Knowledge Combination vs. Meta-Learning

Ivan Bruha
McMaster University, Canada

INTRODUCTION

Research in intelligent information systems investigates the possibilities of enhancing their over-all performance, particularly their prediction accuracy and time complexity. One such discipline, data mining (DM), processes usually very large databases in a profound and robust way (Fayyad et al., 1996). DM points to the overall process of determining a useful knowledge from databases, that is, extracting high-level knowledge from low-level data in the context of large databases. This article discusses two newer directions in this field, namely knowledge combination and meta-learning (Vilalta & Drissi, 2002).

There exist approaches to combine various paradigms into one robust (hybrid, multistrategy) system which utilizes the advantages of each subsystem and tries to eliminate their drawbacks. There is a general belief that integrating results obtained from multiple lower-level decision-making systems, each usually (but not required) based on a different paradigm, produce better performance. Such multi-level knowledge-based systems are usually referred to as *knowledge integration* systems. One subset of these systems is called *knowledge combination* (Fan et al., 1996). We focus on a common topology of the knowledge combination strategy with base learners and base classifiers (Bruha, 2004).

Meta-learning investigates how learning systems may improve their performance through experience in order to become flexible. Its goal is to search dynamically for the best learning strategy. We define the fundamental characteristics of the meta-learning such as bias, and hypothesis space.

Section 2 surveys the various directions in algorithms and topologies utilized in knowledge combination and meta-learning. Section 3 represents the main focus of this article: description of knowledge combination techniques, meta-learning, and a particular application including the corresponding flow charts. The last section presents the future trends in these topics.

BACKGROUND

So far, commonly utilized decision-making systems have been exploiting a single technique, strategy, or topology. Consequently, their accuracy and overall performance have not been so high (Pratt & Thrun, 1997). New data mining (DM) systems utilize results obtained from several lower-level systems, each usually (but not required) based on different paradigm, or combine or refine them within a dynamic process. Thus, such a multi-strategy (hybrid) system consists of two or more individual 'agents' that interchange information and cooperate together.

It should be noted that there are in fact two fundamental approaches for combining the information from multi-data tasks:

1. In *data combination*, the data sets are merged into a single set before the actual knowledge acquisition.
2. In *knowledge* (*theory*) *combination*, or *sensor fusion*, several agents (base classifiers, sensors) process each input data set separately, and the induced models (knowledge bases) are then combined at the higher-level.

When we look at the issue of the multi-strategy systems from the other side, we come to the *meta-learning*. Generally speaking, meta-learning investigates the way the learning systems can increase their performance and efficiency over experience.

The base learners, the ones with a simple inductive paradigm, such as algorithms inducing decision trees or decision sets of rules, or neural nets, generate a hypothesis (concept description) by applying a fixed bias that is implanted in the knowledge base of the learner. The performance usually increases by larger training sets and losing the restrictions on the hypotheses (concept descriptions).

Using other words, a meta-learner searches dynamically for the best learning strategy and consequently, its performance is flexible. There are a few strategies of the meta-learning, however, various researches recognize it in various ways so that one cannot specify exactly which strategy belongs to meta-learning and which not (Vilalta & Drissi, 2002). Also, there is no sharp boundary between knowledge combination and meta-learning; some researches on machine learning (ML) and DM claim that the first is the subset of the latter, some not. Therefore, this article introduces the most common sights to this issue.

Another taxonomy of these systems distinguishes the way of arrangement of datasets and learning paradigms. We thus differentiate:

1. **Different subsets of training data with a single learning paradigm:** Different subsets are either generated when they are collected, or a single (usually larger) database is split to several subsets, following a certain criterion. Each base learner (with the same learning technique) processes different training subset. Typical examples of such a technique is bagging (Breiman, 1996) and boosting (Freund & Schapire, 1997).

2. **Different training parameters with a single learning paradigm:** Each learning algorithm is accompanied by various parameters that have to be setup. We can thus generate several base learners by changing these parameters, and use then the entire dataset for all the base learners, see for example Bruha (2004).

3. **Different learning paradigms:** The entire multi-strategy system consists of several base learners, each with different learning paradigm (learning system inducing decision trees, that inducing set of decision sets, artificial neural net, genetic algorithm, etc.). These base learners then can process the same database (Kotsiantis & Pintelas, 2004; LiMin et al., 2004).

It should be also noted that there is no uniform terminology in the knowledge-intensive systems (including DM, machine learning, and meta-learning); therefore, we use here usually not a single but several most common terms that can be found in literature.

KNOWLEDGE COMBINATION AND META-LEARNING

Knowledge Combination

A large research in ML focuses on improving topology of classifiers by combining various paradigms into one multi-strategy (hybrid) system which utilizes the advantages of each subsystem and tries to eliminate their drawbacks. There is a general belief that integrating results obtained from multiple lower-lever classifiers produce better performance. We can consider the boosting and bagging algorithms (Bauer & Kohavi, 1999) as already traditional topologies of this approach.

Generally speaking, the main advantages of such hybrid systems are: better performance than that of individual lower-level agents included, the ability to process multivariate data from different information sources, and better understanding of internal data processing when a complex task is solved.

Multi-level knowledge based techniques (called *knowledge integration* systems) can be divided into the following three 'subtechniques:'

1. **Knowledge combination/selection:** The input to such a system is usually formed by several knowledge bases (models) that are generated by various DM algorithms (learners). Each model (knowledge base) independently produces its decision about prediction. These results are then combined into a final decision (knowledge *combination*) or the best decision is selected according to a given statistical criterion (knowledge *selection*).

2. **Knowledge merging:** Several models (knowledge bases) are merged into one robust, usually redundant, model by utilizing statistics that accompany these models.

3. **Knowledge modification (also called revision, refining):** The input is an existing 'old' knowledge base and a 'new' database. A DM algorithm revises (modifies, refines) the current knowledge base according to the knowledge which is 'hidden' in the 'new' database. The new knowledge base thus gets over the 'old' knowledge by being updated by knowledge extracted from the 'new' database.

The first project in this field is evidently (Brazdil & Torgo, 1990); their system merges several decision trees generated by ID3 into a robust one. The already mentioned bagging and boosting algorithms can be viewed as representatives of multi-models. Another direction is formed by the system XCS that is a mixture of genetic algorithms and neural nets (Wilson, 1999). There are several extensions of this system, for example, NXCS (Armano et al., 2002). Another hybrid multisystem combines genetic algorithms with decision trees (Carvalho & Freitas, 2000). All these research projects have revealed that knowledge combination improves the performance of the base classifiers. Knowledge modification is quite often utilized in Inductive logic programming (ILP); they usually use the term 'theory refinement' (Haddawy et al., 2003).

(Fan et al., 1996) introduce the methodology of stacked generalizers and meta-combiners. It can be viewed as learning from information generated by a set of base learners, or using other words, as learning of meta-knowledge on the learned information. The base learners (each usually utilizing a different inductive strategy) induce base classifiers; the base classifiers applied to a training set of examples form so-called meta-database; it is then used by the meta-learner to derive a meta-classifier. The two-level structure of classifiers is then used for making decisions about the input objects.

There are many interesting issues in this field, for example, combining statistical/fuzzy data (probability distribution of classes, quality of decision/performance, reliability of each base classifier), cascade classifiers (Gama & Brazdil, 2000).

Related Content

Moving Up the Mobile Commerce Value Chain: 3G Licenses, Customer Value and New Technology
Martin Fahy, Joseph Feller, Pat Finneganand Ciaran Murphy (2004). *Annals of Cases on Information Technology: Volume 6 (pp. 128-156).*
www.irma-international.org/chapter/moving-mobile-commerce-value-chain/44574

Network Effects of Knowledge Diffusion in Network Economy
Zhang Li, Yao Xiaoand Jia Qiong (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 2778-2782).*
www.irma-international.org/chapter/network-effects-knowledge-diffusion-network/13981

Electronic Government: Assimilation Gaps and Implications
Nabeel A. Al-Qirim (2009). *Journal of Information Technology Research (pp. 53-66).*
www.irma-international.org/article/electronic-government-assimilation-gaps-implications/37409

Organizational E-Mentoring and Learning: An Exploratory Study
Vidya V. Haranand Anand Jeyaraj (2019). *Information Resources Management Journal (pp. 58-72).*
www.irma-international.org/article/organizational-e-mentoring-and-learning/216442

Social Interaction with a Conversational Agent: An Exploratory Study
Yun-Ke. Chang, Miguel A. Morales-Arroyo, Mark Chavezand Jaime Jimenez-Guzman (2008). *Journal of Information Technology Research (pp. 14-26).*
www.irma-international.org/article/social-interaction-conversational-agent/3701