Expert Systems With Applications 62 (2016) 333–346

Contents lists available at ScienceDirect

Expert Systems With Applications

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

What makes classification trees comprehensible?

Rok Piltaver[a,b,*], Mitja Luštrek[a,b], Matjaž Gams[a,b], Sanda Martinčić-Ipšić[c]

[a] Department of Intelligent Systems – Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova cesta 39, Ljubljana 1000, Slovenia
[c] Department of Informatics – University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia

ARTICLE INFO

Article history:
Received 13 January 2016
Revised 7 June 2016
Accepted 7 June 2016
Available online 16 June 2016

Keywords:
Classification tree
Comprehensibility
Understandability
Interpretability
End-user survey

ABSTRACT

Classification trees are attractive for practical applications because of their comprehensibility. However, the literature on the parameters that influence their comprehensibility and usability is scarce. This paper systematically investigates how tree structure parameters (the number of leaves, branching factor, tree depth) and visualisation properties influence the tree comprehensibility. In addition, we analyse the influence of the question depth (the depth of the deepest leaf that is required when answering a question about a classification tree), which turns out to be the most important parameter, even though it is usually overlooked. The analysis is based on empirical data that is obtained using a carefully designed survey with 98 questions answered by 69 respondents. The paper evaluates several tree-comprehensibility metrics and proposes two new metrics (the weighted sum of the depths of leaves and the weighted sum of the branching factors on the paths from the root to the leaves) that are supported by the survey results. The main advantage of the new comprehensibility metrics is that they consider the semantics of the tree in addition to the tree structure itself.

1. Introduction

Classifier comprehensibility, which is sometimes referred to as interpretability (Freitas, 2014; Huysmans, Dejaeger, Mues, Vanthienen & Baesens, 2011; Jin & Sendhoff, 2008; Jin, Sendhoff, & Körner, 2005; Maimon & Rokach, 2005b) or understandability (Allahyari & Lavesson, 2011; Pazzani, 2000; Sommer, 1995), is defined as "the ability to understand the logic behind a prediction of a model" (Martens, Vanthienen, Verbeke, & Baesens, 2011) or "how well humans grasp the induced classifier" (Maimon & Rokach, 2005a). It has been recognised as an important classifier property since the 1980 s (Michalski, 1983) and is continuously emphasised (Allahyari & Lavesson, 2011; Freitas, 2014; Huysmans et al., 2011; Martens et al., 2011; Sommer, 1995; Zhou, 2005; Quinlan, 1999). For example, one of the main features of ID3-like algorithms is their ability to generate easy-to-understand decision trees (Michie, 1987). Similarly, Kodratoff (1994) recognises the comprehensibility as a decisive factor when machine learning models are applied in the industry. Comprehensible classifiers are especially important in domains such as credit scoring, medicine, churn prediction and bioinformatics (Freitas, 2014) because they enable domain experts to classify instances without using a computer, explain classifications of individual instances, validate the classifier, confirm hypotheses, discover new knowledge and improve or refine classifiers in collaboration with data-mining experts (Craven & Shavlik, 1995; Zhou, 2005).

Classifier comprehensibility depends on the type of knowledge representation that is employed (Freitas, 2014; Huysmans et al., 2011; Johansson, Niklasson, & König, 2004; Martens et al., 2011; Zhou, 2005). For example, classification trees and rules are considered to be the most comprehensible (Freitas, 2014; Johansson et al., 2004; Martens et al., 2011; Zhou, 2005), while support vector machines, artificial neural networks and ensembles of classifiers are considered to be the least comprehensible (Chorowski, 2012) and, hence, are termed black-box classifiers (Freitas, 2014; Huysmans et al., 2011; Johansson et al., 2004; Zhou, 2005). There are differences in the comprehensibility of the classifiers based on the same type of knowledge representation as well (Martens et al., 2011): the complexity of a specific classifier (measured as the number of leaves in a tree (Maimon & Rokach, 2005a), conditions in a classification rule set (Sommer, 1995), or connections in a neural network (Jin & Sendhoff, 2008; Liu & Kadirkamanathan, 1995)) is often used as a surrogate metric for classifier comprehensibility (Allahyari & Lavesson, 2011; Freitas, 2003; Freitas, 2004; Jin & Sendhoff, 2008; Jin et al., 2005; Johansson et al., 2004; Martens et al., 2011); a lower complexity corresponds to a higher comprehensibility. However, other properties, such as the structure of the model and its visualisation, affect the comprehensibility as well

* Corresponding author.
E-mail addresses: rok.piltaver@ijs.si (R. Piltaver), mitja.lustrek@ijs.si (M. Luštrek), matjaz.gams@ijs.si (M. Gams), smarti@inf.uniri.hr (S. Martinčić-Ipšić).

http://dx.doi.org/10.1016/j.eswa.2016.06.009
0957-4174/

(Göpferich, 2009; Huysmans et al., 2011), but it is not clear how and to what extent. Therefore, the main problem with regard to most classification algorithms is that they do not explicitly consider the comprehensibility (Huysmans et al., 2011, Johansson et al., 2004), while the ones that do usually simplify it to the classifier complexity (Pazzani, 2000). This approach has several drawbacks (Freitas, 2014) and could lead to over-simplistic models (Elomaa, 1994) that are neither accurate nor comprehensible. This consideration is the motivation for our systematic empirical study of tree properties that potentially influence the comprehensibility of classifiers, in which we tackle classification trees, which are probably the most commonly used type of comprehensible classifiers.

We analyse the comprehensibility through the lens of classifier usability, which is actually the property that is important in practice: the easier a classifier is to comprehend, the easier it is to use. Therefore, the classifier comprehensibility and usability can be interchanged in this paper, although in general, the terms are not exact synonyms (Göpferich, 2009). This study is based on data about the performance of users while solving four types of tasks that involve classification trees and their opinions on the task difficulty and the tree comprehensibility, which was obtained using a carefully designed survey. We collected the answers to 98 questions from 69 respondents and analysed them with statistically sound methodology; we provide the interpretation of the results as well as several empirically supported guidelines on how to construct more comprehensible classification trees. We focus mainly on the influence of the tree structure properties (the number of leaves, branching factor, tree depth) on the comprehensibility, but we also analyse the influence of several tree visualisation properties. One of the most important contributions of this study is the investigation of the influence of the question depth, which is equal to the depth of the deepest leaf that is required to answer a question about a classification tree. Another improvement over the related work is the comparison of the performance and opinions about the tree comprehensibility, from novice versus expert data-miners. Finally, we propose two new classification-tree comprehensibility metrics (the weighted sum of the depths of the leaves and the weighted sum of the branching factors on the paths from the root to the leaves). Comprehensibility metrics are required to act as heuristic functions in learning algorithms (Giraud-Carrier, 1998; Piltaver, Luštrek, Zupančič, Džeroski, & Gams, 2014) and to compare the comprehensibility of the classifiers obtained from various algorithms (Piltaver, Luštrek, Zupančič, et al., 2014; Zhou, 2005).

The paper begins with a review of related work. Section 3 explains the survey design and implementation by listing the general design choices, survey bias prevention strategies, analysed classification tree properties, methods for generating the classification trees used in the survey and survey question examples. Section 4 presents and discusses the survey results. First, the survey and survey respondents are described, followed by a discussion on the performance of different survey respondent groups, the influence of the classification tree parameters on the comprehensibility for each of the survey tasks, and the influence of the classification tree visualisation. The paper closes with a summary of the most interesting findings and suggested directions for further research.

2. Related work

Although many papers emphasise the importance of classifier comprehensibility (Freitas, 2014; Kodratoff, 1994; Martens et al., 2011; Michalski, 1983; Michie, 1987; Quinlan, 1999; Sommer, 1995; Zhou, 2005), related work on classifier comprehensibility is relatively scarce (Allahyari & Lavesson, 2011; Pazzani, 2000). The most general related work comes from the field of cognitive science. Cognitive load theory (Sweller, 1988) divides the total amount of mental effort that is used in working memory into three types: the intrinsic cognitive load is inherent to the specific topic and cannot be altered (the complexity of the classification domain); the extraneous cognitive load depends on the way that information or tasks are presented (the classifier representation); and the germane cognitive load is devoted to the processing and construction of mental structures that organise the categories of information and their relationships. Research in this field has developed a way of measuring the perceived mental effort (Paas & Van Merriënboer, 1993), which motivated us to approach the analysis of the classifier comprehensibility with objective measures. Furthermore, it was shown that experience with a specific task reduces the cognitive load, while the lack of it increases the load (Murphy & Wright, 1984). This concept is addressed in our study by comparing the performance of two groups: data-mining experts (as suggested in Freitas (2014)) and novice data-miners.

More specific studies come from the field of text comprehensibility, where numerous methods for determining comprehensibility have been devised (Göpferich, 2009). Schriver (1989) divides them into three groups and concludes that reader-focused approaches provide advantages over text-focused and expert-judgment-focused approaches. In line with this result, we perform an empirical study that is based on a user survey instead of simply measuring the model complexity, as in Allahyari and Lavesson (2011), Freitas (2003), Freitas (2004), Jin and Sendhoff (2008), Jin et al. (2005) and Martens et al. (2011) or using expert-judgements, as in Freitas (2014).

In the IT field, there are a considerable number of empirical studies that investigate the understandability of conceptual models. Our survey design builds upon the following design issues, which are summarised in a review of experiments from this field (Houy, Fettke, & Loos, 2012): the research design, the number of experiment participants, and the observed dependent variables. In addition, our study follows the framework for empirical evaluation of model comprehensibility (Aranda, Ernst, Horkoff, & Easterbrook, 2007) in all of the recommendations, which can be applied to the classifier comprehensibility.

Finally, a few studies address specifically the classification-tree comprehensibility. Freitas (2014) reviews the case for comprehensible classifier models and discusses the advantages and drawbacks of five types of classification knowledge representations, including classification trees. This work motivated us to study the influence of the question depth on the tree comprehensibility, which has not been empirically evaluated before.

The work by Allahyari and Lavesson (2011) is probably the first empirical study of classification-model comprehensibility. This study compares the comprehensibility of classifiers that are learned by three classification-tree and three rule-learning algorithms, based on subjective comparisons of classifier pairs by 50 students. We focus on classification trees because they are more comprehensible than classification rules (Allahyari & Lavesson, 2011). Furthermore, one of our survey tasks follows the design of their study, comparing classifier pairs. They also report that the classifier complexity has a negative correlation with the understandability, and therefore, we extend their work by analysing the influence of several other tree structure parameters. We also improve on their work by using objective measures and additional survey tasks, including data-mining experts as survey respondents, analysing larger trees and using a domain that is familiar to the respondents.

The study by Huysmans et al. (2011) empirically evaluates the comprehensibility of decision tables, trees and rules using subjective opinions (answer confidence) and objective measures of the respondent performance (time to answer and answer accuracy) for three tasks: classify an instance, verify whether a classifier agrees with a statement, and compare two classifiers for their equivalence. The results of this study are in favour of the single-hit de-