Department: Visualization Viewpoints
Editor: Theresa-Marie Rhyne,
theresamarierhyne@gmail.com

Visual Analytics for human-centered Machine Learning

N. Andrienko
Fraunhofer Institute IAIS (Germany) and City, University of London (UK)

G. Andrienko
Fraunhofer Institute IAIS (Germany) and City, University of London (UK)

L. Adilova
Fraunhofer Institute IAIS and Ruhr University Bochum (Germany)

S. Wrobel
Fraunhofer Institute IAIS and University of Bonn (Germany)

***Abstract*—We introduce a new research area in Visual Analytics (VA) aiming to bridge existing gaps between methods of interactive Machine Learning (ML) and eXplainable Artificial Intelligence (XAI), on one side, and human minds, on the other side. The gaps are, first, a conceptual mismatch between ML/XAI outputs and human mental models and ways of reasoning, second, a mismatch between the information quantity and level of detail and human capabilities to perceive and understand. A grand challenge is to adapt ML and XAI to human goals, concepts, values, and ways of thinking. Complementing the current efforts in XAI towards solving this challenge, VA can contribute by exploiting the potential of visualization as an effective way of communicating information to humans and a strong trigger of human abstractive perception and thinking. We propose a cross-disciplinary research framework and formulate research directions for VA.**

■ **THE IMPORTANCE** of involving humans in the process of creating and training Machine Learning (ML) models is currently widely recognized in the ML community [1]. It is argued that humans involved in this process need to understand what the machine is doing and how it uses human inputs; hence, the machine must be able to explain its behavior to the users. Understanding of ML models has also critical importance for deciding whether they can be adopted for practical use. Explainability of models may even be more important than their performance, especially in high-stake domains. In response to the need to explain untransparent ML models ("black boxes") to users, the research field of eXplainable Artificial Intelligence (XAI) has emerged recently [8]. The work in this field was boosted by the European Parliament's adoption of the General Data Protection Regulation (GDPR), which introduces the right of individuals to receive explanations of automatically made decisions relevant to them.

However, there is a tendency to admit that model “explainability” does not necessarily mean that the model is indeed properly explained to humans and understood by them [9,12]. In this paper, we discuss the deficiencies of the common approaches to explaining ML models, mention current efforts towards overcoming these deficiencies, argue why Visual Analytics (VA) [11] should be involved in such efforts, and consider its possible role in helping humans to understand models.

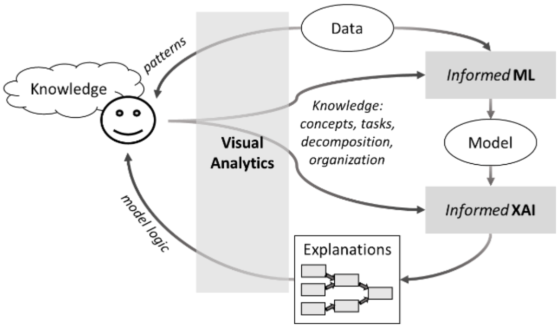


Figure 1. Schematic representation of the research framework for human-centered ML supported by VA.

Based on our considerations, we propose a research framework for developing VA approaches supporting human-centered ML. The basic idea is schematically represented in Fig. 1. Here, the term “informed ML” means involving prior knowledge in the process of deriving models from data, and “informed XAI” means involving knowledge in the process of explaining models to humans. While informed ML uses knowledge that is explicitly represented in a machine-processable form, VA can support acquiring knowledge from a human expert, including expert’s prior knowledge and new knowledge that the expert has obtained through interactive visual data analysis. The knowledge of the expert is externalized through interactive visual interfaces and supplied to the ML and XAI components. Please note that “informed XAI” is a new term that we introduce by analogy with “informed ML”. The contents of Fig. 1 will be explained in more detail later.

We shall begin with providing background information concerning explainability of ML models and deficiencies of common approaches in XAI. After an overview of the relevant research in ML, XAI, and VA, we present the general idea of how VA can contribute to human-centered ML and propose research directions towards realizing this idea.

BACKGROUND

The following definitions and statements are based on a recent survey of the XAI research [8] unless another reference is specified.

In the ML and XAI literature, the terms “explainability” and “interpretability” are used interchangeably. *Interpretability* is defined as the ability to *explain* or to provide the meaning of something in terms *understandable* to a human. The definition assumes implicitly that an explanation is self-contained and does not need further explanations.

An important distinction is made between global and local interpretability. *Global interpretability* means that humans can understand the whole logic of a model and follow the reasoning leading to all possible outcomes. *Local interpretability* means the possibility to understand the reasons for a specific decision.

Among the existing types of ML models, a few are recognized as interpretable and easily understandable for humans, namely, decision tree, rules, and linear (regression) models. A decision tree can be represented graphically, and a human can trace its branches and read logical conditions in the nodes. Rules have the form of logical statements “if … then …”, which are familiar and understandable to humans. Linear models can be interpreted by considering the sign and magnitude of the contribution of each attribute to a prediction.

These model types are considered interpretable by their nature and needing no explanations. The research in XAI is concerned with explaining other types of models that are untransparent to humans. The research addresses three distinct problems. The *black box explanation* problem consists in providing a globally interpretable model which is able to mimic the behavior of the black box. The black box *outcome explanation* problem consists in providing explanations of the reasons for predictions or decisions made by a black box. It is not required to explain the whole logic behind the black box. The *black box inspection* problem consists in providing a representation (visual or textual) for understanding either how the black box model works or why the black box returns certain predictions more likely than others.

The content of this paper partly refers to the first problem, i.e., black box explanation, which is being solved by creating interpretable mimic models. However, the problem we consider here is different. We share the doubts of ML researchers who question the common belief that certain model types are easily understood by humans just because they can be represented in a human-readable form.

DEFICIENCIES OF CURRENT XAI

Some ML researchers argue that the current XAI approaches fail to provide satisfactory explanations that can be well understood by humans, i.e., linked to their mental models. The term "explainability" is contrasted with "explanation" [12] and "causability" [9]. According to Kovalerchuk et al. [12], a model is truly explained if a domain expert accepts it based on both empirical evidence of satisfactory accuracy and the *domain knowledge/theory/reasoning,* which is beyond a given dataset. Instead, XAI methods generate "quasi-explanations", which refer to components and properties of data and specifics of the modelling algorithm but do not explain models in terms of domain knowledge and concepts that humans use in their reasoning. "Causability" [9] is defined as the extent to which an explanation achieves a specified level of *causal understanding*.

The authors of [12] give the following example. Consider a branch of a decision tree or a logical rule "If ($x_1 > 5$) and ($x_2 < 7$) and ($x_3 > 10$) then **x** belongs to class 1". It may be quite accurate in classifying data instances, and a domain expert can understand what it says if attributes x1 to x3 are meaningful in the domain where the data are taken from. However, the domain expert can say that, despite its high empirical confirmation, it is not clear *why this model should work*. The model is not explained in the terms of the domain knowledge such as causal relations known in the domain. This is a quite common situation in ML.

Another example given in [12] refers to linear models, which are also commonly recognized as interpretable. It is typical that linear models involve heterogeneous attributes, such as blood pressure, cholesterol level, temperature, and so on. The weighted summation of such heterogeneous attributes does not have physical meaning. Even when attributes are homogeneous it is still not necessary that the regression models will be meaningful. For instance, what is the meaning of a weighted sum of systolic and diastolic blood pressure measurements?

Additionally, a theoretically interpretable model, similarly to a deep learning model, may involve highly engineered features, such as a cube root of several indicators, which may not have a domain interpretation.

The problem that these examples refer to can be characterized as *conceptual mismatch* between ML/XAI outcomes and human mental models. Another problem, also discussed in [12], is that a model interpretable in theory may be incomprehensible in practice due to its size and complexity. Consider, for example, a decision tree containing hundreds of nodes, as in Fig. 2. A human can trace and understand any small part of it, but the whole tree is beyond the human capabilities for tracing and understanding. Hence, there is a mismatch between the information quantity and the human perceptual and cognitive capabilities.

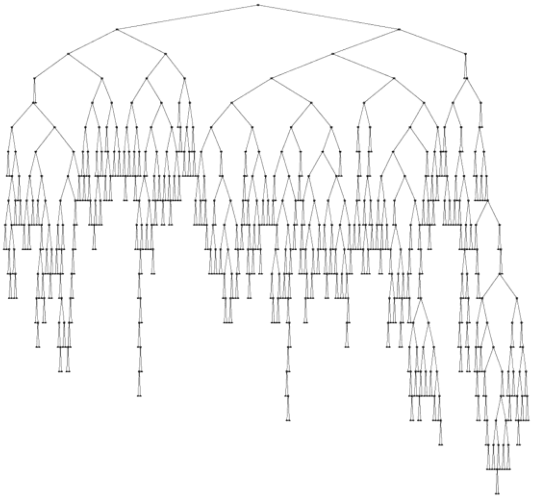


Figure 2. The structure of a decision tree meant to "explain" the logics of an ML model (an example).

There exists research in XAI aimed at making models easier to comprehend. A few representative works are mentioned in the Sidebar 1. However, XAI researchers strive to develop purely algorithmic approaches. VA researchers can complement these efforts by supporting involvement of human knowledge and reasoning.

STATE OF THE ART

Here we briefly overview the state of the art and open problems in ML and VA concerning two sides of human-computer collaboration in development of ML models. One side can be called "Humans for ML": how to make better use of human intellectual capabilities in developing ML models? The other side is "ML for humans": how to ensure that ML results are properly explained to humans in terms of human-relevant concepts rather than machine-specific objects and structures?

Interactive ML

ML acknowledges the value of human feedback in the process of deriving models from data [1]. Most of all, ML researchers are concerned with eliciting training data from human experts [14]. In the ML paradigm known as *active learning*, an algorithm applies some strategy to choose from a pool of unlabeled examples and queries a human "oracle" to provide labels. Apart from practical difficulties in finding suitable strategies [5,6], this approach can cause such problems as human's frustration and unwillingness to repeatedly perform a routine task [1]. Visual-interactive labeling provides users an active role and possibility to apply different strategies [6].

The concept of *interactive machine learning* [10] acknowledges the fact that people may be capable and willing to do much more for development of a good model than just provide data labels. Interactive ML engages human users in a tight interaction loop of iterative modification of data and/or features to improve model performance [2,10]. However, to play such an active role, the users need to understand what the machine is doing and how it uses their inputs. Hence, the machine must be able to *explain* its behavior to the users.

Informed ML

While traditional ML develops methods to derive models purely from data, more and more researchers call for combining data- and knowledge-based approaches, which can reduce the required amount of training data and, at the same time, lead to better model quality, explainability, and trustworthiness. A research field called *informed ML* [16] works on integrating machine learning techniques with processing of conceptual and contextual knowledge. Researchers mostly focus on utilizing knowledge that has been previously prepared and represented in a machine-readable form, such as logic rules, algebraic equations, or concept graphs.

The survey [16] refers to many works on involving knowledge of human experts into the ML pipeline. Expert knowledge may be provided in the form of algebraic equations, probabilistic relations (often represented by Bayesian network structures), or human feedback. The first two forms can be directly used in an ML algorithm. Examples of human feedback are setting preferences, judging relevance, editing algorithm outcomes, and pre-specifying learning targets, such as topics in text documents or data patterns and hierarchies. For obtaining different forms of human feedback, machine learning is increasingly combined with visual analytics [15].

Granular Computing (GC)

Granular computing [15,17] is a paradigm in computer science and applied mathematics that strives to reflect the human ability to perceive the world at different levels of granularity and to switch between these levels. According to [17], there are three basic concepts that underlie human cognition: *granulation*, *organization* and *causation*. Informally, granulation involves decomposition of whole into parts; organization involves integration of parts into whole; and causation involves association of causes with effects. The central concept of GC is an *information granule*, which is a construct composed of data or information items based on their similarity, adjacency, or other relationships. The ultimate objective of information granules is to describe phenomena in an easily understood way and at a certain level of abstraction. Therefore, the ideas of GC align very well with the need of explaining ML models in human-friendly ways [15].

Acknowledging that information granules created and used by humans are fuzzy rather than crisp, the founder of GC L. Zadeh proposed the theory of fuzzy information granulation supported by fuzzy logic [17]. There are also research works in GC applying the theory of rough sets.

Granular computing does not consist of specific methods; it is rather a set of ideas and a way of thinking. The book [15] contains some examples of involving the ideas of GC in building ML models for specific applications. One of the book chapters calls for combining GC with visual analytics.

Visual Analytics (VA)

VA is a natural partner of ML and AI in the research both on involving users in ML processes and on explaining ML to users. Combining human and machine intelligence is the central idea of VA [2]. Sidebar 2 points out the research areas in VA related to ML and refers to representative works.

Most of the research dealing with ML models has been related so far to different aspects of the problem "humans for ML". The area of VA for XAI can be, in principle, categorized as "ML for humans", but the current research in it addresses mostly the needs of model developers rather than domain experts. The visualization of classification rules in RuleMatrix [13] is meant for users with little competence in ML; however, the authors do not consider the problem of comprehensibility of large rule sets with rules involving many conditions.

A series of VISxAI workshops (Visualization for AI Explainability, http://visxai.io/) promotes the creation of interactive visual "explainables" or "explorables" explaining how ML/AI techniques work using visualization. AI explorables are also being created and published by the Google team PAIR (People + AI Research, https://pair.withgoogle.com/explorables/). There are many interesting works allowing users to experiment with models by changing parameters or supplying different inputs. Such experiments, however, do not explain the internal logic of the models. Other works focus on explaining ML concepts and methods rather than models created for specific applications. Both groups of work are more oriented to students and curious public than to domain experts going to use the models in practice.

It can be seen that different research communities are concerned with making ML models understood by users. These communities focus on different aspects of the model explanation problem, such as model complexity, form of representation, level of abstraction, and “what-if” explorability. It seems, however, that satisfactory solutions can only be achieved when the communities join their efforts in tackling the problem. Therefore, we propose an interdisciplinary research framework for human-centered ML.

RESEARCH FRAMEWORK

The idea of the proposed research framework is schematically represented in Fig. 1. It is interpreted as follows. Following the paradigms of interactive ML and informed ML, models are developed in tight interaction of ML algorithms with humans, so that human knowledge and human-defined concepts are transferred to the algorithms and used in building computer models. This process is supported by interactive visual interfaces provided by VA. The knowledge and concepts that have been acquired from the human experts are involved not only in model building but also in generating explanations of the models. The methods for doing this, which still need to be developed, can be called “informed XAI”, by analogy with informed ML. It can be expected that such methods will soon be developed in the XAI area. When they appear, it will be the task of VA to represent their outcomes to users. VA researchers should also think about possible visual and interactive ways of organizing outputs of current XAI methods based on human knowledge.

This research framework refers simultaneously to both perspectives of human-computer collaboration in the creation of computer models, i.e., “humans for ML” and “ML for humans”. These two perspectives are united through the involvement of human expert knowledge. The role of VA is to support acquisition of knowledge from experts and use of the expert knowledge in providing model explanations to the end users.

The research on human-centered ML can built on the achievements and current developments in the areas of interactive ML, informed ML, XAI, and VA. Since VA is interdisciplinary by its nature, it will be the task of VA researchers to **design and develop integrated VA-ML-XAI workflows** implementing the conceptual view of visual analytics activity as the process of model building [4].

Integrated VA-ML-XAI workflows

Two complementary directions for integration can be envisaged. The *first direction* involves applying VA to the data that will be used for model building. The idea is that VA supports the human analyst in organizing the data and defining meaningful concepts at an appropriate level of abstraction. There is a special ML component that learns the concepts and the ways of organizing data items into instances of these concepts. The knowledge thus gained from the human is then used in an ML algorithm that derives a model from the data, which means that the algorithm is designed to utilize this expert knowledge for directing the data-driven learning process, according to the ideas and approaches of informed ML. Additionally to this, the knowledge is used by an XAI component, which generates and organizes explanations according to the human-defined concepts thereby implementing the idea of “informed XAI”.

There exist multiple VA solutions for supporting transfer of knowledge from humans to ML algorithms, e.g., [7]. However, we are not aware of works implementing the next step, in which the knowledge obtained is used for generation of human-oriented explanations.

The *second direction* is interplay of VA and XAI components. The XAI component initially generates detailed low-level explanations. The human analyst uses VA techniques to organize subsets of these explanations into meaningful information granules, in terms of granular computing, and thereby define relevant concepts at suitable levels of abstraction. The

XAI component learns the concepts and the ways of organizing explanations from the analyst and applies this knowledge to other subsets of “raw” explanations under the expert’s supervision. Being trained in this way, the XAI component will later be able to use the learned principles of structuring and abstracting in explaining other ML models of the same type (e.g., classification or regression) in the same domain. This is another way of implementing the idea of “informed XAI”. VA techniques are used to present the resulting explanations to users in effective ways.

To create a theoretical basis underpinning these practical developments, we propose to work on combining the ideas and frameworks of visual analytics and granular computing.

Theoretical research

GC aims to model the human ability to organize and perceive information at different levels of abstraction. VA, in turn, is concerned with supporting abstractive perception of data and information from visual displays. The central concept in VA is a *pattern*, which is a combination of multiple items perceived and considered together as a single entity due to relationships existing between the items [3]. Patterns themselves may also be linked by relationships and on this basis integrated into patterns of a higher level of abstraction.

There is a semantic similarity between the concepts of information granule in GC and pattern in VA. The ultimate goal of VA is similar to that of GC: enable humans to understand phenomena at appropriate levels of abstraction. Therefore, it appears reasonable to link these two research fields. It needs to be investigated what theories and methods of GC can be integrated with techniques of VA, how different types of information granules can be represented visually, and how these types of granules can be formed through human-computer discourse using visual and interactive techniques.

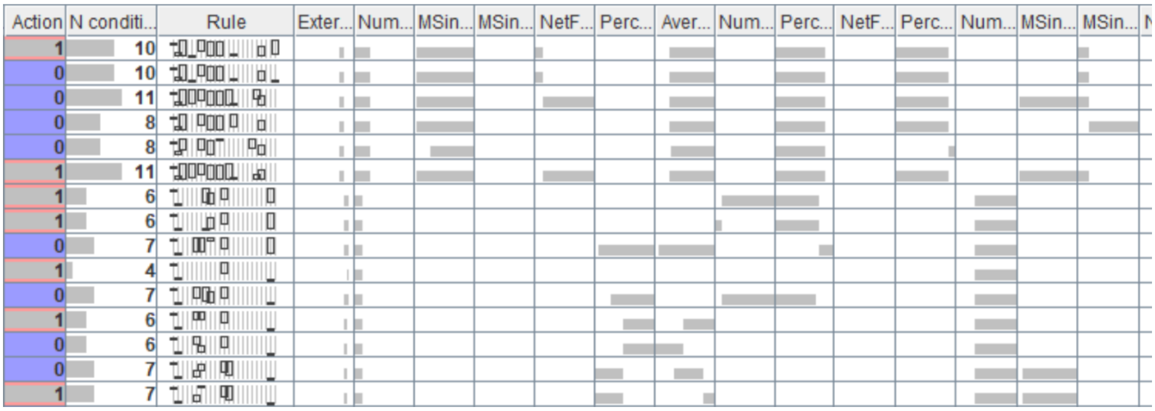
Particularly, GC is concerned with modelling the approximate, fuzzy way of human conceptualization and reasoning. As mathematical apparatuses for this, GC proposes to use fuzzy sets and rough sets theories. These formalisms appear suitable for representing data patterns, such as a cluster or a trend, which usually have an approximate character.

Based on the definition of a data pattern as a system of type-specific relationships between data items [3], it may be possible to generate formal representations of data patterns discovered in the process of visual analysis and roughly outlined or otherwise marked by the analysts. These formal representations can be processed by computers and used in model building. To find suitable ways of representing data patterns, it is also reasonable to consult the literature on knowledge representation in the classical AI.

To provide theoretical foundation to organization and abstraction of low-level XAI outputs, it is necessary to elaborate the pattern theory in more detail for defining possible patterns in such a complex type of information as XAI-generated explanations, e.g., having the form of decision rules or trees. In the next section, we describe some preliminary ideas concerning patterns in a set of rules and possibilities for uniting and generalizing related rules. Please note that these ideas and examples refer to the second direction in the work on implementing integrated VA-ML-XAI workflows.

EXAMPLE: GRANULATION OF RULES

Let us consider decision rules with conditions involving numeric attributes (features). Such rules may be components of an original ML model or of a mimic model constructed by some XAI method to explain a black box model. Each condition of a rule refers to one feature and states that the feature value must be lower or higher than a certain constant, or that it must be within a certain interval. A rule usually contains several conditions connected by the logical operator AND. The outcome, or consequent, of a rule is one element from a finite set of possible classes, decisions, or actions.



Action	N conditi...	Rule	Exter...	Num...	MSin...	MSin...	NetF...	Perc...	Aver...	Num...	Perc...	NetF...	Perc...	Num...	MSin...	MSin...	N
1	10																
0	10																
0	11																
0	8																
0	8																
1	11																
1	6																
1	6																
0	7																
1	4																
0	7																
1	6																
0	6																
0	7																
1	7																

Figure 3. A fragment of a table representing rules.

To see rules in a visual form, we can use a table view like the one shown in Fig. 3. Each table row corresponds to one rule. For each feature, there is a column. Conditions, i.e., intervals of feature values, are represented by horizontal bars, which show the relative positions of the intervals between the minimal and maximal feature values. If a feature is not used in a rule, the corresponding cell is empty.

A single rule can also be represented by a glyph, as shown in Fig. 4. Based on the idea of parallel coordinate axes, a glyph includes vertical axes corresponding to all features occurring in a rule set. Vertical bars represent the value intervals of the features used in the rule. The color of the glyph frame encodes the rule outcome.

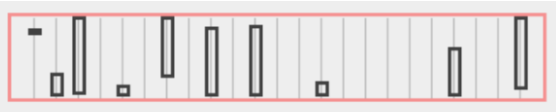


Figure 4. One rule represented by a glyph.

According to the definition of a pattern [3], patterns in a set of rules emerge due to relationships between rules. Relationships between rules are composed from relationships between their conditions and between the outcomes. For the outcomes, two relationships are possible: same or distinct. Relationships between two conditions involving the same feature are relationships between the value intervals specified in the conditions. The intervals can be disjoint, overlapping, coinciding, or one can lie inside the other. Relationships between intervals can be expressed numerically as distances between them. For this purpose, we can use an adapted version of the Hausdorff distance between two subsets of a metric space.

Figure 5 demonstrates a possible visual representation of relationships between rule conditions. Three rules are represented by glyphs. The outcome of the first rule differs from the outcomes of the two others. The first rule is selected. Its conditions are represented in all three glyphs by bars shaded in light blue and drawn on the right of the corresponding feature axes. The relative positions of the framed hollow bars and the shaded bars represent the relationships between the feature value intervals in the conditions of the selected rule and in the other rules.

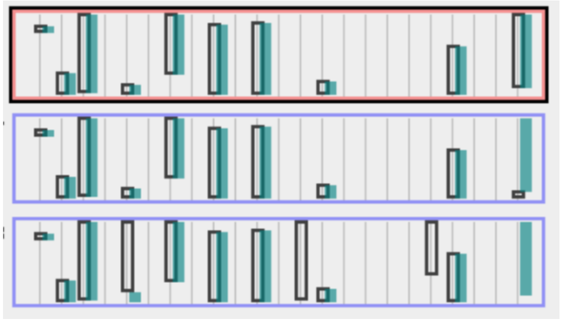


Figure 5. Representation of relationships between rule conditions.

To understand the possible relationships between rule antecedents composed of multiple conditions, let us imagine the multidimensional space of all features involved in all rules. The antecedent of a rule can be imagined as a shape (a hyper-parallelepiped) in this space. When some feature is not used in a rule explicitly, it can be treated as implicitly present with the value interval covering the whole range of feature values from the smallest to the largest. In such a view, relationships between rule antecedents translate to relationships between such multidimensional shapes. Possible types of relationships are set relationships (disjoint, intersect, include, coincide) and metric distance relationships between the shapes. As a numerical expression of these distances, we can use, for example, the mean of the distances in all individual dimensions. This numeric measure of rule similarity can be used to algorithmically find groups (clusters) of close rules, as well as for ordering of rules. Thus, adjacent table rows in Fig. 3 correspond to close rules.

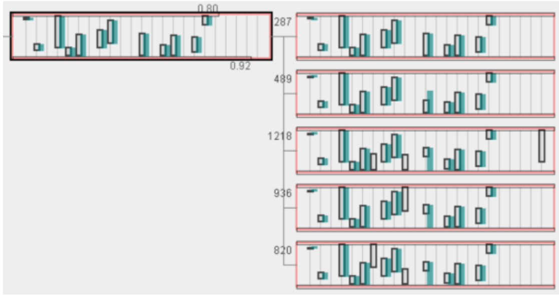


Figure 6. A union of a group of close rules.

An important rule pattern is a cluster of close rules having the same outcome. Such a cluster can be abstracted into a multidimensional shape enclosing it. This envelope shape, in turn, corresponds to a rule that is more general than each member rule of the cluster; we shall call it a *union rule*. We say that the union rule *covers* each original rule of the cluster that has been abstracted. In terms of rule conditions, it means that each interval of feature values of the union rule covers (i.e., coincides with or includes) the value intervals of the same feature of all original rules. Hence, a union rule can be derived from a group of rules by obtaining the unions of the value intervals of the same features. When a union of two or more intervals equals the full range of the feature values, the condition referring to this feature can be omitted from the union rule. Fig. 6 shows an example of a union rule abstracting a cluster of five close rules.

In terms of granular computing, a union rule is an information granule. Union rules can be derived by iterative joining of pairs of close rules. This creates rule hierarchies involving information granules of different degrees of abstraction.

A union rule covering a cluster of close rules with the same outcome may occasionally also cover some other rules with different outcomes. This is similar to enclosing a cluster of points of the same class on a scatterplot by a bounding box: some points of another class may also fit into the box. Hence, a union rule can be an approximate, rough representation of a cluster of similar rules. We shall use the term *rough rule* for a rule covering two or more rules with the same outcome as in this rule and at least one rule with a different outcome. The accuracy of a rule can be numerically expressed as the ratio of the number of covered rules with the same outcome as in this rule to the total number of the rules covered by this rule. The accuracy of a rough rule will thus be less than 1.

Obviously, a rough union rule is less suitable for making predictions than the original group of rules that has been abstracted. However, it may be quite well suitable for explanation of the model logic to a human, since it is normal for human cognition to deal with rough concepts and approximations. A user of an ML model can agree to accept some inaccuracies in exchange for a simper description of the model logic, and the user can choose the minimal accuracy that is still acceptable. Hence, by finding and abstracting clusters of rules with same outcomes, we aim to derive a simpler model that is *descriptive* but not necessarily predictive.

We have conducted multiple experiments on granulation of different ML models consisting of rules or decision trees (a decision tree can be transformed to a set of rules by representing each path from the root to a leaf by one rule). The models were created based on several benchmark datasets using state-of-the-art ML methods. Our goal was to find out how much a model can be simplified by means of rule granulation. We varied the minimal accuracy threshold from 1 to 0.6. Interestingly, even with the threshold equal to 1 some compression is achieved. For example, a 3-class classification model with 109 rules and 818 conditions in total has been reduced to 103 rules with 762 conditions. With the threshold of 0.75 for the same model, we obtained 84 rules with 594 conditions, and the threshold 0.6 gave us 54 rules with 342 conditions. A model with 10 classes containing 202 rules (1739 conditions) was abstracted to 167 rules (1357 conditions) taking the threshold 0.75 and to 139 rules (1062 conditions) taking the threshold 0.6. Similar degrees of compression were achieved in the other experiments.

Based on our experiments, we can conclude that rule granulation is a viable approach to simplification of rule sets. However, its power is limited: the simplified models still contain too many rules and conditions to be treated as easily comprehendible. The reason for this inadequacy is that abstracted rules involve the same low-level features taken from training data as the original rules. A model can be better understood by a domain expert if it refers to higher-level domain-relevant concepts. Such concepts cannot be automatically derived from data but need to be taken from other sources, as it is supposed in the paradigm of informed machine learning [16]. One of the possible sources may be a human expert interacting with a model building algorithm, as shown schematically in Fig. 1. The expert may define concepts based on groups of features, which can be seen as “feature granulation”.

As a simple example of feature granulation, let us imagine creation of a model for diagnosing various allergies. Elementary features may be symptoms like sneezing, runny nose, blocked nose, red eyes, itchy eyes, watery eyes, itchy skin, red rash, and many others. A domain experts may organize the symptoms in groups, such as nasal symptoms, eye symptoms, skin symptoms, etc., and tell the learning algorithm which groups of symptoms are related to respiratory allergies, skin allergies, food allergies, and so on. When the groups of symptoms and groups of allergies defined by the expert are involved in the model or at least used in generating explanations of the model, it can be expected that the explanations will be more structured, more meaningful for domain users, and better understood by them.

CONCLUSION

With this paper, we aim to motivate and trigger research on bridging gaps between machine learning and human mental models using a synergy of approaches from informed machine learning, artificial intelligence, and visual analytics. While substantial amount of research is being conducted in several areas of computer science, the contribution from visual analytics is still low. We believe that VA researchers should take a lead in these efforts, since the goal of combining human and computer intelligence lies at the core of VA. Interactive visual interfaces serve as means of human-computer communication and as facilitators of human abstractive perception of information and derivation of new knowledge, which refines and enriches human mental models [4]. Since human knowledge plays the key role in the proposed framework (Fig. 1), visual analytics researchers are supposed to care about capturing this knowledge and transferring it to computers.

We have outlined several lines of research in VA that fit in the proposed research framework. These include theoretical developments, such as models and methods of information granulation and transformation of data patterns into knowledge structures, and practice-oriented design of workflows involving cross-disciplinary approaches. Progress in these directions will result in methods and systems for building models enhanced by the power of human intelligence and readily accepted by humans as extensions of their mental models and enhancers of their reasoning.

ACKNOWLEDGEMENTS

This research was supported by Fraunhofer Center for Machine Learning within the Fraunhofer Cluster for Cognitive Internet Technologies, by DFG within Priority Programme 1894 (SPP VGI), by EU in project SoBigData++, and by SESAR in projects TAPAS and SIMBAD.

■ REFERENCES

1. S. Amershi, M. Cakmak, W.B. Knox, & T. Kulesza. “Power to the People: The Role of Humans in Interactive Machine Learning”. AI Magazine, vol. 35, no. 4, pp. 105-120. 2014.
2. N. Andrienko, G. Andrienko, G. Fuchs, A. Slingsby, C. Turkay, and S. Wrobel. “Visual Analytics for Data Scientists”. Springer, 2020.
3. N. Andrienko, G. Andrienko, S. Miksch, H. Schumann, & S. Wrobel. “A theoretical model for pattern discovery in visual analytics”. Visual Informatics, vol. 5, no. 1, pp. 23-42, 2021.
4. N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. “Viewing visual analytics as model building”. Computer Graphics Forum, vol. 37, no. 6, pp. 275–299, 2018.
5. J. Attenberg and F. Provost, “Inactive learning? difficulties employing active learning in practice”. ACM SIGKDD Explorations Newsletter, vol. 12, no. 2 (December 2010), 36–41, 2011.
6. J. Bernard, M. Zeppelzauer, M, Lehmann, M. Müller, and M. Sedlmair, “Towards User-Centered Active Learning Algorithms”. Computer Graphics Forum, vol. 37, pp. 121-132, 2018.
7. S. van den Elzen and J. J. van Wijk, "BaobabView: Interactive construction and analysis of decision trees," 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 151-160, 2011.
8. R. Guidotti, A. Monreale, F. Turini, D. Pedreschi, and F. Giannotti. “A Survey of Methods for Explaining Black Box Models”. ACM Computing Surveys, vol. 51. 2018.
9. A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller. “Causability and explainability of artificial intelligence in medicine”. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 4: e1312, Jul-Aug. 2019.
10. L. Jiang, S. Liu, and C. Chen. “Recent research advances on interactive machine learning”. Journal of Visualization, vol. 22, no. 2, pp. 401–417, 2019.
11. D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. “Visual Analytics: Definition, Process, and Challenges”. In: A. Kerren, J.T. Stasko, J.-D. Fekete, C. North (eds) Information Visualization. Lecture Notes in Computer Science, vol 4950. Springer, Berlin, Heidelberg. 2008.
12. B. Kovalerchuk, M.A. Ahmad, and A. Teredesai. “Survey of Explainable Machine Learning with Visual and Granular Methods Beyond Quasi-Explanations”.

In: Pedrycz W., Chen SM. (eds) “Interpretable Artificial Intelligence: A Perspective of Granular Computing”. Studies in Computational Intelligence, vol. 937, pp. 217-267. Springer, 2021.

13. Y. Ming, H. Qu, and E. Bertini. “RuleMatrix: Visualizing and Understanding Classifiers with Rules”. IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1 (Jan. 2019), 342–352, 2019.
14. R. Monarch. “Human-in-the-Loop Machine Learning. Active learning and annotation for human-centered AI”. Manning Publications, 2021.
15. W. Pedrycz, and S.-M. Chen (eds.) “Interpretable Artificial Intelligence: A Perspective of Granular Computing”. Springer, 2021.
16. L. von Rueden et al. Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems, in IEEE Transactions on Knowledge and Data Engineering, 2021. doi: 10.1109/TKDE.2021.3079836.
17. L.A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic”. Fuzzy Sets and Systems, vol. 90, pp. 111-127, 1997.

Natalia Andrienko, is a lead scientist at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Results of her research have been published in two monographs, ”Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach” (2006) and ”Visual Analytics of Movement” (2013). Natalia Andrienko has been an associate editor of IEEE Transactions on Visualization and Computer Graphics (2016-2020) and is now an associate editor of Visual Informatics.

Gennady Andrienko, is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Gennady Andrienko was a paper chair of IEEE VAST conference (2015–2016) and associate editor of IEEE Transactions on Visualization and Computer Graphics (2012–2016), Information Visualization and International Journal of Cartography.

Linara Adilova, is a PhD student in Ruhr University Bochum and a research scientist at Fraunhofer IAIS. She has been working and publishing on multiple research directions, e.g., distributed (federated) learning, autonomous driving. Her main research focus lies in theory and mathematics of deep learning.

Stefan Wrobel, is Professor of Computer Science at University of Bonn and Director of the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS. His work is focused on questions of the digital revolution, in particular intelligent algorithms and systems for the largescale analysis of data and the influence of Big Data/Smart Data on the use of information in companies and society. He is the author of a large number of publications on data mining and machine learning, is on the Editorial Board of several leading academic journals in his field and is an elected founding member of the “International Machine Learning Society”.

SIDEBAR 1: XAI EFFORTS FOR IMPROVING MODEL COMPREHENSIBILITY

A so-called “user-centric XAI framework” [3] aims to link XAI approaches to theories describing human reasoning and decision making, which have been developed in psychology and philosophy. The framework is intended to inform XAI researchers about human cognitive patterns that should be taken into account in designing XAI methods. The authors care most of all about the use of XAI for mitigation of human cognitive biases and improvement of human reasoning and decision making rather than about the improvement of XAI itself.

There exist research works on structuring and abstracting information for increasing model comprehensibility. One example is an approach to identifying the contribution of groups of features to the predictive accuracy of a model [1]. It uses a predefined hierarchy of features and tries to ascertain the level of resolution at which the importance of the features and feature groups can be determined. Another example is integration of multiple decision tree models into a more general model [2]. The proposed approaches are purely algorithmic. VA researchers can complement these efforts by supporting involvement of human knowledge and reasoning.

REFERENCES

1. K. Lee, A. Sood, and M. Craven, “Understanding Learned Models by Identifying Important Features at the Right Resolution”, AAAI, vol. 33, no. 01, pp. 4155-4163, Jul. 2019.
2. P. Strecht, J. Mendes-Moreira, and C. Soares. “Inmplode: A framework to interpret multiple related rule-based models”, Expert Systems. 2021
3. D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing Theory-Driven User-Centric Explainable AI. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Paper 601, 1–15. 2019.

SIDEBAR 2: VA RESEARCH ON COMBINING VA AND ML

There are several research areas in VA related to ML:

- ML in VA: incorporation of ML methods in VA systems and workflows to complement human reasoning and advance data analysis [1].
- Predictive VA: synergistic use of ML and VA techniques for development of predictive models [4].
- VA-assisted ML: leveraging VA techniques in ML workflows [5,7].
- VA of ML models: VA support to model inspection, i.e., the process of understanding, diagnosing, and refining an ML model [2,3].
- VA for XAI: interactive visual interfaces to XAI methods [6].

REFERENCES

1. A. Endert, W. Ribarsky, C.Turkay, B.L.W. Wong, I.T. Nabney, I. Diaz-Blanco, and F. Rossi, “The State of the Art in Integrating Machine Learning into Visual Analytics”. Computer Graphics Forum, vol. 36, no. 8, pp. 458-486, 2017.
2. F. Hohman, M. Kahng, R. Pienta and D. H. Chau, "Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers," IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 8, pp. 2674-2693, 1 Aug. 2019.
3. S. Liu, X. Wang, M. Liu, and J. Zhu. “Towards better analysis of machine learning models: A visual analytics perspective”, Visual Informatics, vol. 1, no. 1, pp. 48-56, 2017
4. Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski. “The state-of-the-art in predictive visual analytics”. Computer Graphics Forum, vol. 36, no. 3, pp. 539–562, 2017.
5. D. Sacha, M. Kraus, D.A. Keim, and M. Chen. “VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning,” in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 385-395, 2019.
6. T. Spinner, U. Schlegel, H. Schäfer, and M. El-Assady. "explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 1064-1074, 2020.
7. J. Yuan, C. Chen, W. Yang, M. Liu, J. Xia, and S. Liu. “A survey of visual analytics techniques for machine learning”. Computational Visual Media, vol. 7, pp. 3–36, 2021.