# Learning Compound Decision Functions for Sequential Data in Dialog with Experts

Wojciech Jaworski

Faculty of Mathematics, Computer Science and Mechanics
Warsaw University, Banacha 2, 02-07 Warsaw, Poland
wjaworski@mimuw.edu.pl

**Abstract.** In this paper, we investigate the problem of learning the decision functions for sequential data describing complex objects that are composed of subobjects. The decision function maps sequence of attribute values into a relational structure, representing properties of the object described by the sequence. This relational structure is constructed in a way that allows us to answer questions from a given language. The decision function is constructed by means of rule system. The rules are learned incrementally in a dialog with an expert. We also present an algorithm that implements the rule system and we apply it to real life problems.

**Keywords:** Rough sets, Sequential pattern recognition.

## Introduction

There are two types of sequential data:

– Data describing changes of objects in some evolving process.
– Data describing structural objects.

The first approach is suitable for physical phenomena, e.g., modeled by differential equations. In this case, surroundings of each point in the sequence is a state of object. The second approach is natural while analysing the data generated as the result of purposeful actions composing objects from simpler objects. Such a property have for example textual data, voice, recorded parameters of car on road.

In this paper, we investigate the problem of the decision function learning for data that belongs to the second mentioned type. The considered decision function maps any sequence of attribute values into a relational structure, representing properties of object described by the sequence. This relational structure is constructed in a way that allows us to answer questions from a given language. The decision function is constructed by means of a rule system. In a consequence, the decision function is compound (consists of vast number of rules), and has compound domain. So it belongs to a huge hypothesis space and according to statistical learning theory [11] cannot be learned only from data.

The function is learned in a dialog with an expert. The expert provides us the domain knowledge: He explains his decisions in natural language. We approximate, in a sense, his language by the rule system. The rule system must be compatible with the expert's way of thinking. Then the rules acquired from

S. Greco et al. (Eds.): RSCTC 2006, LNAI 4259, pp. 627–636, 2006.

the expert are tested on data. Finally, the expert is inquired about the cases than does not match to the rules or are classified improperly. In this way, we extract successive fragments of decision functions which converge to the ideal description of the problem.

Due to the sequential character of data the rule system differs from the one used with data represented by tables. Each object is described by a sequence of attribute values. Complex objects can be decomposed into simple ones, which correspond to the split of the attribute value sequence into smaller parts. However, the distance from the beginning of sequence does not distinguish objects. Decision rules are not applied to the attributes according to their absolute position in sequence. Only the relative position of attributes is important for the rule to recognize the pattern.

Rules are used to recognize objects. Successful rule application means that object described by its construction belongs to the upper approximation of the problem. Information about this object is added to the data as a new attribute. The relational structure that describes the object is assigned to the attribute. The other rules may use such an attribute to recognize more complex objects.

The problem discussed in the paper is relates to the objectives of the Information Extraction (IE) [2,6,10].

IE is a subdiscipline of the Natural Language Processing. Its task is to find information, in text written in a natural language, needed to fill a table with description of some event. The attributes of the event are defined a priori.

The main differences follow the characteristics of the data. IE bases on the fact that in modern languages words are marked out with spaces and meaning of ambiguous parts of document can be determined using heuristic methods [1].

Our aim is to process sequential data in general. This implies that we can't take advantage of any a priori defined partition of the data sequence. We concentrate on solving the problem of ambiguity without the necessity of choosing one out of the contradicting interpretations.

We discover the relational structures during the process of learning instead of using a prori defined table.

We adopted the idea of syntax and semantic parsers well known in computer science [3,4,5,8]. Yet we propose our own approach to representation of rules and parser (rule-applying algorithm).

In Section 1, we formally define objects and attributes. In Section 2, we describe rules system properties. In Section 3 and 4, we define rules. In Section 5, we discuss the representation of data. In Section 6, we propose efficient algorithm for applying rules. In Section 7, we discuss the problem of learning the decision rules. In Section 8, we present applications.

## 1 Objects, Attributes, Meanings and Relational Structures

We are given a set of *objects* $\mathcal{U}$, a set of *attributes* $\mathcal{A}$, a set of *meanings* $\mathcal{M}$ and a set of *relational structures* $\mathcal{E}$.