UNIVERSITÉ
LIBRE
DE BRUXELLES

Learning multidirectional functional models with MERCS

Hendrik Blockeel

Topic : Artificial Intelligence

Julien Baudru - 000460130

March 31, 2022

Contents

1 Introduction

This document the synthesis of the seminar given by Hendrik Blockeel during which he briefly presented a new machine learning algorithm that his team and he have developed, this presentation was entitled : **Learning multidirectional functional models with MERCS**.

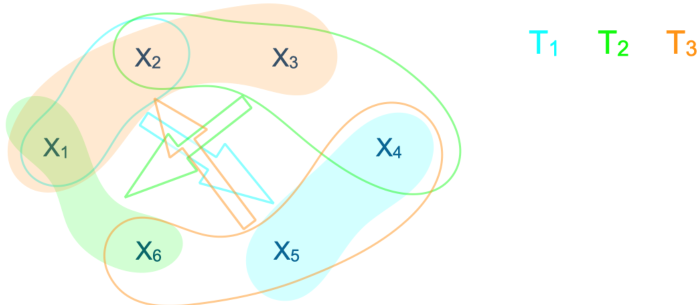

Figure 1: Illustration taken from the speaker's slides

During this seminar the speaker began by introducing the notion of versatile algorithms, i.e. that they indifferently accept multiple types of inputs. In particular, he focused on giving example of the PGM models, which are well known for their versatility. He then explained theoretically how he and his team had built their model named **MERCS**, a machine learning algorithm based on decision trees. It is in fact not a simple decision tree but rather a set of trees, each taking several different source variables to predict a single target variable, different for each tree. Thus, he highlighted the particularities of their model compared to other existing models as well as the performances obtained by it. After that, he explained that besides being able to make predictions, their model was also able to recover missing data in a dataset, he detailed the different versions through which this algortihmes went through to reach this task. At the same time, he compared the performance obtained by each of the different versions with the states-of-the-art methods. And finally, he showed a variant of their model named **AD-MERCS** which is able to detect anomalies in datasets. Moreover, their model also allowed them to highlight the limitations of the benchmarks often used by other anomaly detection models.

(1567 characters)

2 Biography

Hendrik Blockeel was born in 1970 in Kortrijk, Belgium. He started his academic career with a master's degree in Computer Science at the KU Leuven, which he obtained in 1993. He then obtained a second master's degree in 1994 in Artificial Intelligence at the same university. Four years later he obtained the title of Doctor by defending his thesis entitled *Top-down Induction of First-Order Logical Decision Trees*[1]. From a professional point of view, he was a research assistant from 1994 to 1999, then became assistant professor from 2005 to 2009 and finally professor in October 2009. During these years at the KU Leuven he taught among others the following subjects Machine Learning and Inductive Inference, Fundamentals of Computer Science (translated from Dutch) and Data Mining in Practice. In addition, he was General Chair at the Thirteenth International Symposium on Intelligent Data Analysis (IDA 2014). He has also participated in numerous international conferences as a presenter, organizer, and reviewer. In addition to this, he is also an editorial member of four journals: Applied Artificial Intelligence, Machine Learning, Intelligent Data Analysis and New Generation Computing. Still on the subject of scientific journals, he has also been a reviewer for no less than 15 journals. Finally, during his career, Hendrik Blockeel was awarded two prizes, the first one, the Jos Schepens Memorial Fund, for his PhD thesis in 1999 and the second one, the Best demonstration award at ECML/PKDD in 2009 for his demonstration entitled *A Community-based Platform for Machine Learning Experimentation.*

3 The major points discussed

The main points discussed during this seminar are listed and detailed below. They mainly concern the particularity of model created by the speaker and his team.

- **Versatile machine learning:** During this introductive section, the presenter explained the concept of versatile machine learning. He defined it as follows, unlike traditional models that are assigned to a well defined task which is known in advance, versatile models can be used for many different purposes. Versatile models can be designed in two ways, either the model learns all the possibilities in advance (offline) but there can be a lot of them which can make learning time consuming, or it learns them on the fly (online) but this can be slow. The solution proposed here is to build a single model being able to predict any variable from any other set of variables. This model M is therefore able to predict in any direction, i.e. from a set of variables X and Y, M will be able to predict Y from X (direct case) or to derive a specialized function $f_{x \to y}$ from M (inderect case). Finally, to conclude this section, the speaker presented a classic example of a versatile model called a probabilistic graphical model (PGM). The latter are models based on probabilistic inference, for example Bayesian networks or Markov networks. In these models, the random variables are represented by the nodes and the statistical dependency between them are represented by edges.

- **Building versatile models using decision trees:** During this part of the presentation, the speaker detailed how he and his team were able to build a versatile model using prediction trees. First of all, he highlighted the limitations of PGM models, which were the following: although these models use probabilistic reasoning, the probabilistic inference can be very costly in terms of resources (NP-Hard), moreover this type of model is overkill most of the time. Indeed, in the majority of cases we only want to get the closest answer and not the exact answer. This is where decision trees come in, they have many advantages compared to the PGM model, among others they scale very well, they are able to learn quickly, they allow quick predictions and it is easily possible to explain the predictions obtained by them. However, a decision tree used alone does not provide as good results as other models such as Neural Networks (NN) or Support Vector Machines (SVM). To overcome this problem, the authors have proposed to use a set of decision trees instead of a single one. The main idea behind this method is to learn several trees with different versions of the datatsets and then combine the results obtained by them. Then, in order to make this set of decision trees versatile, the method applied is the following: we will create a set of trees for which each tree will predict a target variable from multiple source variables, so each k times, any variable will be predicted by a set of k trees. In the specific case of the **MERCS** (Multi-directional models of regression and classification trees), the model created by this team, to predict p target variables from m source variables, this model will need $\frac{mk}{p}$ trees (see Figure 1). Finally, the presenter proceeded to a presentation of the results obtained by their model on different datasets in comparison with the BayesFusion model, the following results were obtained: Their model has an accuracy comparable to this model, it has a slightly better induction time than the latter and finally it has a prediction time 100 times better.

- **Handling missing values with MERCS:** In this section the speaker explained how their model was able to recover missing values at the time of prediction. To retrieve these values the **MERCS** model uses other trees than those initially used for predictions. He then detailed the techniques used during the development of the different versions of **MERCS**. The first version used the SKL method, the second version used the MRAI method (the different trees operate a vote to find the missing value), the third version used the IT method and finally the last version used the RW method (random walks). These different methods were finally compared to the well-known MissForest imputation method to show that they were globally as efficient.

- **Anomaly detection with MERCS:** In this last section, which was rushed through due to lack of time, Hendrik Blockeel explained how the **MERCS** model was able to detect anomalies in a data set. The main idea is to predict a data point from the set of other data and to compare this prediction with the actual data point and thus detect whether this data point seemed normal or not. A notable example of anomaly detected by their model during experimentation is the scorpion in the Zoo database. Indeed, this animal does not correspond to the pattern normally present for the other animals (See Figure 2). The main conclusion that emerged from their experiments with **AD-MERCS** is the limitations of the benchmarks often used for the detection of anomalies are highlighted. Indeed, these benchmarks focus mainly on a particular type of anomaly.

Animals that ...	*typically ...*	*but a scorpion ...*
do not lay eggs	have teeth	does not have teeth.
do not lay eggs and are not aquatic	give milk	does not give milk.
do not give milk	lay eggs	does not lay eggs.
have no backbone	have no tail	has a tail.

Figure 2: Illustration taken from the speaker's slides

4 Other papers

4.1 Related papers

Among the papers related to the subject presented by Hendrik Blockeel during this seminar, we can notably retain the one written by these colleagues of the KU Leuven Laurens Devos, Wannes Meert, and Jesse Davis entitled *Fast Gradient Boosting Decision Trees with Bit-Level Data Structures*[2]. In this paper the three authors present an alternative to the gradient boosting model for decision trees. The main innovation brought by the authors consists in the acceleration of the execution time required by this type of model. To do so, they used a bitset and bitslices data structure to take advantage of the SIMD instructions of modern CPUs, thus allowing to accelerate the learning from 2 to 10 times compared to before without losing predictive quality. This paper is close to the topic of the seminar for several reasons, firstly it deals with a new method concerning decision trees and secondly the authors of this paper are close to the presenter.

The second paper selected for this section is *A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis*[3] by Kyoungok Kim and Jung-sik Hong. In this paper, the authors described a regression algorithm for processing inputs of different types using decision trees. The innovation of this proposed method is that in addition to being able to treat indifferently numerical inputs or categorical variables, it does not increase the computation cost. This article is very close to the subject treated during this seminar because like the **MERCS**, the algorithm proposed here is versatile in the sense that it accepts several families of inputs and that it is a machine learning algorithm based on decision trees.

4.2 Closely related papers

Among the many articles closely related to the subject treated here, we can mention the one entitled *Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection*[5]. In this paper the authors present one of the state-of-the-art anomaly detection methods, indeed on the basis of the F1-score this method manages to improve by 14% the best scores obtained on several public benchmark datasets. The main innovation of this paper is the use of deep neural network architecture and the use of Gaussian Mixture Model (GMM) instead of the Expectation-Maximization (EM) algorithm usually used for anomaly detection. This paper is indeed related to the topic of the seminar as it also presents a new method for anomaly detection, however the model described by the presenter handles high dimensional inputs better than GMMs.

As another article closely related to the topic of the seminar I chose the one by Daniel J. Stekhoven and Peter Bühlmann, entitled *MissForest: non-parametric missing value imputation for mixed-type data*[4]. In this paper the authors present a method to recover simultaneously missing data of any type. This method named MissForest is based on the well-known random forest technique. This article is indeed related to the seminar because the **MERCS** algorithm is also able to perform this type of task, moreover this algorithm is as good as the MissForest method.

5 Other groups

The field of machine learning being a hot topic, a huge number of researchers from all over the world are working on this particular subject. The most prominent research groups in the field include, among others: The Facebook AI Research group (FAIR), The Deep learning group of University of Oxford or the Google research group called DeepMind. The latter focus mainly on deep learning, however, on the specific topic of decision trees it seems that the research groups being the precursors in this area are the following. The Department of Statistics at the University of Wisconsin who would have invented the concept in 1963 and the Institute of Computing Science in the Poznań University of Technology who would have been the first to publish an article on the subject. We can quote Breiman, Stone, Friedman, and Olshen who seem to have invented the first Decision tree learning named CART.

6 Open science

The author of the conference as well as the colleagues with whom he co-authored his various papers seem to practice open science, indeed all the publications of Hendrik Blockeel are available for free download via his personal website. However, as it was the case for the previous seminars and as it is often the case in the scientific community the source code of the different innovations seems to be accessible nowhere.

7 Two scientific questions

Rereading the section on how **MERCS** is built, given that each tree manages to predict a target variable from multiple sources and that the algorithm uses k trees, it is obvious that the method presented here brings a better robustness in case of missing data. Hence my next question : **At what percentage of missing data is the MERCS aglorithm still able to provide correct predictions?**

My second question is more about the multi-purpose nature of the model, indeed the one described here is said to be versatile. So I wonder : **If it is possible with this model, how an image for example was treated by the model in order to detect an anomaly in practice?** Are all inputs (image, audio, text) normalized in a certain way in input or does **MERCS** use a technique similar to multi-model, using different trees for each type of input? Or is the model closer to what some people call multi-view learning, opting for a kind of data fusion?

8 Criticism(s) about the seminar

Overall I found this seminar well structured and clearly explained. Many examples were given allowing the listener to better understand the explanations given. In addition, each illustration presented during the seminar serves the purpose of the presenter and are not there only for decorative purposes. If I had to find a criticism to address to this seminar, it would be related to my second question. Indeed, it would have been a bonus to know how in practice the model was able to handle different families of inputs.

References

[1] Hendrik Blockeel and Luc De Raedt. "Top-down induction of first-order logical decision trees". In: *Artificial Intelligence* 101.1 (1998), pp. 285–297. ISSN: 0004-3702. DOI: https://doi.org/10.1016/S0004-3702(98)00034-4. URL: https://www.sciencedirect.com/science/article/pii/S0004370298000344.

[2] Laurens Devos, Wannes Meert, and Jesse Davis. "Fast gradient boosting decision trees with bit-level data structures". eng. In: vol. 11906. Brefeld, U. Springer International Publishing, 2020, pp. 590–606. ISBN: 978-3-030-46150-8. URL: https://doi.org/10.1007/978-3-030-46150-8.

[3] Kyoungok Kim and Jung-sik Hong. "A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis". In: *Pattern Recognition Letters* 98 (2017), pp. 39–45. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2017.08.011. URL: https://www.sciencedirect.com/science/article/pii/S0167865517302672.

[4] Daniel J. Stekhoven and Peter Bühlmann. "MissForest—non-parametric missing value imputation for mixed-type data". In: *Bioinformatics* 28.1 (Oct. 2011), pp. 112–118. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr597. eprint: https://academic.oup.com/bioinformatics/article-pdf/28/1/112/583703/btr597.pdf. URL: https://doi.org/10.1093/bioinformatics/btr597.

[5] Bo Zong et al. "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection". In: *International Conference on Learning Representations*. 2018. URL: https://openreview.net/forum?id=BJJLHbb0-.