Multimodal Classification: Case Studies

Andrzej Skowron[1], Hui Wang[2], Arkadiusz Wojna[3], and Jan Bazan[4]

[1] Institute of Mathematics
Warsaw University
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

[2] School of Computing and Mathematics
University of Ulster at Jordanstown
Northern Ireland, BT37 0QB, United Kingdom
h.wang@ulst.ac.uk

[3] Institute of Informatics
Warsaw University
Banacha 2, 02-097 Warsaw, Poland
wojna@mimuw.edu.pl

[4] Institute of Mathematics
University of Rzeszów
Rejtana 16A, 35-310 Rzeszów, Poland
bazan@univ.rzeszow.pl

Abstract. Data models that are induced in classifier construction often consist of multiple parts, each of which explains part of the data. Classification methods for such multi-part models are called multimodal classification methods. The model parts may overlap or have insufficient coverage. How to deal best with the problems of overlapping and insufficient coverage? In this paper we propose a hierarchical or layered approach to this problem. Rather than seeking a single model, we consider a series of models under gradually relaxing conditions, which form a hierarchical structure. To demonstrate the effectiveness of this approach we consider two classifiers that construct multi-part models – one based on the so-called lattice machine and the other one based on rough set rule induction, and we design hierarchical versions of the two classifiers. The two hierarchical classifiers are compared through experiments with their non-hierarchical counterparts, and also with a method that combines k-nearest neighbors classifier with rough set rule induction as a benchmark. The results of the experiments show that this hierarchical approach leads to improved multimodal classifiers.

Keywords: hierarchical classification, multimodal classifier, lattice machine, rough sets, rule induction, k-nearest neighbors.

1 Introduction

Many machine learning methods are based on generation of models with separate model parts, each of which explains part of a given dataset. Examples include decision tree induction [20], rule induction [7] and the lattice machine [33].

J.F. Peters and A. Skowron (Eds.): Transactions on Rough Sets V, LNCS 4100, pp. 224–239, 2006.

A decision tree consists of many branches, and each branch explains certain number of data examples. A rule induction algorithm generates a set of rules as a model of data, and each rule explains some data examples. The lattice machine generates a set of hypertuples as a model of data, and each hypertuple covers a region in the data space. We call this type of learning *multimodal learning* or *multimodal classification*.

In contrast some machine learning paradigms do not construct models with separate parts. Examples include neural networks, support vector machines and Bayesian networks.

In the multimodal learning paradigm the model parts may overlap or may have insufficient coverage of a data space, i.e., the model does not cover the whole data space. In a decision tree the branches do not overlap and cover the whole data space. In the case of rule induction, the rules may overlap and may not cover the whole data space. In the case of lattice machine the hypertuples overlap and the covering of the whole data space is not guaranteed too.

Overlapping makes it possible to label a data example by more than one class whereas insufficient coverage makes it possible that a data example is not labeled at all. How to deal best with the overlapping and insufficient coverage issues?

In this paper we consider a hierarchical strategy to answer this question. Most machine learning algorithms generate different models from data under different conditions or parameters, and they advocate some conditions for optimal models or let a user specify the condition for optimal models. Instead of trying to find the 'optimal' model we can consider a series of models constructed under different conditions. These models form a hierarchy, or a layered structure, where the bottom layer corresponds to a model with the strictest condition and the top layer corresponds to the one with the most relaxed condition. The models in different hierarchy layers correspond to different levels of pattern generalization.

To demonstrate the effectiveness of this strategy we consider two multimodal classifiers: one is the lattice machine (LM), and the other one is a rough set based rule induction algorithm RSES-O. We apply the hierarchical strategy in these two classifiers, leading to two new classification methods: HLM and RSES-H.

HLM is a hierarchical version of the lattice machine [33]. As mentioned earlier, the lattice machine generates hypertuples as model of data, but the hypertuples overlap (some objects are multiply covered) and usually only a part of the whole object space is covered by the hypertuples (some objects are not covered). Hence, for recognition of uncovered objects, we consider some more general hypertuples in the hierarchy that covers these objects. For recognition of multiply covered objects, we also consider more general hypertuples that cover (not exclusively) the objects. These covering hypertuples locate at various levels of the hierarchy. They are taken as neighborhoods of the object. A special voting strategy has been proposed to resolve conflicts between the object neighborhoods covering the classified object.

The second method, called RSES-H, is a hierarchical version of the rule-based classifier (hereafter referred to by RSES-O) in RSES [22]. RSES-O is based on rough set methods with optimization of rule shortening. RSES-H constructs