered relevant.

A CBR System contains four knowledge containers [34]: (1) the vocabulary knowledge describing the case and the problem domain, (2) the retrieval knowledge including the similarity measure and the indexing method, (3) the adaptation knowledge and (4) the case base itself.

However, most CBR systems, designed to operate over long periods and that must handle large volumes of data and cases encounter problems in the retrieval and adaptation phases. The latter can be costly in terms of time [9]. To maintain the quality of system (i.e.) the speed of the retrieval process, maintenance of the case based reasoning system and particularly of the knowledge containers becomes necessary.

1.1. Maintenance of the CBR system

There are two types of maintenance studies, the maintenance policies and the integration of maintenance with case-based reasoning processes [46]. Reinartz et al. [33] define two phase's steps and tasks (Reviews and Restore) necessary to integrate maintenance into a CBR process. Heister and Wilke [21] propose an architecture integrating maintenance component in a CBR system that is composed of knowledge container and modeling tools.

* Corresponding author.
E-mail addresses: brigitte.morello@femto-st.fr (B. Chebel-Morello), noureddine.zerhouni@femto-st.fr (N. Zerhouni).

http://dx.doi.org/10.1016/j.knosys.2015.07.034
0950-7051/

During maintenance, the contents of each of the four knowledge containers may be revised [45] in order to improve the performance objectives, e.g., the quality of the proposed solution [3].

In past decades, a lot of studies have been done in retrieval knowledge maintenance promoting the performance of similarity measurement. Craw et al. [11] developed a method to optimize CBR retrieval. Bonzano et al. [8] combines introspective learning for feature weighting in CBR. Feature weights for a set of cases are adjusted dynamically during case retrieval by Zhang and Yang [57]. Zhang and Yang [56] maintain the important measures of different features of the case base, by integrating a learning network method for feature weighting within the CBR system. To strengthen the retrieval performance, many works are interested in the case base container and combine problems of feature selection and case organization [54,2,58].

Salamó and López-Sánchez [49,50] and Salamó and Golobardes [47] focused on feature weighting and instance selection methods based on Rough Set Theory (RST).Those works mainly pay attention to the feature weighting and similarity measurement techniques partially or simultaneously

Regarding the maintenance of adaptation knowledge and the case-base, Shiu et al. [51] transform the case-base to a set of small case bases each associated with adaptation rules generated by fuzzy decision tree [45].

Iglezakis and Roth-Berghofer [22] suggest that a CBR system cannot be maintained without scanning the case-base. Maintenance activities are activated only via the case-base which plays a major role in the maintenance of CBR systems. This explains why the bulk of the work done in this field is primarily based on Case-Base Maintenance (CBM). In fact, the case-base is the knowledge container that is the most sensitive to changes in the CBR system and its consultation is essential in order to set maintenance operations in motion [23].

1.2. Case-based maintenance

CBR systems are large scale case bases. It is thus necessary to keep a compact and competent case base [49], to maintain the quality of the case base and the speed of the retrieval process [52]. New content cases must be added, existing cases may need to be revised, and out-of-date cases must be deleted; this is a classic example of the case-base maintenance problem [16].

The k-Nearest Neighbors classifier frequently used in Case-based reasoning remains as (i) one of the most well-known algorithms for supervised non-parametric classification [14] in Pattern Recognition, data mining and Case based maintenance [10,16] (ii) as a benchmark for experimental studies in machine learning [42]. To improve the quality of the case-base, many researchers develop reduction of cases method in the CBR community [55,52,22], or instance reduction in machine learning community and particularly in instance based learning [5]. Reinartz [32] proposes a unifying view on Instance Selection, we base on this proposition in order to first draw the link between the work done in instance reduction and in CBM, review it, and then propose a new approach to reconstructing the case-base.

According to Smiti and Elouedi [52] there are 2 types of CBM policies: **partitioning of the case base** which builds a case-base structure [54,9], and **CBM optimization** which uses an algorithm to delete or update the whole CB [19], (Smyth and McKenna, 1995), [37].

In this study we present, an instance based learning case in the context of classification, where all descriptors are indicated and the solution represents a class.

Different investigations in CBM are reviewed taking into account the framework of Reinartz [32]. Indeed, CBM policies following Reinartz's framework can be presented as clustering (also known as case-base partitioning) and prototyping (also known as case-base optimization) steps. We propose a classification of the case-base optimization taking into account selection criteria and search procedure to complete the taxonomies of CBM algorithms presented in machine learning. Indeed, the most fundamental criteria that allow evaluating the case base quality in case-base optimization are the performance and competence.

We are especially interested in the latter in order to achieve the objectives related to search-time problems as well as to reduce case-base size while preserving its quality. Case-base quality depends on a number of criteria that are discussed in the same section. Section 2 ends with a summary of the state-of-the-art study. Several observations and conclusions are given allowing us to propose our CBM method based on the principle that to maintain a case-base, it is necessary to evaluate and optimize its quality by structuring it and auto-incrementing it with new cases while maintaining its structure.

Section 3 describes a novel method based on the structuring of a case base and its auto-increment. Indeed, after having structured the case base, it will be incremented by new cases. However, this integration must be made under specific conditions in order to ensure system quality. Consequently, several questions arise. Which case is to be retained among those solved? How is the case to be indexed? How can it be introduced with respect for the structure of the expert base, to allow a delicate updating of expertise? What is its contribution to the improvement of expert-base quality? An auto-increment algorithm will therefore be proposed in the same section. Section 4 addresses a comparative study of existing methods conducted on particularly significant benchmarks according to competence and performance criteria. The suggested method is applied in Section 5 via a supervised industrial system of pallet transfer (SISTRE).

2. Related work on case-base maintenance

2.1. Case-base maintenance and instance reduction methods

Maintenance in CBR involves different operations: outdated, redundant or inconsistent cases may be deleted; groups of cases may be merged to eliminate redundancy and improve reasoning power; cases may be re-described to repair incoherencies [37]. Furthermore, case-base maintenance implements policies for revising the organization or contents (representation, domain content, accounting information, or implementation) of the case base in order to facilitate future reasoning [24]. Note that this definition considers the information defining an indexing scheme to be an intrinsic organizational component of the case base itself. Case-base maintenance involves revision of indexing information, links between cases, and/or other organizational structures and their implementation [37,44].

In this context, maintenance is based on applying update policies for case-base representation and is implicated in their reorganization in order to facilitate future reasoning in response to sets of performance objectives. The state-of-the-art of case-base maintenance relies on the various methods used. Case-base maintenance aims to reduce case search time while improving the performance of the system. Time is reduced by minimizing case-base size or by the partitioning it into several smaller parts.

2.2. Unified framework on instance reduction technique and CBM

Hereafter, we present the different points of comparison between these methods and develop the crucial points. There are several different ways to categorize existing case-base maintenance methods. In this study we take an interest in case size