Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Simić, Srđan Daniel

SVEUČILIŠTE JURJA DOBRILE U PULI
FAKULTET INFORMATIKE

ZAVRŠNI RAD

Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Srđan Daniel Simić

Pula, Rujan 2019.

SVEUČILIŠTE JURJA DOBRILE U PULI
FAKULTET INFORMATIKE

ZAVRŠNI RAD

Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Srđan Daniel Simić

JMBAG: 0303056019

Kolegij: Modeliranje poslovnih procesa

Znanstveno područje: Društvene znanosti

Znanstveno polje: Informacijske i komunikacijske znanosti

Znanstvena grana : Informacijski sustavi, informatologija

Mentor: doc. dr. sc. Darko Etinger

Komentor : dr. sc. Nikola Tanković

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani ____________________________, kandidat za prvostupnika ____ ovime izjavljujem da je ovaj Završni rad rezultat isključivo mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljeni način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

____

U Puli, __________________

Pula, Rujan 2019.

IZJAVA O KORIŠTENJU AUTORSKOG DJELA

Ja, ____ dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj Završni rad pod nazivom

____

____

koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, __________________

Potpis

____

Sažetak

Pregršt dostupnih heterogenih tranzakcijskih podataka te nepredak unutar područja strojnog učenja služe kao ključne sile za razvoj kompleksnih algoritama koji uspijevaju postići preformanse ljudske razine na sve većem broju zadataka.
Zbog svoje ne linearne strukture sastavljene od mnogo slojeva kompjutacije, ti modeli visoke preciznosti obično su primjenjeni kao crne kutije; bez dubljeg razumjevanja o njihovim unutarnjim mehanizmima. To smanjuje transparentost procesa donošenja odluka i često može sadržati skrivene predrasude koji su potencijalno skriveni u podacima. Unutar ovog rada predlažemo alat za generiranje modela za donošenje odluka koji je u skladu s Decision Model & Notation (DMN) standardom baziran na tehnikama smanjivanja kompleksnosti. Ansambl klasifikatora stabla odlučivanja slojevite arhitekture je predložen kako bi kontrolirao "*bias-variance trade-off*". Evaluirali smo preformanse predložene metode na nekoliko javno dostupnih setova podataka usko vezanih za socijalno osjetljivo odlučivanje.
Unutar ovog rada zanima nas automatizacija donošenja odluka koristeči DMN zajedno sa poznatim BPMN alatom za automatizaciju procesa. Primarnih ishod ovog istraživanja je okvir za modeliranje procesa odlučivanja baziranog na prijašnjim ishodima odluka primjenom tehnika strojnog učenja na način bijele kutije. Razumjevanje izlaza modela i zašto je do tog izlaza došlo ključno je za izbjegavanje slučajnih predrasuda.
Naši doprinosi su sljedeći: (1) ansambl metoda strojnog učenja koja je semantički kompatibilna s DMN specifikacijama, time i razumljiva čovjeku, (2) okvir za primjenu ove nove metode na već postojećim zapisima odluka, (3) algoritam za automatsku pretvorbu predloženih modela u DMN artefakte.

Literatura:

Simić S.D., Tanković N., Etinger D. (2020) Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach. In: Ahram T., Karwowski W., Pickl S., Taiar R. (eds) Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing, vol 1026. Springer, Cham

DOI :https://doi.org/10.1007/978-3-030-27928-8_120

Pregled bibliografske jedinice broj: 1021188

Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Simić, Srđan Daniel; Tanković, Nikola; Etinger, Darko
Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach // *Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing, vol 1026.* / Ahram T. ; Karwowski W. ; Pickl S. ; Taiar R. (ur.).
Cham: Springer, 2020. str. 789-794 doi:10.1007/978-3-030-27928-8_120 (predavanje, međunarodna recenzija, cjeloviti rad (in extenso), znanstveni)

Naslov
Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Autori
Simić, Srđan Daniel ; Tanković, Nikola ; Etinger, Darko

Vrsta, podvrsta i kategorija rada
Radovi u zbornicima skupova, cjeloviti rad (in extenso), znanstveni

Izvornik
Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing, vol 1026. / Ahram T. ; Karwowski W. ; Pickl S. ; Taiar R. - Cham : Springer, 2020, 789-794

ISBN
978-3-030-27928-8

Skup
IHSED 2019: Human Systems Engineering and Design II

Mjesto i datum
Minhen, Njemačka, 16-18.09.2019.

Vrsta sudjelovanja
Predavanje

Vrsta recenzije
Međunarodna recenzija

Ključne riječi
Machine learning ; Automated decision making ; White-box models

Sažetak
Plethora of available heterogeneous transactional data and recent advancements in machine learning are the key forces that enable the development of complex algorithms that can reach human-level performance on an increasing number of tasks. Given the non-linear structure composed of many layers of computation, these highly accurate models are usually applied in a black-box manner: without a deeper understanding of their inner mechanisms. This hinders the transparency of the decision-making process and can often incorporate hidden decision biases which are potentially present in the data. We propose a framework for generating decision-making models conforming to Decision Model & Notation standard based on complexity-reducing techniques. An ensemble of decision-tree classifiers in a layered architecture is proposed to control the bias- variance trade-off. We have evaluated the performance of the proposed method on several publicly available data-sets tightly related to socially sensitive decision-making.

Izvorni jezik
Engleski

Znanstvena područja
Računarstvo, Informacijske i komunikacijske znanosti

POVEZANOST RADA

Ustanove
Sveučilište Jurja Dobrile u Puli

Autor s matičnim brojem:
Darko Etinger, (285273) (https://www.bib.irb.hr/pregled/znanstvenici/285273)
Nikola Tanković, (356610) (https://www.bib.irb.hr/pregled/znanstvenici/356610)

CERTIFICATE OF PARTICIPATION

Srđan Daniel Simić
Juraj Dobrila University of Pula, Croatia
ssimic@unipu.hr

September 26, 2019

To Whom it May Concern,

This is to certify that Srđan Daniel Simić attended the 2nd International Conference on Human Systems Engineering and Design: Future Trends and Applications (IHSED 2019) which was held at Universität der Bundeswehr München, Munich, Germany, September 16-18, 2019 and presented the following:

Paper ID: 133

Presentation Title: Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Sincerely,

Dr. T. Ahram
Scientific Program Chair
IHSED 2019 Administration

Questions? Please send to support@ihsed.org
Conference website: http://ihsed.org

Automated Decision Modeling with DMN and BPMN: A Model Ensemble Approach

Srđan Daniel Simić[1], Nikola Tanković[1], and Darko Etinger[1]

[1] Juraj Dobrila University of Pula, Faculty of Informatics, Pula, Croatia
{ssimic, ntankov, detinger}@unipu.hr
https://fipu.unipu.hr/fipu/en

Abstract. Plethora of available heterogeneous transactional data and recent advancements in machine learning are the key forces that enable the development of complex algorithms that can reach human-level performance on an increasing number of tasks.
Given the non-linear structure composed of many layers of computation, these highly accurate models are usually applied in a black-box manner: without a deeper understanding of their inner mechanisms. This hinders the transparency of the decision-making process and can often incorporate hidden decision biases which are potentially present in the data. We propose a framework for generating decision-making models conforming to Decision Model & Notation standard based on complexity-reducing techniques. An ensemble of decision-tree classifiers in a layered architecture is proposed to control the bias-variance trade-off. We have evaluated the performance of the proposed method on several publicly available data-sets tightly related to socially sensitive decision-making.
Keywords: Machine learning · Automated decision making · White-box models

1 Introduction

Understanding the behavior of intelligent agent algorithms is the key step in understanding what effects their outputs or decisions will have on our business processes. Generally, there is a need to reaping their benefits without accessing their harmful parts. Thus, a whole new field of science on intelligent machines as actors with behavior and their whole surrounding ecology was recently proposed [1]. The ecology part of machine learning algorithms is mostly considered to be their evolution which is affected through interaction with their surroundings and exchanging new information among themselves and their business operators.

In this paper, we are interested in the decision-making automation by applying Decision Model & Notation (DMN) [2] coupled with well-known BPMN process automation tools [3]. The primary outcome of this research is a framework for modeling decision-making processes based on the previous decision outcome data by applying machine learning techniques in a white-box manner. Understanding the model outputs, and the reasoning behind that output is the key requirement in avoiding unintentional bias.

Our contributions are the following: (1) a machine-learning ensemble method which is semantically compatible with the DMN specification, and thus human understandable, (2) a framework for applying this novel method on the existing decision logs - transactions of decision outcomes, (3) an algorithm for the automatic transformation of our proposed model into DMN artifacts.

The rest of the paper is organized as follows: Section 2 brings the definitions and motivation behind our contributions, Section 3 disseminates our contributions. Section 4 evaluates the proposed methods on three decision-making data-sets, and Section 5 concludes the paper with suggestions for future research.

2 Background and Motivation

The implementation of algorithmic operational decisions, an important driving force for a modern organization, should not be left solely to the software engineering teams. Automated decision-making should be carefully implemented, evaluated and continuously updated with the organizational business goals. Hence the implementation of optimal business decisions should be handled on the management level since documenting business requirements and offloading the implementation to software engineers is less efficient and error-prone due to misunderstandings. There is a level of model-driven automation present, but still it has not reach wider adoption [4].

The transition from a classical to a digital economy has brought new challenges: the three V's of Big Data: volume, velocity, and variety make the process of decision-making a challenging one and in a need for a continuous adaptation. Machine learning models in the role of predictive analytical models are the cornerstone in seeking insight into internal and external organizational factors.

Replacing multiple decision points embedded in business processes by pro- viding a method to derive DMN decision tables from the corresponding machine learning model was implemented in previous works [5]. But an important challenge still remains: how to extract knowledge from increasingly complex machine learning models that try to optimize the results based on historical transactions data in which unavoidable past decision biases are deeply concealed? Failure in efficiently implementing machine learning models can lead to undesired con- sequences [6]. The goal of Business Decision Management should be to gather insight from predictive methods, evaluate the biases and find the best trade-off in order to confidently deal with the risk of model drifts, regulatory compliance, and decision outcome performance [7].

3 Decision framework proposal

The overall framework steps can be observed in Figure 1. The prerequisites for the proposed framework is the implementation of the ETL process (extract-transform-load) of data from the relational database into the Decision Tool modeling environment.

The framework is composed of three automated steps: (1) Feature proposition - based on the data provided, important features are provided to the business analysts (2) DMN model proposition - a layered machine-learning model is recommended based on the chosen features, (3) DMN table generation - automated generation of DMN tables from the proposed layered model.

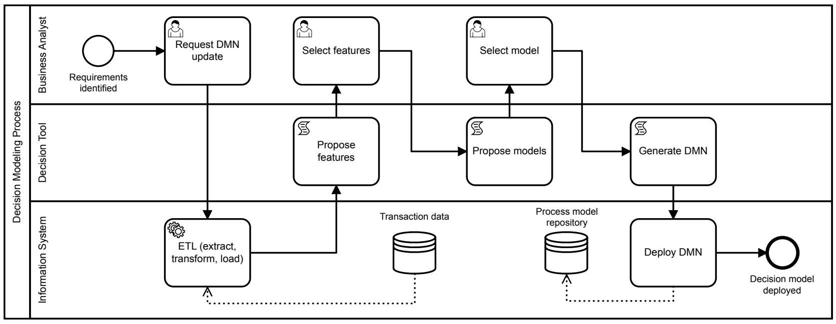

Fig. 1. The workflow of proposed Decision Framework

3.1 Feature proposition task

The importance of individual features $f_i \in \Phi$, is determined based on training a Gradient Boosting Machine (GBM) classifier [8] on the data resulting from the ETL process. GBM consists of a desired number (n) of weak decision tree learners $wl_i, i \in [1, n]$ where feature importance corresponds to the number of splits throughout wl_i decision trees, or on the average gain of the feature when it was used in the model. The score of each feature $S(f_i)$ is reported in the interval [0, 1]. Decision tool then reports the proposed features to the user in the descending order of importance $S(f_i)$ and automatically suggests an optimal number of features. Optimal number of features is generating in a following way: $\bar{\Phi} = \{f_i \mid S(f_i) > \psi * \bar{S}\}$ where $\bar{S}$ is the mean feature importance and ψ is the inclusion parameter, which we recommend to set in the range $[0.1, 0.5]$ based on the case studies we conducted. In this step special care should be given to exclude the features that should not contribute to the model according to regulations in the given domain (e.g. gender, ethnicity).

3.2 Model proposition task

Based on the agreed set of features with the business analyst, the decision framework tool is able to propose an ensemble of N_L decision-tree models: $\mathbb{E} = \{tm_i | i \in [1, N_L]\}$. The parameter N_L is configurable and reflects the overall complexity of the desired ensemble model. The models are organized in a layered style so that model

tm_i suggests a decision $D(tm_i, \mathbf{x}_i)$ for an input vector $\mathbf{x}_i$ only if the purity of the decision leaf $l(tm_i, \mathbf{x}_i)$ is higher than the configurable threshold π_i for that level, that is $purity(l(tm_i, \mathbf{x}_i)) >= \pi_i$. Otherwise, the decision is offloaded to the next model $tm(i+i)$. In case that $i+1 > N_L$ the final decision is left to the human operator. Such a layered composition achieves two goals: (1) It a achieves a human-understandable composition of decision-tree models that can automatically be converted to DMN models, (2) It offloads the decisions where there is lower uncertainty of the outcome to human operators.

3.3 DMN model generation

Algorithm 1 contains the pseudo-code for generating the rules of DMN models for each decision-tree tm_i on each level. Final DMN models, one per each layer are composed into the final runtime BPMN process where the final decision in the case of uncertainty is left to human operators. Figure 2 depicts a case where $N_L = 2$.

Algorithm 1: Create DMN table rows

```
procedure generateTableRow(mlDictionary, annotation, isLastTable)
Input: mlDictionary is dictionary containing decision tree features, thresholds and logic
    for className, value in mlDictionary do
        newRule ← create new sub element of decision table
        if isLastTable == True then
            newAnnotation ← create annotation for rule with value annotation
        foreach featureName in value do
            thresholdSignList ← list of thresholds for each featureName
            if len(thresholdSignList) == 0 then
                ruleCell // empty in this case

            else if len(thresholdSignList) == 1 then
                ruleCell ← contains threshold and threshold sign
                // if threshold sign is "is" or "not" threshold becomes string value
            else
                if thresholdSignList[0] is "<=" then
                    if threshold[0] > threshold[1] then
                        ruleCell ← create range cell
                if thresholdSignList[0] is ">" then
                    if threshold[0] < threshold[1] then
                        ruleCell ← create range cell
                if thresholdSignList[0] is "not" or "is" then
                    ruleCell ← if positive signs in list, write them first, negation last
        end
    end
    newOutputCell ← rule sub element, contains className
```

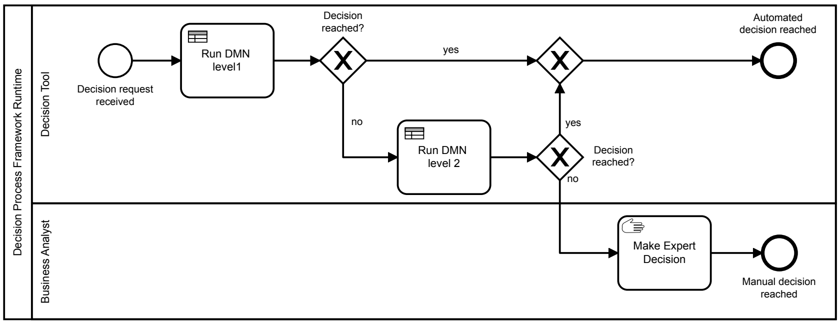

Fig. 2. Decision Process Framework Runtime

4 Evaluation

We have evaluated our layered ensemble model on three distinct data-sets: (1) Credit risk assessment dataset[1], (2) Medical insurance claims dataset[2] and (3) Risk of prisoner recidivism dataset[3] and compared the results of the layered model against the simple decision-tree classifier and the black-box GBM model. The results are expressed through a *F1* score [9] and presented in Table 4. Results are obtained through a 10-fold leave-one-out cross-validation technique. We can observe that from our preliminary results, our layered model with $N_L = 5$ obtains more accurate results than simple decision-tree models. Such results are closer to more complex, but incomprehensible GBM models with 300 trees in ensemble. A much more comprehensive evaluation is left for future studies. It is our goal not to compare only the accuracy and the recall of the model (*F1* score) but also the complexity and present the trade-offs when tuning the parameters of our layered model. Each decision generated by the GBM model in these examples is calculated from the whole set of 300 decision trees, where our approach can map a decision only to a single tree (from a specific layer). We argue that this is a significant step forward in model comprehensibility.

Table 1. Preliminary evaluation results based on the F1 score.

	Decision-tree	**Layered model**	**GBM**
Credit-risk	0.700 ± 0.023	0.804 ± 0.028	0.830 ± 0.030
Insurance claim	0.942 ± 0.021	0.952 ± 0.018	0.987 ± 0.011
Recidivism	0.644 ± 0.036	0.650 ± 0.019	0.654 ± 0.022

[1] Available at https://www.kaggle.com/uciml/german-credit
[2] Available at https://www.kaggle.com/easonlai/sample-insurance-claim-prediction-dataset
[3] Available at https://github.com/propublica/compas-analysis

5 Conclusion

In this paper we have presented a layered decision-tree model that is semantically compatible with the DMN specification. We have proposed a Decision-making framework and a tool for automatically building layered models and translating them into the within BPMN processes. In a brief preliminary evaluation, we have also showed that our approach reaches higher accuracy than the simple decision tree models, support-vector machines and linear models. In the future research we plan to conduct a more thorough evaluation of our proposed methods, as well as provide an optimization framework in tuning our model parameters.

References

1. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J-F., Breazeal, C., Crandall, J.W., Christakis, N.A., Couzin, I.D., Jackson, M.O., Jennings, N.R., Kamar, E., Kloumann, I.M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D.C., Pentland, A., Roberts, M.E., Shariff, A., Tenenbaum, J.B. and Wellman M.: Machine behaviour. Nature, 568(7753):477–486, apr 2019.
2. Decision Modelling Notation (DMN) 1.2 Specification, 2019. Version 1.2.
3. Business Process Model and Notation (BPMN) 2.2 Specification, 2014. Version 2.2.
4. Tanković, N., Vukotić, D. and Žagar, M.: Rethinking model driven development: analysis and opportunities. In Proceedings of the ITI 2012 34th International Conference on Information Technology Interfaces, pages 505–510. IEEE, 2012.
5. Etinger, D., Simić, S.D. and Buljubašić, L.: Automated decision-making with DMN: from decision trees to decision tables. In Proceedings of the 42nd Inter- national Convention on Information and Communication Technology, Electronics and Microelectronics. Croatian Society for Information and Communication Technology, Electronics and Microelectronics, 2019.
6. Dressel J. and Farid H.: The accuracy fairness, and limits of predicting recidivism. Science Advances, 4(1), jan 2018.
7. Sinur J., Geneva, R., Debevoise, T. and Taylor J.: The MicroGuide to Process and Decision Modeling in BPMN/DMN: Building More Effective Processes by Integrating Process Modeling with Decision Modeling. CreateSpace Independent Publishing Platform, oct 2014.
8. Ayyadevara, K.V.: Gradient Boosting Machine. In Pro Machine Learning Algorithms, pages 117–134. Apress, 2018.
9. Goutte, C. and Gaussier, E.: A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In Proceedings of the 27th European Conference on Advances in Information Retrieval Research, ECIR'05, pages 345– 359, Berlin, Heidelberg, 2005. Springer-Verlag.