Talking to your Data

Talking to your Data: Interactive and interpretable data mining through a conversational agent*

Isel Grau[1][0000−0002−8035−2887], Luis Daniel Hernandez[1], Astrid Sierens[1], Simeon Michel[2], Nico Sergeyssels[2], Vicky Froyen[3][0000−0002−5649−5888], Catherine Middag[2][0000−0001−5732−0281], and Ann Nowe[1][0000−0001−6346−4564]

[1] Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium
[2] Gezondheidszorg, Design & Technologie, Erasmushogeschool Brussel, Belgium
[3] Collibra NV, Belgium

Abstract. In this demo, we showcase the "Talking to your Data" system. The key idea of this system is to support data governance and data mining in a novel way. We aim to bring the use of interpretable machine learning techniques closer to the business analysts by using natural language. We have developed a conversational agent and a data mining backend that supports the analysis of data. Our approach facilitates solving prediction tasks and also provides explanations for these predictions. Furthermore, we make possible the interaction for including the feedback of the business analysts in the models.

Keywords: conversational agents · decision tree · subgroup discovery · interactive machine learning · explainable artificial intelligence

1 Introduction

The Collibra project [6] aims to develop a platform for supporting data management through smart engagement using a conversational agent. The goal is to go beyond the level of reports, incorporating interpretable data mining models to gain new insights into the data. Here, by interpretability we refer to the transparency of the model, i.e. the model is referring to terms familiar to the user and the user can understand the reasoning of the model [4].

By adding the human in the process of building or fine-tuning a machine learning model, the users' understanding and trust of the system, as well as the accuracy of learned systems, can be improved. Decisions trees (DT) [8] are one of the most widely used intrinsically interpretable machine learning techniques [7]. While the lesser-known (but also interpretable) subgroup discovery techniques are focused on generating descriptions of interesting patterns in data [5]. Other works have proposed interactive machine learning tools for building and visualizing machine learning models [10, 9], but they rely on traditional graphical user interfaces.

* Supported by the Innoviris TeamUp project "Driving collective data governance through smart engagement platforms".

2 System description

In this work, we propose the use of a conversational agent for the interaction with data and interpretable machine learning techniques. The conversational agent was implemented using the RASA library [1], which facilitates the dialogue management and language understanding/generation modules. Our systems relies in two backend services, the Collibra Data Governance Platform [2] and a data mining backend. The Collibra Platform manages all requests regarding reports, data editions, and permissions, while the data mining backend processes all machine learning-related tasks, such as learning, prediction, interpretation, and edition of the models. The system architecture is depicted in Figure 1.

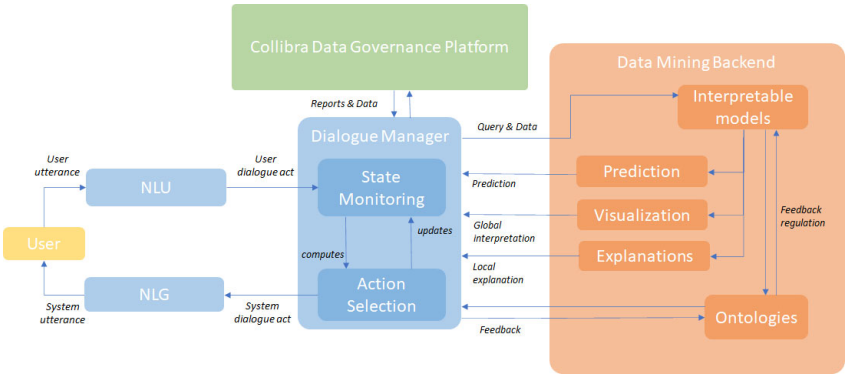

Fig. 1. System architecture of "Talking to your Data", involving the conversational agent (blue), the Collibra Platform backend (green) and the data mining backend (orange).

Our system supports common operations with data that are needed during the exploratory phase of the data mining process. For example, loading or merging datasets, requesting the possible values of a feature, and performing group-by operations offering aggregation statistics. For the predictive phase, we allow to train decision tree models and subgroup discovery algorithms. The latter also providing the possibility of intervening during the optimization process [3]. After the machine model is built, it can be questioned in natural language for obtaining predictions, even with incomplete information. Perhaps the most relevant features are the possibility to obtain explanations over the predictions in the form of rules and to modify those rules based on the feedback of the user, thus changing the trained model with the knowledge of the expert. For this last feature, we rely on ontologies associated with the datasets, which allows controlling the vocabulary, finding alternative features, and reusing the calculations already performed by the classifier. *System requirements for demonstration: Two screens and internet connection.* Video available at: https://youtu.be/SaigB3usp6U

References

1. Bocklisch, T., Faulkner, J., Pawlowski, N., Nichol, A.: Rasa: Open Source Language Understanding and Dialogue Management (2017), http://arxiv.org/abs/1712.05181
2. Collibra: Collibra Data Governance: Organize and understand your data — Collibra, https://www.collibra.com/data-governance
3. Dzyuba, V., van Leeuwen, M.: Interactive discovery of interesting subgroup sets. In: International Symposium on Intelligent Data Analysis. pp. 150–161. Springer (2013)
4. Grau, I., Sengupta, D., Garcia Lorenzo, M.M., Nowé, A.: An Interpretable Semi-supervised Classifier using Rough Sets for Amended Self-labeling. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE (2020)
5. Helal, S.: Subgroup Discovery Algorithms: A Survey and Empirical Evaluation. Journal of Computer Science and Technology **31**(3), 561–576 (2016). https://doi.org/10.1007/s11390-016-1647-1
6. Loeckx, J., Grau, I., Sergeyssels, N., Michel, S., Froyen, V., Middag, C., Nowe, A.: Driving Collective Data Governance through Smart Engagements Platforms (Collibra) (2018)
7. Molnar, C.: Interpretable Machine Learning. Leanpub (2019)
8. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
9. Van Den Elzen, S., Van Wijk, J.J.: BaobabView: Interactive construction and analysis of decision trees. In: VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings. pp. 151–160 (2011). https://doi.org/10.1109/VAST.2011.6102453
10. Ware, M., Frank, E., Holmes, G., Hall, M., Witten, I.H.: Interactive machine learning: Letting users build classifiers. International Journal of Human Computer Studies **55**(3), 281–292 (2001). https://doi.org/10.1006/ijhc.2001.0499