

Explainable AI Beer Style Classifier

Jose M. Alonso¹, A. Ramos-Soto^{1,2}, C. Castiello³, and C. Mencar³

¹ Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS),
Universidade de Santiago de Compostela, Santiago de Compostela, Spain
{josemaria.alonso.moral,alejandro.ramos}@usc.es

² Department of Computing Science, University of Aberdeen, United Kingdom,
alejandro.soto@ac.abdn.uk

³ Department of Informatics, University of Bari “Aldo Moro”, Bari, Italy,
{ciro.castiello,corrado.mencar}@uniba.it

Abstract. This paper describes how to build an eXplainable Artificial Intelligent (XAI) classifier for a real use case related to beer style classification. It combines an opaque machine learning algorithm (Random Forest) with an interpretable machine learning algorithm (Decision Tree). The result is a XAI classifier which provides users with a good interpretability-accuracy trade-off but also with explanation capabilities. First, the opaque algorithm acts as an “oracle” which finds out the most plausible output. Then, we generate a textual explanation of the given output which emerges as an automatic interpretation of the inference process carried out by the related decision tree, if the outputs from both classifiers coincide. We apply a Natural Language Generation Approach to generate the textual explanations.

Keywords: Explainable Artificial Intelligence, Classification Task, Natural Language Generation

1 Introduction

Ethical and legal issues become essential to guarantee the success of (artificial intelligence) AI applications which interact with humans into real-world usage [1]. Accordingly, a new European General Data Protection Regulation (GDPR) just took effect in May 2018. Among other issues, the new GDPR is somehow related to a “right to explanation” [2, 3].

Researchers have tackled how to model explainability since the pioneer expert systems [4]. More recent studies refer to the structure and function of explanations [5]. Moreover, the generation of explanations is also a hot topic in the field of decision-support and recommendation systems [6]. We observe that machine learning (ML) researchers, in the broad sense, are more and more aware of the need to consider explainability while designing ML algorithms [7].

The rest of the manuscript is organized as follows. Section 2 introduces the architecture of the XAI classifier. Section 3 describes how our proposal was evaluated by humans in an on-line survey. Finally, Section 4 concludes this paper and pinpoints future work.

2 The Design of the XAI Classifier

First of all, we gather all the available information (data and expert knowledge) related to the classification problem under consideration. Then, we build the components of our XAI classifier in an off-line stage as follows:

1. We build an accurate yet non-interpretable ML-based classifier. It acts as an “oracle” to guide the classification procedure. In this paper, we consider the Random Forest (RF) algorithm [8]. It produces an ensemble of decision trees (C4.5), each one depending on the values of a random vector sampled independently. Notice that C4.5 [9] is known for its ability to generate classifiers allowing easy interpretation by humans, no matter their background. Given a data sample, the inference procedure starts from the root of the tree and follows a branch by evaluating conditions on attributes for each intermediate node; eventually, the class corresponding to the leaf node is returned. Even though the single classifiers in RF are deemed interpretable, their random combination is hardly interpretable.
2. We build one interpretable ML-based classifier: J48 is the Weka method implementing C4.5 [10].

Then, given a data example to be classified (on-line stage), the system decision is computed as follows. If the class selected by the “oracle” (ML class) is supported by the interpretable classifier, then this class is considered as the output class. The related explanation is based on the automatic interpretation of the related branches in the tree. Otherwise, the system informs the user that it is not able to provide him/her with an explanation of the given decision.

A Natural Language Generation (NLG) module produces the textual explanation associated to the output class. Notice that NLG is a well-known area inside the computational linguistics field [11]. In Data-to-Text applications, NLG plays a key role for generating text that is readable and understandable by humans [12]. Many NLG systems have been developed on the basis of the generic methodology and the architectural framework proposed in [13].

We have already shown the benefits of using Natural Language for explaining classification tasks [14, 15]. Here, we use the open source software rLDCP [16], which is an R package based on the Computational Theory of Perceptions [17].

We design the granular linguistic model (GLMP) which describes how to interpret in natural language the classification procedure carried out by a decision tree previously learnt from data. The GLMP is a hierarchical network of perception mappings (PMs) and computational perceptions (CPs). Each CP is defined as a tuple (A, W_A, R_A) where A is a vector of linguistic expressions, W_A and R_A are, respectively, vectors of validity and relevance degrees associated to A . Each PM is defined as a tuple (U, y, g, T) where U is a vector of inputs (being numerical values at the lowest level of the hierarchy and CPs in the rest of levels), y is the output CP, g is the aggregation function which translates inputs into output, and T is the text generation algorithm (in the simplest case just a text template).

Given a decision tree, the GLMP is built bottom-up. In the lowest level of the hierarchy, there is a PM associated to each input variable. In the second level, there is a PM associated to each branch in the tree. At the top of the hierarchy, there is a PM which aggregates all the branches and gives the output class.

3 Use Case

We design an XAI classifier regarding 8 beer styles (Blanche, Lager, Pilsner, IPA, Stout, Barleywine, Porter, and Belgian Strong Ale) in terms of 3 attributes (color, bitterness and strength). The dataset is made up of 400 instances [18].

We applied 10-fold cross-validation as evaluation methodology (accuracy is measured in terms of classification rate). Table 1 reports the main values related to all the classifiers considered in our approach. On the one hand, the performance of classifiers is described in terms of average accuracy. On the other hand, to describe the performance of the interpretable classifier (J48), the accuracy is accompanied by a basic interpretability metric (NB) which counts the number of branches in a tree. Gain is measured as the accuracy improvement provided by the RF-based XAI approach with respect to RF considered alone. EXP-J48 represents the percentage of examples for which the explanation of the XAI classifier is based on the tree generated with J48.

As it can be appreciated in Table 1, the XAI classifier increases accuracy (gain is +1.25) while the percentage of examples without associated explanation (1.5%) is almost negligible. Notice that the XAI classifier developed in this paper outperforms the systems developed in [18] from the point of view of accuracy: an expert system (8 rules with 3 inputs) with 81.25% of classification rate; and a fuzzy rule-based classifier (9 rules with 2 inputs) with 87.5% of classification rate. Moreover, our XAI classifier also offers to users added-value textual explanations.

We assessed the quality of the generated explanations in an on-line survey which was announced via email and via social networks. We showed five different samples in five consecutive screens to each interviewee. Notice that we first presented the decision made by the XAI classifier along with a related picture. Then, we included the related explanation along with additional details (see Fig. 1). Afterwards, we asked interviewees to assess each explanation by rating the following statements on a five-level Likert scale: (S1) “The information above helps you understand the system’s decision”; (S2) “The explanation is clear and does not need to be re-written in a clearer way”; and (S3) “Providing decision along with explanation helps you to trust the system”.

The main targets of our poll were persons with a Bachelor of Science (BSc) and native in English (we considered people living in English-speaking countries

Table 1. Averaged classification results over 10-fold cross-validation.

RF	J48		RF-based XAI classifier		
Accuracy	Accuracy	NB	Accuracy	Gain	EXP-J48
95	94.75	10.7	96.25	+1.25	98.5

	Decision	Explanation
	The beer is Blanche	It is very likely that this beer is Blanche, because its color is pale, its bitterness is low, and its strength is session.
	Additional Details	
Color	<ul style="list-style-type: none"> Color is pale because it is equal to 2 in terms of the Standard Reference Method (SRM). SRM gives values from 0 to 45. Linguistic scale: [Pale, Straw, Amber, Brown, Black]. 	
Bitterness	<ul style="list-style-type: none"> Bitterness is low because it is equal to 18 in terms of International Bittering Units (IBU). IBU gives values from 7 to 250. Linguistic scale: [Low, Low medium, Medium high, High]. 	
Strength	<ul style="list-style-type: none"> Strength is session because it is equal to 0.049 in terms of Alcohol By Volume (ABV). ABV gives values from 0.035 to 0.136. Linguistic scale: [Session, Standard, High, Very high]. 	

Fig. 1. Example of output given by the XAI classifier.

as native persons). We got feedback from 26 anonymous interviewees (84.61% with a BSc). The average score was: S1=3.61; S2=2.98; S3=4.05. Notice that average score around 4 for S3 validates our XAI classifier as a useful system for the target audience. In addition, the lower score associated to S2 suggests that the system may be further enhanced by refining the generated text.

4 Conclusions and Future Work

We have presented a first pilot of a beer style XAI classifier. Preliminary experimental results are encouraging. The designed XAI classifier overwhelms (regarding accuracy) other classifiers reported in the literature for the same dataset, while providing the added-value of textual explanations. The utility and operational value of the generated explanations were evaluated in an on-line survey.

Even though we have shown qualitatively the feasibility of our proposal, there is room for further research. We plan extending our survey to a wider audience and exploring ways to enhance the generated explanations. In this sense, we will analyze deeper the representation possibilities of GLMP and the overall NLG framework, with the aim of providing users with clearer and more actionable explanations. Moreover, we will study how to enhance our framework with further reasoning capabilities and argument technology.

Acknowledgements

J.M. Alonso is supported by RYC-2016-19802 (Ramón y Cajal contract) and A. Ramos-Soto is funded by the “Consellería de Cultura, Educación e Ordenación Universitaria” (under the Postdoctoral Fellowship accreditation **481B 2017/030**).

References

1. US Public Policy Council, A.: Statement on Algorithmic Transparency and Accountability (2017)
2. Goodman, B., Flaxman, S.: European union regulations on algorithmic decision-making and a “right to explanation”. In: ICML Workshop on Human Interpretability in Machine Learning (WHI), New York, NY (2016) 1–9 <https://arxiv.org/abs/1606.08813>.
3. Wachter, S., Mittelstadt, B., Floridi, L.: Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law* **7**(2) (2017) 76–99
4. Buchanan, B., Shortliffe, E.: Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley (1984)
5. Lombrozo, T.: The structure and function of explanations. *TRENDS in Cognitive Science* **10**(10) (2006) 464–470
6. Tintarev, N., Masthoff, J.: Explaining recommendations: Design and evaluation. In Ricci, F., Rokach, L., Shapira, B., eds.: *Recommender Systems Handbook*. Springer, Boston, MA (2015) 353–382
7. Biran, O., Cotton, C.: Explanation and justification in machine learning: A survey. In: *Proceedings of the IJCAI-17 Workshop on Explainable AI (XAI)*, Melbourne, Australia (2017) 8–13
8. Ho, T.: The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(8) (1998) 832–844
9. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA (1993)
10. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *ACM SIGKDD explorations newsletter* **11**(1) (2009) 10–18
11. Gatt, A., Krahmer, E.: Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* **61** (2018) 65–170
12. Ramos-Soto, A., Bugarin, A., Barro, S.: On the role of linguistic descriptions of data in the building of natural language generation systems. *Fuzzy Sets and Systems* **285** (2016) 31–51
13. Reiter, E., Dale, R.: *Building natural language generation systems*. Cambridge University Press (2000)
14. Alonso, J.M., Ramos-Soto, A., Reiter, E., van Deemter, K.: An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems. In: *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. (2017) 1–6
15. Alonso, J.M., Mencar, C.: Building cognitive cities with explainable artificial intelligent systems. In: *Proceedings of the Workshop on Comprehensibility and Explanation in AI and ML, 16th International Conference of the Italian Association for Artificial Intelligence, Bari, Italy* (2017) 1–10
16. Conde-Clemente, P., Alonso, J.M., Trivino, G.: rLDCP: R package for text generation from data. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, Italy (2017) 1–6
17. Zadeh, L.A.: A new direction in AI: Toward a computational theory of perceptions. *Artificial Intelligent Magazine* **22**(1) (2001) 73–84
18. Castellano, G., Castiello, C., Fanelli, A.M.: The FISDeT software: Application to beer style classification. In: *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, Italy (2017) 1–6