

Název práce: Metody srovnávání pravidel získaných z dat

Autor: Vojtěch Hlaveš

Katedra: Katedra softwarového inženýrství

Vedoucí diplomové práce: Ing. RNDr. Martin Holeňa, CSc.

E-mail vedoucího: martin@cs.cas.cz

Abstrakt: Dobývání znalostí z dat patří k nejrychleji se rozvíjejícím informačním technologiím. Jedním z nejpoužívanějších způsobů strukturovaného vyjádření znalostí jsou speciální typy tvrzení v nějaké logice, označované jako pravidla. Jelikož přístupy založené na různých teoretických principech lze i ze stejných dat extrahovat naprosto rozdílné množiny pravidel, je žádoucí mít k dispozici metody, které budou umět tyto soubory pravidel měřit a srovnávat.

Příspěvkem k rozvoji těchto metod je i tato práce. Nejprve je předveden nový univerzální způsob měření souborů pravidel různých typů, pro který je ukázáno, že rozšiřuje metodu měření souborů klasifikačních pravidel, která byla již známa. Dále je provedeno zobecnění ROC prostorů, které je využito k demonstrování navrženého způsobu měření souborů pravidel při zkoumání závislosti kvality souborů pravidel na různých parametrech použitých metod na dobývání znalostí.

Klíčová slova: Dobývání znalostí, Pravidla, Soubory pravidel, Míry kvality

Title: Methods for comparing rules extracted from data

Author: Vojtěch Hlaveš

Department: Department of Software Engineering

Supervisor: Ing. RNDr. Martin Holeňa, CSc.

Supervisor's e-mail address: martin@cs.cas.cz

Abstract: Data mining has been rapidly developing recently. There is a great variety in types of structured knowledge, but the representation as a sentence of some formal logic is one of the most common. This structure is usually called rules. Because different methods usually produce different sets of rules, the importance of measures which can compare quality of sets of extracted rules grows.

This thesis participates in developing such measures. Firstly, a new and universal approach to measure whole sets of rules is demonstrated and it is shown that this approach extends an already known method of measurement sets of classification rules. Further, ROC spaces are generalized and used to present the proposed approach, which also reveals dependence quality of sets of rules on various parameters of methods extracting data.

Keywords: Data mining, Rules, Sets of rules, Measures of quality