



VIAL: a unified process for visual interactive labeling

Jürgen Bernard^{1,4} · Matthias Zeppelzauer² · Michael Sedlmair³ · Wolfgang Aigner^{2,5}

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

The assignment of labels to data instances is a fundamental prerequisite for many machine learning tasks. Moreover, labeling is a frequently applied process in visual interactive analysis approaches and visual analytics. However, the strategies for creating labels usually differ between these two fields. This raises the question whether synergies between the different approaches can be attained. In this paper, we study the process of labeling data instances with the user in the loop, from both the machine learning and visual interactive perspective. Based on a review of differences and commonalities, we propose the “visual interactive labeling” (VIAL) process that unifies both approaches. We describe the six major steps of the process and discuss their specific challenges. Additionally, we present two heterogeneous usage scenarios from the novel VIAL perspective, one on metric distance learning and one on object detection in videos. Finally, we discuss general challenges to VIAL and point out necessary work for the realization of future VIAL approaches.

Keywords Information visualization · Visual analytics · Machine learning · Labeling · Active learning · Classification · Regression · Similarity search · Visual interactive labeling · Labeling

1 Introduction

A central topic in data science is the understanding of data and the discovery of knowledge from data. Research has addressed this issue from different perspectives. On the one hand, machine learning (ML) provides a rich tool set for the automatic indexing, organization, and categorization of huge amounts of data. On the other hand, visualization (VIS) aims at the organization and presentation of data as well as knowledge discovery in a visual interactive way. While both disciplines have their respective strengths for data analysis, they have an even stronger potential when they are combined in visual analytics (VA) approaches [32,84]. Still, the complementary strengths are often not fully exploited.

Building upon approaches investigating combinations of ML with VIS in general, this work explicitly addresses the

common goal of *labeling* tasks. We refer to labeling as the assignment of labels y to given input instances x (i.e., objects, elements, or samples). Labels can be used to find functions f that either map instances to labels, i.e., $f(x) = y$, or define relations between instances, i.e., $f(x_1, x_2) = y$. Labels in this context can be of different types, such as categorical labels [39] in classification tasks, numerical labels [17] in regression tasks, relevance scores [82] in retrieval tasks, as well as labels that represent a relation between two instances (e.g., for learning similarity and distance measures between objects) [15].

A fundamental difference between ML and VIS approaches is the way this goal is achieved. ML most often operates fully automatically, i.e., instances with previously defined labels are fed into a supervised ML algorithm which in turn learns the function f from the data. Once trained, the algorithm can be applied for the labeling of new data. As such ML methods are predominantly model-centric. In turn, the VIS perspective emphasizes the knowledge generation of the user, e.g., by visual interactive labeling interfaces. As such, the VIS perspective is most often user-centric.

In the presence of unlabeled data, incremental ML methods provide iterative solutions to learning ML algorithms, and to refine them continuously [47]. Active learning (AL) [71] is a special type of incremental supervised ML

✉ Jürgen Bernard
juergen.bernard@gris.tu-darmstadt.de

¹ TU Darmstadt, Darmstadt, Germany

² St. Pölten University of Applied Sciences, St. Pölten, Austria

³ Jacobs University, Bremen, Germany

⁴ IGD, Fraunhofer Institute for Computer Graphics Research, Darmstadt, Germany

⁵ TU Wien, Vienna, Austria

where the user is integrated into the learning process to guide the training. As such, a connecting factor between AL and VIS is the user-in-the-loop principle. In AL an algorithmic model proactively asks the user for feedback about unlabeled data to improve a learning model. Therefore, the user is queried for labels of instances the learner is unsure about, which will potentially improve the quality of the model. Typical outputs of the AL process are datasets enriched with labels as well as models learned in the process. Compared to VIS, however, in AL the active role of the user remains rather marginal. In most cases, AL neither incorporates the user's abilities to *select* meaningful instances (e.g., reflecting visual patterns in data), nor does the AL process foster the user's knowledge generation process.

In the VIS community labeling is an important task as well. Many approaches accept feedback from users for data instances of interest as input to learn and further support the users' information need. Important tasks supported by visual interactive interfaces are the analysis of model results, the identification and selection of instances, as well as labeling per se. Example labeling interfaces accept numerical interestingness scores to train regression models [17] or user-defined class labels to train classifiers [39]. More complex labeling techniques allow, for example, the manipulation of spatial proximity of data instances [12] to make statements about (similarity) relations between instances. In contrast to ML, VIS approaches seem to prefer user-centered over model-based criteria. Potential drawbacks of VIS approaches relate to (a) the usefulness of purely user-selected instances for labeling to build accurate, robust, and generalizable algorithmic models, and to (b) the emphasis on knowledge generation while the enrichment of data with labels is often neglected.

We assume that the model-centered AL and the user-centered VIS perspectives have complementary and unexploited synergies for labeling tasks. Building upon and extending notions of “interactive learning” presented in pioneer approaches combining AL and VIS [39,74], we investigate the strengths of both and propose an abstracted and unified process in a visual analytics (VA) context that we refer to as *visual interactive labeling* (VIAL). Our line of approach complies with established process models in visualization and VA [21,22,41,96], resembling the abstract data and interaction flow, as well as user-based knowledge generation [83]. While these models offer a high degree of abstraction, we extend and substantiate these general process toward labeling tasks. Process models and surveys in AL exist as well [71]. However, these models often fall short in visual interfaces as well as knowledge generation support [83].

Most related approaches either indicate the combination of model-based and user-based labeling [39,84] or propose methodologies or concepts for the interactive propagation of user feedback for labeling tasks [18,52]. Although these approaches are inspiring, they are specific toward a data type,

employed technique, application goal, user group, or target variable y . In contrast, the rationale of our unified VIAL process is to abstract from concrete approaches and to propose a general and conceptual labeling workflow. Furthermore, one aspect of the labeling process remains largely uncharted—the three types of output: labeled data, trained models, and gained knowledge. VIAL, on the contrary, obtains a data-, model-, and user-centric perspective with three outputs: data, models, and knowledge.

In this work, we present a conceptual cross-disciplinary process that combines the AL and the VIS perspective. We explain the six crucial steps of the VIAL process, point out their interplay, and describe how AL and VIS can contribute to the respective step. In addition, we discuss the major design and development challenges in every step from both the AL and the VIS perspective. Future approaches may benefit from the VIAL process in two ways. First, we provide an integrated view of AL and VIS in a VA setting that may inspire novel innovative approaches that go beyond the borders of the individual disciplines. Second, the outlined challenges help to overcome inherent hurdles in the VIAL process and to make informed design decisions.

We present related work in Sect. 2. In Sect. 3, we introduce the VIAL process, followed by two usage scenarios in Sect. 4 building upon the VIAL process. We discuss limitations and potential future work in Sect. 5 and conclude with Sect. 6.

This article is an extension of a previous workshop paper on the same topic [20] and provides a theoretical framework for our empirical studies in this area [8]. The article at hand has been fundamentally reshaped in many ways including a substantial extension of related work, the integration of usage scenarios to better illustrate the application of the model, the extension, and more in-depth descriptions of the model's steps including research challenges and possible solutions. Moreover, it includes more details in most sections of the paper, particularly the discussion, limitations, and future work.

2 Background and related work

In the ML and VIS literature, we find different and partly complementary approaches to data labeling. A major contribution of the VIAL process is to join both strategies to develop a broader and more general perspective. In the following, we describe the labeling process in ML (AL) and VIS and point out differences.

2.1 Labeling in machine learning

Labeling of datasets is an important prerequisite for the training of supervised ML models. Recent developments have shown that labeled datasets are not only important

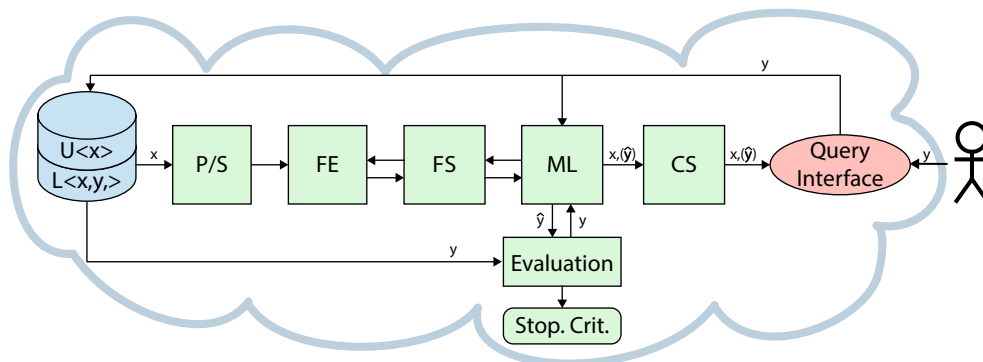


Fig. 1 The abstracted AL process. A data source contains unlabeled (U) and labeled (L) data. Data are first preprocessed (e.g., normalized or segmented, “P/S”). Next, feature extraction “FE” and feature selection “FS” are applied to obtain a useful data representation for ML. A learning model “ML” is trained and evaluated iteratively. Candidate

suggestion “CS” strategies query new labels y from the oracle which are used to iteratively adapt the ML until a stopping criterion is met [57]. Users provide labels but are not involved in CS. In addition, knowledge generation by users is not part of the model

to enable the objective comparison of ML approaches but are further necessary to successfully train today’s complex classifiers, such as deep neural networks (DNNs) [46]. To reduce the effort of labeling large datasets (of potentially millions of instances) different strategies have been developed (e.g., captcha-based information collection [106], web-based annotation systems [62], paid micro-tasks on platforms like mechanical turk [9], and game-based approaches [61,89]).

Pre-labeling In supervised ML the dataset is usually generated and annotated completely in an offline process before classifier training starts. To differentiate this type of labeling process from other more interactive and online labeling processes, we refer to this process as *pre-labeling*. Pre-labeling is an expensive and time-consuming process that should—in the best case—be performed by domain experts. In the absence of a sufficiently large number of experts, webcrawling represents a relatively inexpensive and promising alternative. This however may introduce problems such as inconsistent labelings across several annotators, unreliable and wrong labels. Thus, special filtering methods need to be applied to obtain robust labels. A drawback of pre-labeling is that once annotated labels cannot be refined during the learning process, e.g., according to the interpretation of the data by a user.

Incremental learning and active learning An alternative to pre-labeling is the use of incremental and online learning methods that enable the continuous refinement of a previously learned model as new labeled data become available [47]. AL is special type of incremental ML that explicitly incorporates user knowledge into the learning process [71]. In AL, an algorithmic model proactively asks the user (referred to as the oracle) for feedback during the learning process

(usually in the form of labels) to iteratively improve the learned model. Since user interactions are time-consuming and thus expensive, AL aims at minimizing the amount of required user interactions by querying only that information that will improve the accuracy of the given model in a best possible way. The general AL workflow is illustrated in Fig. 1. To find the potentially most useful unlabeled instances for learning, different strategies for candidate selection have been introduced which are discussed in detail in a number of surveys on AL [57,71,88,98]. These strategies can be partitioned into five groups: (1) uncertainty sampling, (2) query by committee, (3) error reduction schemes, (4) relevance-based selection, and (5) purely data-centered strategies.

Active learning strategies *Uncertainty sampling* aims at finding the instances that the learner is most uncertain or unsure about. Widely used strategies search for those instances near decision boundaries of margin-based classifiers [99] (large-margin-based AL) [88] or for instances with high entropy of class probabilities [69]. *Query by Committee (QBC)* [79] strategies measure the uncertainty of a committee of classifiers. Instances are considered interesting when the committee disagrees with respect to their labeling [50]. *Error reduction* strategies select those instances which may change the underlying classification model most. Selection criteria are expected model change [70], risk reduction [59], or variance reduction [37]. *Relevance-based* strategies [94] focus on instances which have the highest probability to be relevant for a certain class. This strategy fosters the identification of positive examples for a class which are, for example, useful in the context of ranked search results [98]. *Data-driven* strategies work independent of the learning model. Many techniques work with a density-based selection criterion which is a promising strategy for initiating an AL process

in the case when no labels are available at all (cold start problem) [1]. In contrast, diversity-based criteria foster the selection of dissimilar instances for labeling to increase the information gain for the learner [27].

Benefits and limitations of active learning Incremental learning methods like AL are especially useful in cases where large portions of the data are unlabeled, where manual labeling is expensive, and in cases of online learning where new unlabeled data need to be processed continuously. One drawback of AL is that strategies are primarily driven by the learned model—users are not considered in the *identification and selection* of instances, but only in the labeling itself. Hence, the selection of instances is neither based on expert knowledge, nor humans’ ability to quickly identify patterns such as clusters and outliers. This leads to two limitations. First, the integration of the user needs, intentions, and tasks is impeded, and second, the AL process does hardly support the generation of user knowledge. The VIAL process is more general and extends AL toward the user-based candidate selection by the integration VIS techniques [39,74].

2.2 Labeling in interactive visualization

In VIS, many labeling approaches assign users an active role, using visual interfaces as the means for candidate *identification and selection*. We assess a strong emphasis on knowledge generation. However, in contrast to AL, VIS has gained little attention to the creation of labeled data. We structure related works in process models and techniques for visual interactive labeling.

Process models for interactive visualization We draw connections to both general as well as task-specific process models. The VIAL process is derived from general process models for VIS [22,96] and VA [21,41], resembling the abstract data and interaction flow, as well as user-based knowledge generation [83]. However, abstract frameworks often do not reflect the characteristics of specific tasks [84] such as labeling. The VIAL process extends these models for labeling tasks. In particular, with the VIAL process the output of the process is refined considerably. With *user knowledge*, *labeled data*, and *trained models*, the VIAL process explicitly defines three types of output.

Recently, more specific process models have been proposed for different labeling tasks. Höferlin et al. [39] explicitly focus on interactive classification techniques, facilitated with AL strategies. The term “Interactive Learning” is proposed, resembling the rationale to keep the user in the loop. We build upon the techniques employed in the pioneering approaches and represent abstractions in our conceptual VIAL process. Bernard et al. [17] propose a process where users play an active role in selecting data

instances and assigning labels. From this work we take the idea to support data-centered, model-centered, and user-centered criteria for label suggestion. A similarity modeling approach for mixed data presents techniques for the interpretation of user feedback and discusses pitfalls for the design of labeling approaches [18]. Mamani et al. [52] propose a visualization-assisted methodology for interacting with instances to transform feature spaces. Finally, previous work included an empirical user study to compare the performance of either AL or VIS [8]. Overarching insights were the complementing strengths of both approaches, suggesting the combination of both in a unified process. Although existing methodological approaches are inspiring, they are specific toward a data type, application goal, user group, or target variable. The VIAL process abstracts from concrete approaches. In addition, VIAL obtains a data-, model-, and user-centric perspective with three outputs: data, models, and knowledge.

Visual interactive labeling techniques We differentiate VIS techniques for labeling tasks by the type of the label to be defined by the user, such as categorical labels, numerical labels, as well as labels that represent a relation between two instances. The type of the data to be categorized differs from video [39], text [38], time-oriented data content [4,81], to two-dimensional data represented in scatterplots [11,63]. In many cases that accept categorical labels, classification models are learned in parallel or as a downstream task. Categorical labeling tasks differ in the number of groups to be assigned. Binary tasks accept yes/no or relevant/irrelevant statements, while multi-labeled tasks consist of three or more classes. Numerical labeling tasks form a second type of labels. Users are enabled to assign fine grained interestingness or relevance scores, e.g., in the context of relevance feedback [66,82]. As an alternative, numerical labels can be used to characterize patient well-being with regression models [17]. Finally, we shed light on labeling approaches that allow the assignment of relations between instances, e.g., in the sense of similarity scores. Using the example of similarity scores, related approaches aim to learn meaningful similarity (or more generally distance) functions by metric learning [104] that reflect the relations expressed by the user [15]. Interaction designs such as the re-allocation of instances in a 2D manifold are used to enable users to express complex relations between instances in the visual space [12,18,52].

3 The visual interactive labeling process

Based on a review of related work in AL and VIS, we propose the VIAL process. We unify the main building blocks into an iterative process consisting of six steps shown in Fig. 2. The VIAL process is special in its focus on exploratory instance

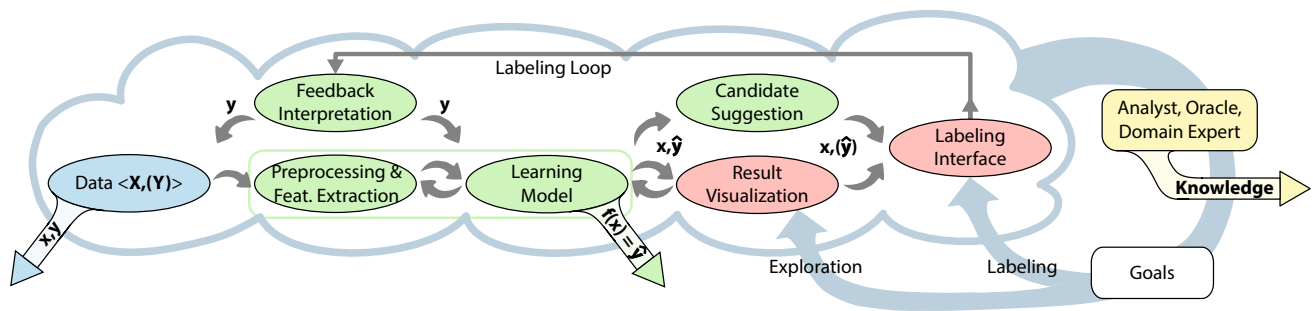


Fig. 2 The VIAL process. Four algorithmic models (green) and two primary visual interfaces (red) are assembled to an iterative labeling process. To resemble the special characteristics of the AL and the VIS perspective, the VIAL process contains a branch (from “Learn-

ing Model” to “candidate suggestion” and “result visualization,” since both are complementary). The VIAL process can be applied for data exploration and labeling tasks. The output of the VIAL process is three-fold: labeled data, learned models, and gained knowledge

selection and labeling tasks, as well as its emphasis on three output types, i.e., labeled data, trained learning models, and gained domain knowledge. All three outputs can be of equal importance, depending on the preference in a particular application scenario. As such the VIAL process provides a novel perspective on labeling tasks that can lead to novel innovative labeling approaches, as well as interfaces and tools for instance selection and labeling.

In the following, we describe each of the six steps in detail. For each step, we present the particular challenges from the ML and VIS perspectives together with benefits and challenges that may emerge when the strengths of AL and VIS are combined in a unified process. In addition, we shed light on pioneer approaches that already address some of these challenges to ease the design of future VIAL solutions. We want to point out that the VIAL process is extensible toward more general problems in ML and VIS, such as visualization of complex parameter spaces originating, for example, from deep neural networks, model visualization, progressive visual analytics, and user-centered design aspects. In section (Sect. 5), we discuss future directions of VIAL toward these general aspects.

3.1 Preprocessing and feature extraction

Preprocessing is a fundamental step in almost every data analysis approach that needs to be handled with care [42]. We refer to preprocessing as the cascade of operations that needs to be applied to ensure the usability of the input data by the models applied in later stages of the process. Preprocessing includes, for example, the identification of erroneous (measurement) data, filtering noise and outliers, and dealing with missing data. We combine the preprocessing step with the mapping of real-world objects into more abstract representations (usually referred to as feature extraction). Existing labeling approaches either directly adopt semantically interpretable attributes of data instances as features

(e.g., the age of a person, the GDP of a country) or apply complex derived descriptors [19] yielding abstract and often high-dimensional feature spaces, e.g., histograms of images or learned basis functions derived from the input signals. Processing and feature extraction are necessary to map the raw input data into a common (potentially high-dimensional) feature space on top of which a model (e.g., a classifier) can be trained. This abstraction is essential to enable (interactive) labeling approaches.

Challenges The analysis of raw data is associated with a series of challenges. In a numerical dataset, a single outlying value that is erroneously represented with a multiple of 1.000 can cause serious problems for downstream models if not detected and cleansed appropriately. In general, many of these problems are associated with quality considerations [42]. Data preprocessing is required to derive secondary data, as a means to yield structured and curated data which are more appropriate for algorithmic models [64]. Using time series data as an example, taxonomies of dirty data [35] can foster the awareness for relevant quality challenges. Accordingly, these taxonomies can be used as a guideline to achieve cleansed data. Visual interactive approaches for preprocessing raw data [13] can support this step, e.g., in close collaboration of designers and domain experts. In a design study including similarity search for a retrieval system, the authors report on a cascade of ten preprocessing steps until data quality was achieved and the users’ information need was accomplished in a meaningful way [3]. Desirable goals for effective data preprocessing include achieving meaningful data representations, comparing raw data with model results, choosing appropriate model parameters, guaranteeing data quality, generalizability of preprocessing workflows, as well as involving users in the preprocessing step [6].

A challenging design consideration in the feature extraction process is whether internal feature representations should be visible to the user. From a VIS perspective trans-

parent feature spaces can be beneficial for the knowledge generation process [36,44] (third output in the VIAL process). The visualization of semantically interpretable features may be particularly beneficial for non-experts. Semantically interpretable features allow for the intuitive exploration of the feature space, ease the identification of important features for the learning success, may support knowledge generation, and may steer the learning into the intended direction [39,40]. The visualization of non-semantic features, however, such as Fourier or Cosine transform coefficients (of, for example, images), is difficult to grasp even for experts. One possible drawback of visible features in a labeling approach is self-biasing. In a recent collaboration with medical experts, we observed how visible features attracted the attention of a physician in an unfavorable manner. As a result, the model to be learned solely reflected the characteristics of these particular features rather than the intrinsic structures of the dataset reflected by all features [17].

Designing and learning useful feature representations that are representative and discriminative for the underlying data is a challenging task in general [2,55]. In the context of VIAL, high-dimensional feature representations such as those obtained from deep neural networks [46] pose special challenges and are currently subject to intensive research [101,105]. The integration of deep learning methods in the VIAL process will highly depend on future progress made in the visualization of such techniques as well as on the interplay of ML and VIS in general [86].

Due to the iterative nature, cases may exist where not the model state but also learned feature representations may change in labeling process. Changing features may, however, confuse the user, which can be addressed with the VIS perspective. VIAL approaches may, for example, provide visual representations showing the evolution of the features, or support the interactive adaption of features [44]. Finally, in VIAL the visualization of the features themselves could further be used as an indicator for training progress, evaluation, and success [39]. This requires, however, that the features are semantically interpretable.

3.2 Learning model

The choice of learning models primarily depends on the data and the labeling task at hand. Classifiers [39] support labeling tasks with binary or categorical labels. Regression models [53] can be used to learn numerical target variables. In addition, the VIAL process includes other types of user-defined labels, such as similarity relations between two or more instances (i.e., metric learning) [7,29]. In the VIAL process the learning model is directly coupled with visual interfaces facilitating analytic reasoning and model refinement [32,84]. The VIAL process is iterative by nature. Thus, it requires learning models that are instantly re-trainable (ide-

ally in real-time) and that adapt their internal parameters to changes in the training data. Good candidates for VIAL are supervised incremental and online learning methods [47].

The trained models represent a primary output of the VIAL process, building the basis for downstream applications.

Challenges Many learning models (i.e., their internal state) are difficult to visualize [36,85]. Traditional machine learning models such as, for example, support vector machines are defined in high-dimensional spaces with complex (potentially nonlinear) decision boundaries. Neural networks challenge visualization with their millions of different internal parameters, their complex connections, and the absence of an explicitly given decision model [101].

Another important issue is to select a suitable termination criterion for learning and labeling to avoid unnecessary labeling effort. In a recent user experiment, we observed that many participants asked when the labeling will be finished [8]. Visualization of the learning status may be one means to inform users about the progress and to avoid potential frustration. Examples are linecharts of the accuracy of classifiers, or the representation of the variety of instances in the dataset that have already been used for labeling [8,17]. From a modeling perspective various alternative termination criteria can be used to inform the user in a visual way. Due to the limited capacity of most classifiers [43,90] the learning progress converges at some point in time. Termination criteria can be both intrinsic (e.g., model change) or extrinsic (e.g., classification accuracy) [71]. A traditional VA criterion is measuring quality aspects that help analysts to validate model convergence [36]. From a ML perspective, a useful strategy for a termination criterion is to detect overfitting of the classifier. Overfitting means that the classifier adapts too much to the training data and thereby loses its ability to generalize to new data. Suitable measures help to avoid overfitting [77].

Another challenge is the choice of the learning model itself. A series of VA approaches addresses this problem, e.g., by the visual comparison of competitive feature sets and algorithmic models [44,53]. In addition to the general problem of algorithm choice, we shed light on the challenge associated with complex learning functions. In fact, categorical labels are usually needed for supervised classifiers, and numerical labels often for regression-based approaches. However, the design space for learning models increases in situations where relations between instances $r(x_1, x_2)$ or distance (or similarity) functions $d(x_1, x_2)$ have to be modeled. An example is metric learning where distance relationships between instances are learned from the data directly [7]. Research has shown that visualization is able to support this process [12,29].

3.3 Result visualization

The visualization of model results corresponds to the VIS perspective on the labeling process. According to the VIS principles, result visualization is particularly suited to amplify knowledge generation [22,83] as one output of the labeling process. Beneficial examples for the VIAL process include visual representations of classification results [74], regression models [53], or results of learned distance functions [12]. We identify three primary benefits for the VIAL process. First, result visualization can facilitate exploration tasks supporting hypotheses and insight generation about the data as well as the knowledge generation process [83]. Second, tightly coupled learning models and result visualizations enable user-centered model refinement [68,92]. Third, result visualization allows users to select meaningful candidates for labeling and thus serves as a complement to model-based AL heuristics for the suggestion of candidates [39,74].

Challenges We draw the connection to general challenges in the visual representation of high-dimensional data. Visual interactive interfaces supporting overview and detailed visualizations are one option to tackle this issue. Dimensionality reduction [86] and data aggregation techniques [28] help to condense the data, for the price of individual challenges. Examples for individual challenges include the applicability, quality, or uncertainty of algorithms in connection with their parameters. Using dimensionality reduction as an example, it cannot be guaranteed that intrinsic structures of the data are retained after the reduction and that new structural patterns are introduced that are not present in the original space.

A particular design challenge for labeling approaches is whether and how predicted labels should be visualized. A recent user experiment showed that users change their strategies in selecting instances considerably when additional information about predicted models are depicted [8]. Showing unlabeled instances with labels predicted by the model supports the comparison of current situation with targeted label information, i.e., users have a means to improve the learning model in a visual interactive way. Users can assess whether the learning model is able to explain the already labeled portion of the data. However, showing predicted labels may also cause biases. In contrast, representing instances without their predicted labels helps to focus on data-intrinsic properties such as structural information (e.g., patterns, clusters, or outliers) [8]. In this connection, visual data exploration might support instance selection and labeling (see Fig. 2).

A central role in VIS is the direct manipulation of learning models and the analysis of respective outcomes. To facilitate this goal, VA approaches such as parameter space analysis support [76], or techniques for the visual comparison [33] of different model outputs can be leveraged.

Another class of challenges addresses the visualization of the model state itself, e.g., for the identification of dependencies between hyperparameters defining the model and input variables. While the standard machine learning process does not require visualizations of the model state, VIAL can benefit from the awareness of knowledge between the built model, the data input, and possibly the semantics of a given analysis scenario. The identification of shortcomings in the training (e.g., bias toward a class, misleading features, overfitting) is another beneficial but challenging goal to be addressed with model visualization. As an example, in practice the visualization of decision trees is one way to support both the awareness of dependencies between data and model, as well as the selection of meaningful labeling candidates [92].

3.4 Candidate suggestion

Automated candidate suggestion (as in AL) and the visualization of model results (from VIS, see Sect. 3.3) represent two complementary alternatives for the *identification and selection* of labeling candidates. From an AL perspective, users are queried in a model-centered way that is assumed to improve the model accuracy most [71]. In turn, in the VIS perspective the user is typically assigned an active role in the candidate selection process, e.g., by leveraging the gained knowledge about patterns in the data. With the VIAL process, we seek to join both perspectives and propose to either include AL-based strategies as guidance concepts in visual interfaces, or to leverage visual interactive interfaces for the analysis, verification and steering of AL strategies.

Pioneer VIAL implementations [17,39,74] provide both options and give an idea of the potential of combined candidate suggestion and selection strategies. However, approaches that actually combine both strategies to form a generic and unified candidate selection strategy are still missing. As such, research in hybrid approaches remains an open topic and a promising direction of future research.

Challenges A major challenge in candidate suggestion stems from the AL process, i.e., the selection of the most beneficial candidates for labeling. A rich set of techniques for candidate suggestion exists with differences in, for example, the query strategy [71] or the computation costs [73] (capability for interactive execution [54]). The applicability of individual AL heuristics depends on the data, the types of labels, and the ML model [71,98], as well as on the interplay of model-based and user-based candidate selection. From a VIAL perspective the suitability of AL strategies for different user, data, and task characteristics remains an open issue.

One central aspect is the cold start problem in AL when no labeled instances are available at all. User-based strategies have shown to outperform AL in early phases of the labeling process; however, questions arise about break-even

points and the trade-off between both strategies in the course of the remaining labeling process [8]. In addition to the latter, pioneer user experiments started comparing AL with the user-based selection of candidates with respect to different factors such as multi-instance labeling, annotation time, quality measures, and selection strategies [8,45,51,72,74]. It remains subject to future work to use the gained knowledge and design hybrid approaches for combined candidate selection.

Considering the need for very large labeled datasets a downstream challenge is the efficient inference of gathered label information to yet unlabeled instances and the exploitation of unlabeled data for the learning. This goal can be addressed by semi-supervised learning which is a branch of supervised learning that tries to improve performance by exploiting information from unlabeled instances [26]. In VIAL, semi-supervised learning could be integrated in two ways: (1) to learn more robust models by taking unlabeled instances into account and (2) by inferring labels from labeled to unlabeled instances. Both aspects help to reduce the required user interactions in the VIAL process.

3.5 Labeling interface

The goal of the labeling interface is to accept labels y from the user which can either be assigned to instances x directly or be used to characterize relations between instances. This data-centered output of the VIAL process can be used to enrich datasets, e.g., for the creation of ground truth data. Every time a user labels an instance in the labeling interface, the labeling loop can be triggered, possibly leading to an improved learning model (see Fig. 2). This iterative user-in-the-loop approach is supported from both the AL and the VIS perspective and is resembled in the VIAL process. Particularly the VIS perspective requires meaningful visualization and interaction designs to support labeling in a meaningful way.

Challenges One challenge is the visual mapping of labeling candidates, i.e., the visual representation of instances in the visual space [22]. In some cases the visual representation of instances is straightforward, e.g., for handwritten digits or other image data. However, other data types build upon complex data characteristics that are more difficult to visualize. Examples from the literature include visual representations of unknown patient histories [17], abstracted features [44], relations between clusters and metadata [14], or poses in human motion capture data sequences [4]. In fact, in order to submit qualified feedback, users must be able to grasp the characteristics of queried instances. In case users already know individual instances, visual identifiers can be used, e.g., national flags for countries [18] or images of soccer players [15]. In other cases where users already have

an intrinsic knowledge of the labeling alphabet, e.g., object classes like cats and dogs, visual identifiers are also suitable. A recent VA approach using the example of personal image collections is PICTuReVis, allowing visual interactive labeling (and classification) of personal image data [91]. In more complex situations where, for example, multimodal data or unknown instances (e.g., of a new class) have to be identified and labeled, visual identifiers become insufficient. More abstract representations and visualizations need to be designed to support decisions. Special interaction designs may be required to explore the instances in detail, e.g., in case of 3D objects or complex data objects with many attributes.

Another challenge relates to rather exploratory situations where users are not aware of all inherent phenomena of a dataset. Labeling only small portions of instances users have knowledge about may remain uncovered areas of the data space. As a result, learning models may hardly reflect all characteristics of the dataset. The visualization of uncertainty information of the learning model can be one means to amplify the users' ability to select meaningful candidates from a model perspective. A recent user experiment showed that visualizing intermediate class distributions predicted by a classifier supported the participants in the selection of useful instances for model refinement [8].

Finally, the interaction design raises challenges in complex learning situations where labels are less distinct and exhibit complex semantics. Categorical and numerical labels can easily be submitted with straightforward text boxes, sliders, or a set of pre-defined labeling options. Although this works well for many cases, more complex learning situations may exceed the limits of baseline interaction designs. For example, applications where models learn relations between more than two instances require more sophisticated interaction designs. One class of approaches where this problem was addressed uses 2D layouts where instances can be arranged spatially (spatialization [29]) to obtain user feedback. These examples exploit the user feedback to learn similarity relations of the data for dimensionality reduction, clustering, or visual interactive similarity search [12,18,30,31,100].

3.6 Feedback interpretation

An often neglected question is how to interpret complex user feedback and pass it to the learning model [86]. We assume that the difficulty in interpreting feedback is related to the complexity of user interaction. For simple labeling tasks such as selecting a category, feedback interpretation may be straightforward. For more complex tasks the situation becomes more challenging, e.g., when users create topologies with multiple instances and the feedback interpreter exploits the relations between multiple instances [12,31]. Similarly, *implicit* user feedback is more difficult to interpret where user behavior is observed and conclusions are drawn

from behavior without explicit queries. In general, meaningful feedback interpretation can facilitate the data-centered and model-centered outputs of the VIAL process.

Challenges We elaborate challenges in feedback interpretation from two perspectives: the *concrete interaction* and the *abstract user intent*. The first perspective arises in more complex interaction paradigms that go beyond text boxes and slider controls used for feedback submission. Two example interactions are (1) spatial re-arrangements of instances in 2D as described in Sect. 3.5 [12] and (2) ranking items in list-based interfaces [97]. To address this challenge, we draw the connection to *observation-level interaction*, a sensemaking technique based on semantic exploration of data [31,100]. The technique explicitly addresses the question how users can interactively express their reasoning on observations about instances instead of directly modifying parameters of analytical models. For user feedback based on re-arrangements of instances in 2D at least three different mechanisms exist [18], i.e., the absolute positions of instances, the relative positions (distances), as well as the vicinity of nearest neighbors. The authors postulate that the labeling interface should be designed in a way that it informs users about how their feedback is interpreted by the system. If the feedback concept is based on ranked items [97], pointwise, pairwise, or listwise approaches can be used to interpret the user feedback and improve the learning model [48].

Challenges from the second perspective are related, but take the discussion deeper into human–computer interaction. Mental models [24] of users communicated through visual interactive interfaces [49] open large spaces for interpretation and thus may deviate from the measured feedback. Implicit feedback falls into this category [56], as well as data from sensor devices such as eye tracking [10] which may be addressed in future VIAL approaches to capture user feedback and to derive labels continuously. Another challenge associated with implicit feedback is the trust of labels generated from such processes. While user-generated labels usually form ground truth data for learning tasks, implicit feedback, however, may require additional uncertainty modeling.

4 Usage scenarios

In Sect. 3, we presented the six phases of the VIAL process using relevant pioneer approaches as examples [4,8,15,17,38,39,74,100]. In this section, we discuss two real-world VIAL approaches in detail which represent heterogeneous and complementary examples. This line of approach is inspired by Sacha et al. [83] using, for example, Jigsaw [75] as an explicit example. Our rationale is to demonstrate the range of possible applications, to shed light on solutions that actually address

the described challenges, and to validate our process model and the interplay between AL and VIS.

To achieve high variability, the selected use cases differ in at least one aspect in each of the six phases on the VIAL process, i.e., (1) semantical vs. non-semantical features, (2) label types and learning models, (3) visualizations of the model state, (4) AL model used for candidate suggestion, (5) the type of visual representation of instances, and (6) feedback interpretation and propagation. The first approach applies VIAL for the generation of categorical labels in the context of object detection in videos. The second approach applies VIAL for learning similarities between complex objects (soccer players). Both approaches are concrete instances of the VIAL process and address all the six steps. Note, however, that not every VIAL step is exploited with the full complexity the general VIAL process allows. The gap that remains between the concrete examples and the general VIAL process are good indicators for future work.

4.1 Visual interactive labeling for video classification

4.1.1 Introduction

In visual classification and recognition, classifiers are usually pre-trained on large amounts of annotated video data and are then applied to new data (e.g., face detection in surveillance applications). Pre-training is possible when the target object is known a priori and when sufficient training data for this object has been collected and annotated. Pre-training is, however, not possible for ad hoc queries defined by the user because in general no pre-defined classifiers and labeled datasets exist a priori for such queries. Höferlin et al. present an interactive learning-based approach for the detection of user-specified objects in a video that allows the user to actively select instances for labeling and thereby to steer the learning process [39]. Additionally, the trained model is visualized by VA methods to provide feedback and to increase the trust of the user in the model. The authors demonstrate their method for the task of person identification in videos. An illustration of the approach is depicted in Fig. 3. This work is an early example which resembles all steps of the VIAL process to a certain degree. For this reason, we select this work as a first usage scenario and discuss it from the perspective of the VIAL process.

4.1.2 The VIAL process

(1) Preprocessing and feature extraction The data employed by Höferlin et al. are surveillance videos from a multi-camera tracking dataset. According to the authors no special preprocessing is performed on the videos and they are taken as they come from the dataset. The authors addressed the learn-

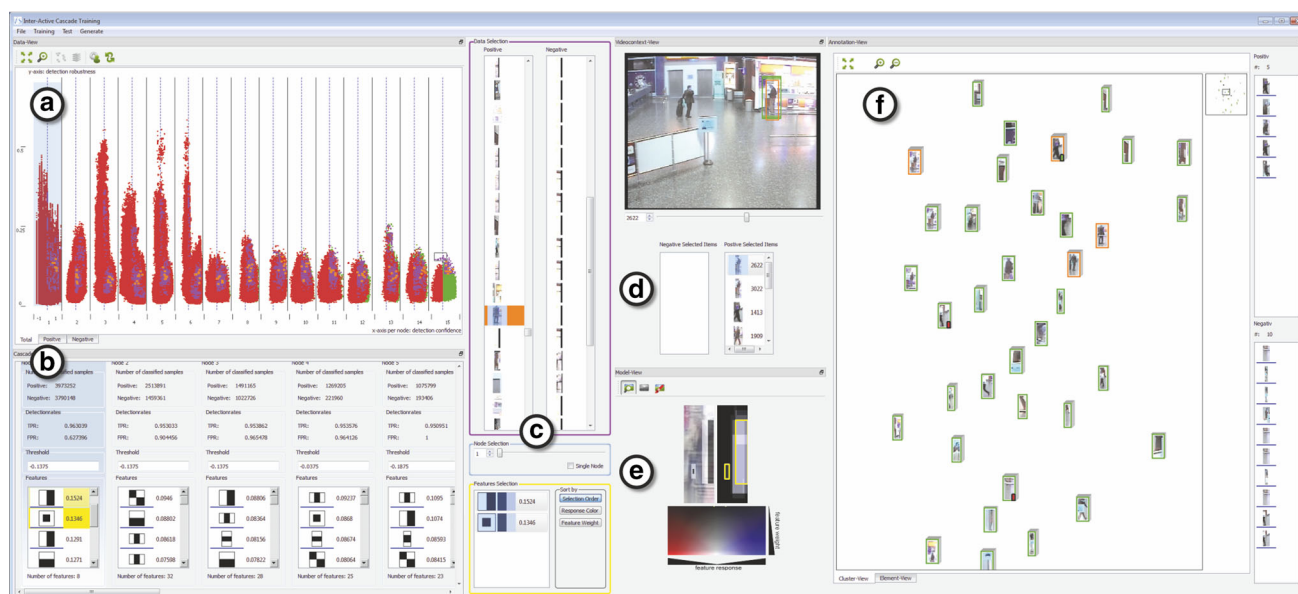


Fig. 3 A screenshot of the VIAL approach by Höferlin et al. [39]. The interface is split into several interlinked areas. *a* The scatterplots for each strong classifier of the cascade, *b* detailed view of each strong classifier including the respective features selected by the classifier, *c*

selected instances by the user (from the scatterplots), *d* visualization of selected instances in the original video frame (context), *e* visualization of rectangle features on selected instances, *f* a 2D view for instance labeling. Reprint with permission of IEEE

ing of representations by employing rectangle features which gained great popularity with the strongly influencing work of Viola and Jones [93] on face detection and which are applicable to arbitrary objects. Rectangle features represent image filters composed of black and white areas. They act as basic edge and corner filters and thus can be directly interpreted visually by humans (see Fig. 3b for examples). Thus, they are well suited for a visualization of the internals of the learning model which is built on top of them in the course of the VIAL process.

(2) Learning model The major challenge in model learning is to identify which combination of filters are most characteristic for a certain type of object. To find a suitable representation, the learning model selects rectangle features dynamically and trains a cascade of classifiers from these features. Each input instance is fed through the entire cascade. Only if it passes all classifiers in the cascade with positive prediction, it is declared as a positive example. Otherwise, it is rejected. This asymmetric scheme is necessary to account for the fact that there are by far more negative instances than positive instances in typical visual detection tasks.

For model learning, the authors employ a slightly modified version of the boosting algorithm proposed by Viola and Jones [93]. Adaptions to the training algorithm were necessary to enable fast iterative re-training of the algorithm (online Adaboost) which is particularly important in VIAL to provide instant feedback to the user. Therefore, the authors

solved the first challenge related to the learning model (see Sect. 3.2). The second challenge is the specification of a stopping criterion. In the approach the point in time when training should be terminated is defined by the user. To support the user in this decision, performance metrics such as true positive rate (TPR) and false positive rate (FPR) are provided (see Fig. 3b). Although this is a sound approach, it has the drawback that the user needs a certain knowledge of the performance metrics and some basic principles of ML. The use of visually interpretable features, their visualization, and the possibility to provide relevance feedback on them to guide the learning process (see below) represent basic concepts of the VIAL process.

(3) Result visualization Challenges addressed in the context of result visualization are the visual representation of the model state, its learned representation, and designing visual techniques that enable the user to guide the learning process. Höferlin et al. design several visualizations to show the internal workings of the machine learner. First, for each classifier in the cascade a scatterplot is shown that displays the decision boundary and the positively and negatively predicted instances (left and right of the boundary, see Fig. 3a). From these scatter plots single or even multiple instances can be selected. The selected items are visualized in Fig. 3c. Note that this visualization does not provide the true label of already annotated instances. Adding this information could further improve the guidance in the selection and exploration

process. In a second view (Fig. 3b), the selected rectangle features (image filters) are displayed for each strong classifier together with their importance. This view helps the user to inspect which types of image structures (e.g., vertical lines, horizontal lines, corners) the classifier relies on and whether these patterns also reflect what the user expects to be important structures for the searched for objects (inclusion of domain knowledge). To further verify what the classifiers learn, the user can overlay the rectangular features over selected instances, see Fig. 3e, and provide relevance feedback to the learner.

(4) Candidate suggestion The major VIAL challenge that Höferlin et al. address in candidate suggestion step is the automated selection of the most important instances to be labeled. For this purpose [39] implement an AL strategy based on uncertainty as an alternative to user-driven candidate selection. The authors compute the uncertainty of the learner for each instance and present the most uncertain instances to the user for annotation [34,95]. This process is initiated only on demand by the user and is thus an optional step.

In the concrete example the selected instances are directly shown to the user (see Fig. 3f). An interesting extension following the idea of VIAL would be to highlight these instances also in the result visualization, i.e., in the scatterplots of each classifier. Furthermore, visualizing the most uncertain instances in the (projected) feature space or relative to the decision boundary would allow the identification of ambiguous features which may lead to wrong decisions. The usage scenario, complemented with the latter described aspects, indicates the full scope of the VIAL process. VIAL aims at a stronger integration of AL and VIS, for example, by the combination of AL-based and user-based strategies for candidate selection. This has hardly been investigated so far and is a challenging topic for future research [8].

(5) Labeling interface The primary VIAL challenge addressed in the context of the labeling interface is the visual mapping of candidate instances and their arrangement. The authors visualize the instances for labeling in 2D in a similarity-preserving way (see Fig. 3f) to support the exploration of the space. Similar objects are grouped into clusters and can be labeled simultaneously which saves time. Thumbnails are used for the visual mapping, which is natural as the instances represent images. For each thumbnail the predicted class is shown as well as the true class if instances have already been labeled. To provide contextual information, selected instances are further shown in the context of the video frame they stem from, see Fig. 3d.

The interface enables the quick tagging of instances. The approach could be extended with an encoding of uncertain instances where the true label is different from the predicted

one. Such items are particularly interesting as they may indicate that the learning model is not able to capture the complexity of the data (e.g., too simple decision boundary, underfitting). An important problem emphasized in the VIAL process and solved by the particular labeling interface is the cold start problem, i.e., the initialization of the learner when no labeled instances exist at all. In the concrete example, the user is simply asked to query positive and negative instances (used to build an initial model).

(6) Feedback interpretation The last step to complete the VIAL loop is feedback interpretation. In the approach, feedback interpretation is rather straightforward as (1) user feedback is explicit and (2) feedback consists of pre-defined mutually exclusive categories (i.e., person or no person). Thus, the provided label is simply passed to the machine learner as provided by the user. We want to note, however, that even in this simple example there is a certain room for interpretation and a certain fuzziness. Should, for example, an instance which shows a person only partially (e.g., one half or less) be considered a positive instance? And where to draw the border between positive and negative instance, i.e., how much of a person must be at least visible to consider an instance positive? There are in most labeling tasks such ambiguities which need to be treated in some way. VIAL provides a means to make such ambiguities explicit to the user and to raise awareness toward such issues.

4.1.3 Application example

Höferlin et al. demonstrate the capabilities of their system on the task of person identification in videos. The VIAL process starts with completely unlabeled videos. In an initial step the user annotates persons in the videos (positive examples) by drawing rectangles. Next, negative examples are generated in the same way. After an initial training on these examples interactive labeling is performed to improve the result. It shows that after a few cycles a similar performance level can be reached than with pure AL (uncertainty sampling) but, however, with much less training examples. This indicates the potential of combining user interaction and active learning in the VIAL process.

The approach generates all three outputs of the VIAL process. Labeled data are obtained in the form of positively and negatively labeled image regions. The learned model is achieved together with a representation that can be reused for the detection of the target object. The gained knowledge of the user is twofold. First the user gains insights into the inner workings of the machine learner. Second, the user gets knowledge about the application domain, its challenges (difficult cases), ambiguities (e.g., patterns which are easily confused with the target object), and misleading image structures which leads to a better understanding of the task.

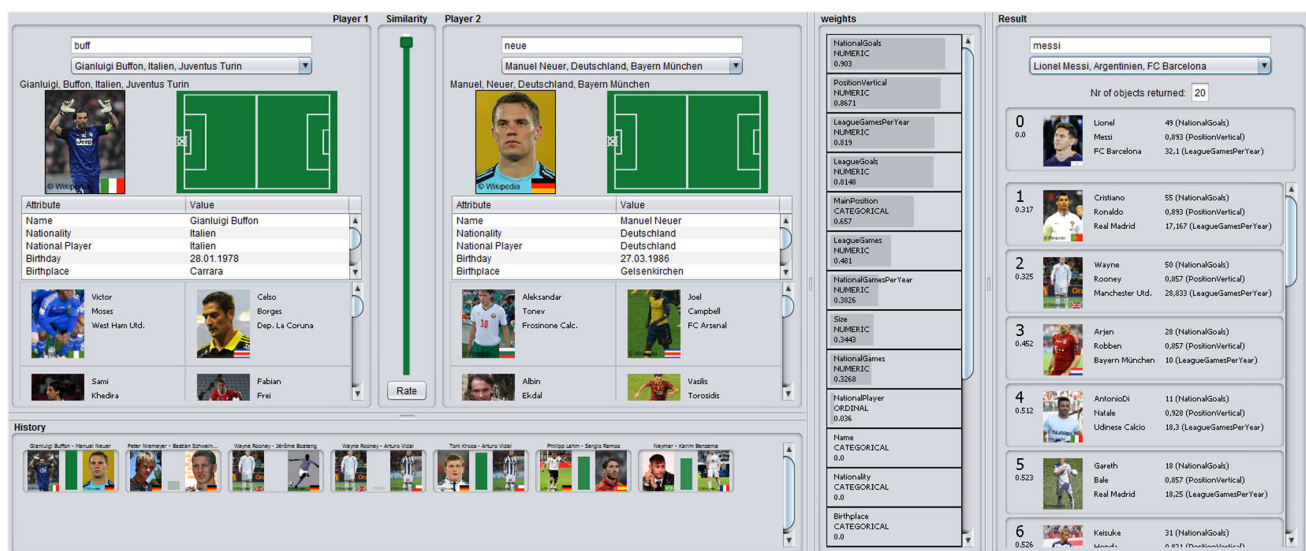


Fig. 4 Visual interactively labeling the similarity between pairs of soccer players [15]. An active learning model suggests other players to be labeled next. The learned similarity function can be validated in a retrieval component. Most recently, the user has defined a 100% similarity between the two goal keepers Gianluigi “Gigi” Buffon and Manuel

Neuer. At the bottom previously labeled players can be seen including their similarity scores (bars in green-to-gray). The retrieval algorithm was applied with Lionel Messi; according to the user-defined similarity function most similar soccer players are Cristiano Ronaldo, Wayne Rooney, and Arjen Robben

4.2 Visual interactive labeling for similarity modeling

4.2.1 Introduction

The second usage scenario describes a similarity learning approach applied for soccer players [15] (see Fig. 4). Experts as well as non-experts can create a personal similarity function for soccer players in a visual interactive way. In general, such similarity functions can subsequently be used for downstream algorithms to address retrieval, clustering, or other ML tasks. In this approach, soccer players are characterized by a series of attributes gathered from Wikipedia; all attributes are semantically interpretable. A visual interactive interface allows the definition of similarity relations between pairs of soccer players. A similarity modeling component interprets the user feedback and builds a model that can be applied on the entire dataset. A retrieval interface supports the visual validation of (intermediate) similarity models for soccer players. Overall, the approach resembles all six steps of the VIAL process.

4.2.2 The VIAL process

(1) Preprocessing and feature extraction The dataset of the approach consists of (European) soccer players represented by a set of attributes (name, position on the field, goals, national games, size, etc.). The feature vector used in the approach consists of primary attributes extracted from

Wikipedia (DBpedia) and secondary attributes derived from primary attributes. As such, all features are semantically interpretable, even by non-experts. The approach addresses the challenge to deal with mixed data, i.e., the features can be either be numerical, categorical, or binary. Challenges in the data preprocessing process were the acquisition of data, as well as dealing with missing values. The authors’ design foresees an explicit visualization of existing features (attributes) in order to facilitate the visual comparison of player characteristics in detail, see Fig. 4 (left). This design decision comes with the cost of self-biasing caused by a focus on individual attributes (see Sect. 3.1). Simple representations make the approach applicable even for non-experts, e.g., a straightforward glyph design depicts the position of players on a soccer field.

(2) Learning model The function to be learned is $f(x_1, x_2) = y$ where y is a numerical value between 0 (dissimilar) and 1 (very similar), and x are the players. A two-step approach first correlates every feature of the dataset against the user labels, and then, pairwise distances for all pairs of players in the dataset can be calculated. Every attribute is correlated individually with different correlation measures for numerical, categorical, and binary features. Individual treatment of features is fast (allows rebuilding the model in real-time), and results are easily interpretable, i.e., a weight is calculated for every feature. However, the approach does not address dependencies of attributes between each other. Different distance measures for numerical, categorical, and binary features are

applied, until a global similarity value is calculated, successively. Differing weights of features indicate model changes and build the basis for an intrinsic and comprehensible termination criterion. The model-centered output of the approach is an algorithmic measure that reflects user-defined similarity.

(3) Result visualization In line with the VIAL principles, the approach provides visual interfaces for both model visualization and model result validation. A straightforward list-based interface with barcharts depicts the weights of individual features, see Fig. 4 (halfway right). In this way, users can analyze the fitness of every feature to the given similarity feedback. A retrieval interface applies the similarity function and represents its results visually, see Fig. 4 (right). Similarity scores for every retrieved player give an indication for the similarity relations of retrieval results with respect to a queried player. As such, users have a means to interact with the data, explore the dataset, and gain knowledge, for example, about previously unknown players. The visual retrieval result can also be used for the selection (and labeling) of new pairs of players. With the retrieval component, the authors omit the challenge to represent the data of the high-dimensional feature space in an abstract way. One drawback of the strategy is a missing overview of the similarity relations of the entire dataset.

(4) Candidate suggestion The approach combines the selection of candidates with AL and VIS, as recommended by VIAL process. The AL strategy builds upon a data-centered criterion, i.e., it exploits structure of the data [73]: suggested candidates have highest distances in the vector space to all labeled instances. A set of players is suggested and depicted in the list-based AL interface. VIS principles are implemented in three different ways. First, a textual search allows querying players known to the users. Second, the history can be used to re-select (an re-label) players, see Fig. 4 (bottom left). Third, players retrieved with the retrieval component of the tool can be selected for labeling.

(5) Labeling interface According to the VIAL process, the approach builds upon a highly iterative labeling strategy. The labeling interface visualizes two players, a slider control in between allows the assignment of a similarity score, see Fig. 4 (left). A tabular interface depicts details about attributes (features) of every player as a means to grasp the characteristics of queried instances. The glyph with the position on the soccer field allows direct comparison of this (probably important) attribute. Images of soccer players serve as visual identifiers of the players. At least experienced users are able to lookup known players without difficulties.

(6) Feedback interpretation With the slider control users can explicitly define the similarity score between two players.

However, the complexity of the approach is the similarity relation in itself ($f(x_1, x_2) = y$). The authors address this problem with correlation measures used to relate every feature to the given similarity feedback. Features with high weights match the similarity characteristics of the set of labeled players. Pairs of players in combination with assigned similarity scores define the data-centered output of the approach.

4.2.3 Application example

In Fig. 4 the approach can be seen in practice. The history of labeled pairs of players at the bottom reveals that 7 pairs have been labeled yet. Most recently, two of the best goal keepers in the world were labeled with a 100% similarity (green bar in the labeling interface). The weights of the similarity model (list-based bar visualization center-right) indicate that the notion of similarity of the user is mainly based on national goals, the (vertical) position on the field, as well as the numbers of league games and goals per year. According to the learned similarity function, the soccer player most similar to the queried player Lionel Messi is Cristiano Ronaldo.

5 Discussion and future work

In this work, we carved out the benefits of joint approaches using AL and VIS for labeling data instances. While we focused on the conceptual baseline, the quantification of success of the VIAL process remains future work. When describing the six core steps of the VIAL process, we went for a broad overview of techniques, but also for challenges and existing solutions. In addition, we presented explicit and holistic reflections of two application examples in the context of VIAL. Future work includes implementations of the VIAL process in application examples with the aim for a broad range of real-world settings. In this connection, collaborations with experts seem to be a promising approach, allowing the application of principles from user-centered design and design studies. Another line of future work involves points of reference such as deep learning or progressive visual analytics to be included in the VIAL process more explicitly. Finally, future work includes evaluations of the challenges described in the six steps.

5.1 Deep learning

Deep learning has gained a lot of attention recently. We discuss possible connection points to the VIAL process. Neural networks in general are suited for VIAL approaches as they can be trained and refined iteratively. However, a number of limitations may hamper the use of deep networks in the VIAL process.

First, training and refinement usually takes considerable time which makes real-time training hardly feasible today. The capability of algorithms for interactive execution [54] is a challenge for many ML approaches in general, so it is for the VIAL process. Second, large numbers of labeled data are necessary for initial training which strongly conflicts with the cold start problem that is present in many situations. This requirement does not meet the scope of the VIAL process; rather, deep learning may be a well-suited downstream process when training data are available as a result of the VIAL process. In addition, a solution for the latter problem may be *transfer learning* where existing models trained for similar tasks are fine-tuned incrementally instead of training from scratch from the given data [5,58,102]. This approach saves training time considerably and could be a first step toward the integration of deep learning into VIAL. Transfer learning requires that a compatible model already exists, such as the model-based output of the VIAL process. Third, a big and still open challenge in the context of VIAL is how to visualize the *representations* learned by the network and the *model* itself. Basic visualization techniques for representations learned from images have been introduced recently [65,101,103,105]. However, the challenge remains of how to visualize thousands of such learned representations and their hierarchical relations. For more abstract and complex data (e.g., multimodal data) which have complex relations and are difficult to interpret visualization becomes significantly more challenging [60]. We thus conclude that an integration of deep learning into VIAL is at the current state of research hardly feasible; however, research that tries to mitigate current limitations is currently ongoing.

5.2 Visualization of active learning models/strategies

Recent experiments have shown that candidate suggestion by a machine learner and candidate selection by the user may differ significantly and that user-based selection can outperform AL [8]. One strength of VIAL is the combination of both strategies. This combination opens up a novel design space for combined selection and suggestion strategies. From the perspective of AL the major novelty lies in the visualization of active learning strategies. Such a visualization should, on the one hand, enable to validate a given strategies against the intuition of the user. On the other hand, it makes the learning process more transparent for the user and fosters a deeper understanding of the model and its actual “needs” (in terms of labeled instances). Furthermore, it bears a potential to improve existing or develop novel AL strategies by learning from the user and his or her selection strategies.

The type of AL strategy strongly influences the visualization approach. Purely data-driven approaches such as density-based selection [99] can be visualized more easily

because density is a concept well perceivable by humans. For relevance-based approaches [94], visual identifiers or glyphs (in case of more abstract data) may be good measures to visualize suggested instances and to let the user verify their actual relevance for a given category. For strategies that are driven by model-specific properties, such as error reduction schemes [59] and uncertainty sampling [69] but also for query by committee [79] the visualization becomes more complex, because internals of the models need to be visualized to explain to the user why a certain instance is at the current point of learning the most important one for the learner. The development of visualization techniques for this purpose is a challenging direction for future work.

5.3 Progressive visual analytics

The VIAL process intrinsically relies on the smooth combination of iterative user input and computational models. With the growing complexity of the computational components, this approach needs to balance a critical trade-off though. On the one hand, the user expects a fluent interaction with immediate feedback from the labeling system and the underlying computational models. Delays as small as seconds can substantially interfere with and interrupt the user’s workflow [25]. Thus, fast and fluent responses from the system are crucial. On the other hand, however, increasing complexity of models often comes with increasing computational requirements, that is, longer run times. The user would thus need to wait for results longer.

To deal with this trade-off, researchers have suggested to work on progressive visual analytics solutions [80]. The basic idea is, instead of waiting for the final results of computational components, to quickly compute intermediate results, offer these to the user, and refine them over time [23,54,87]. An interesting strategy, for instance, might be to train a classifier first on a subset of the original input data only (to keep response times short). Once trained, data from the remaining set can be used to refine the model. Restricting the amount of data used for a single training pass could lead to a reduction in disturbing delays and therefore to a smoother user experience. Such approaches will be crucial for successful implementations of the VIAL process, specifically if the underlying models come with a high computational complexity.

5.4 User-centered design and evaluation strategies

The current state of pioneer VIAL approaches demonstrates the applicability of VIAL for various combinations of datasets, application fields, label types, and learning models. Most existing approaches have an emphasis on techniques, as well as on data and control workflows, proving the applicability of VIAL. One aspect with remaining potentials is

designing VIAL approaches for specific user groups. Following the design study principle [78], future approaches may emphasize methodology for the characterization of users and their application problem, an iterative design phase, and a careful reflection of lessons learned in the design process. Similarly, more general principles from user-centered design can be borrowed to facilitate both usefulness and usability.

Similarly, the evaluation of VIAL approaches is still in an early state. We observe that several existing works already validate learning models, as well as the model conversion process [15–18,67]. However, many other types of evaluation are possible. One strategy for future work may be the validation of all six steps proposed in the VIAL process. This principle helps to validate that individual modules provided in the interactive and iterative workflow work together seamlessly. Another promising direction for evaluation strategies is the assessment of strategies for candidate selection either performed by AL, by users, or by both. A recent experimental study identified that the strategies for the selection of candidates even differs for various users [8]. This opens new evaluation strategies for the assessment of various user-based strategies, as well as in the corporation of analytical guidance.

6 Conclusion

We presented the VIAL process that adopts and extends the process model from AL and VIS and therefore combines the strengths of model-centered active learning with user-centered visual interactive labeling. In the same way, the VIAL process mitigates the missing knowledge generation in AL as well as shortcomings in data creation observed in VIS. Based on a review of the AL process as well as pioneer VIS approaches for labeling, we identified six core steps assembled to the VIAL process. For every step, we described both the AL and the VIS perspective, discussed respective challenges, and outlined possible solutions. In two inspiring examples, we indicated the applicability of the VIAL process also for other application domains. Finally, we discussed four core aspects in the broader sphere of VIAL and indicated connecting points referred to as future work.

In summary, VIAL provides a novel perspective on labeling by bringing together two complementing approaches with a common goal. The VIAL process opens up a space for the design future labeling approaches that has not been explored so far. In particular, the combined candidate selection and suggestion—which is the point where AL and VIS meet—represents a challenging direction of research which bears the potential for significant improvements in existing labeling strategies. VIAL may lead to faster model convergence and more generalizable models in shorter time. In this connection, we amplified the potentials of combining the

strengths of both humans and algorithmic models in a unified process, and outlined possible future work addressing this direction.

Acknowledgements This work is an extended version of a previous EuroVA paper [20] entitled “A Unified Process for Visual Interactive Labeling.” This work was supported by the Deutsche Forschungsgemeinschaft (DFG), Project No. I 2850, Lead Agency Verfahren (DACH) “Visual Segmentation and Labeling of Multivariate Time Series (VIS-SECT),” the Austrian Research Promotion Agency (FFG), Project Nos. 7179681, 7189193, the Austrian Ministry for Transport, Innovation and Technology under the initiative “ICT of the future” via the project “VALiD” (Project No. 845598), the Austrian Research Fund (FWF) via the projects “KAVA-Time” (Project No. P25489-N23), and “VisOn-Fire” (Project No. P27975-NBL), as well as by the Lower Austrian Research and Education Company and the Provincial Government of Lower Austria (NFB), Department of Science and Research via the project “IntelliGait” (Project No. LSC14-005).

References

- Attenberg, J., Provost, F.: Inactive learning? Difficulties employing active learning in practice. *SIGKDD Explor. Newsl.* **12**(2), 36–41 (2011). <https://doi.org/10.1145/1964897.1964906>
- Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
- Bernard, J., Daberkow, D., Fellner, D., Fischer, K., Koepler, O., Kohlhammer, J., Runnwerth, M., Ruppert, T., Schreck, T., Sens, I.: Visinfo: a digital library system for time series research data based on exploratory search—a user-centered design approach. *Int. J. Digit. Libr. (IJoDL)* **1**, 37–59 (2015). <https://doi.org/10.1007/s00799-014-0134-y>
- Bernard, J., Dobermann, E., Vögele, A., Krüger, B., Kohlhammer, J., Fellner, D.: Visual-interactive semi-supervised labeling of human motion capture data. In: *Visualization and Data Analysis (VDA)* (2017). <https://doi.org/10.2352/ISSN.2470-1173.2017.1.VDA-387>
- Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: *ICML Workshop on Unsupervised and Transfer Learning*, pp. 17–36 (2012)
- Bernard, J.: *Exploratory Search in Time-Oriented Primary Data*. Dissertation, Ph.D. Technische Universität Darmstadt, Graphisch-Interaktive Systeme (GRIS), Darmstadt (2015). <http://tuprints.ulb.tu-darmstadt.de/5173/>
- Bellet, A., Habrard, A., Sebban M.: A Survey on Metric Learning for Feature Vectors and Structured Data. *CoRR* [arXiv:1306.6709](https://arxiv.org/abs/1306.6709) (2013)
- Bernard, J., Hutter, M., Zeppelzauer, M., Fellner, D., Sedlmair, M.: Comparing visual-interactive labeling with active learning: an experimental study. *IEEE Trans. Vis. Comput. Graph. (TVCG)* (2017). <https://doi.org/10.1109/TVCG.2017.2744818>
- Buhrmester, M., Kwang, T., Gosling, S.D.: Amazon’s mechanical turk: a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* **6**(1), 3–5 (2011)
- Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T.: State-of-the-art of visualization for eye tracking data. In: *EuroVis (STAR)* (2014), Eurograph. <https://doi.org/10.2312/eurovisstar.20141173>
- Behrisch, M., Korkmaz, F., Shao, L., Schreck, T.: Feedback-driven interactive exploration of large multidimensional data supported by visual classifier. In: *IEEE Visual Analytics Science and Technology (VAST)*, pp. 43–52 (2014)

12. Brown, E.T., Liu, J., Brodley, C.E., Chang, R.: Dis-function: Learning distance functions interactively. In: IEEE Visual Analytics Science and Technology (VAST), pp. 83–92. IEEE (2012)
13. Bernard, J., Ruppert, T., Goroll, O., May, T., Kohlhammer, J.: Visual-interactive preprocessing of time series data. In: SIGRAD, Swedish Chapter of Eurographics, vol. 81 of Linköping Electronic Conference Proceedings, Linköping University Electronic Press, pp. 39–48 (2012). http://www.ep.liu.se/ecp_article/index.en.aspx?issue=081;article=006
14. Bernard, J., Ruppert, T., Scherer, M., Schreck, T., Kohlhammer, J.: Guided discovery of interesting relationships between time series clusters and metadata properties. In: Knowledge Management and Knowledge Technologies (i-KNOW), pp. 22:1–22:8. ACM (2012). <https://doi.org/10.1145/2362456.2362485>
15. Bernard, J., Ritter, C., Sessler, D., Zeppelzauer, M., Kohlhammer, J., Fellner, D.: Visual-interactive similarity search for complex objects by example of soccer player analysis. In: IVAPP, VISIGRAPP, pp. 75–87 (2017). <https://doi.org/10.5220/0006116400750087>
16. Bernard, J., Sessler, D., Berisch, M., Hutter, M., Schreck, T., Kohlhammer, J.: Towards a user-defined visual-interactive definition of similarity functions for mixed data. In: IEEE Visual Analytics Science and Technology (Poster Paper) (2014). <https://doi.org/10.1109/VAST.2014.7042503>
17. Bernard, J., Sessler, D., Bannach, A., May, T., Kohlhammer, J.: A visual active learning system for the assessment of patient well-being in prostate cancer research. In: VIS Workshop on Visual Analytics in Healthcare, pp. 1–8. ACM (2015). <https://doi.org/10.1145/2836034.2836035>
18. Bernard, J., Sessler, D., Ruppert, T., Davey, J., Kuijper, A., Kohlhammer, J.: User-based visual-interactive similarity definition for mixed data objects-concept and first implementation. *J. WSCG* **22**, 329–338 (2014)
19. Baeza-Yates, R.A., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, Longman (1999)
20. Bernard, J., Zeppelzauer, M., Sedlmair, M., Aigner, W.: A unified process for visual-interactive labeling. In: Sedlmair, M., Tominiski, C. (eds.) EuroVis Workshop on Visual Analytics (EuroVA), Eurographics (2017). <https://doi.org/10.2312/eurova.20171123>
21. Chen, M., Golan, A.: What may visualization processes optimize? *IEEE Trans. Vis. Comput. Graph. (TVCG)* **22**(12), 2619–2632 (2016). <https://doi.org/10.1109/TVCG.2015.2513410>
22. Card, S.K., Mackinlay, J.D., Shneiderman, B. (eds.): Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, San Francisco (1999)
23. Choo, J., Park, H.: Customizing computational methods for visual analytics with big data. *IEEE Comput. Graph. Appl. (CG&A)* **33**(4), 22–28 (2013)
24. Craik, K. (ed.): The Nature of Explanation. Cambridge University Press, Cambridge (1943)
25. Card, S.K., Robertson, G.G., Mackinlay, J.D.: The information visualizer, an information workspace. In: SIGCHI Conference on Human Factors in Computing Systems (CHI). ACM, pp. 181–186 (1991). <https://doi.org/10.1145/108844.108874>
26. Chapelle, O., Schölkopf, B., Zien, A.: Semi-Supervised Learning. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, MA (2006)
27. Dagli, C.K., Rajaram, S., Huang, T.S.: Leveraging active learning for relevance feedback using an information theoretic diversity measure. In: Conference on Image and Video Retrieval, pp. 123–132. Springer, Berlin (2006). https://doi.org/10.1007/11788034_13
28. Elmqvist, N., Fekete, J.-D.: Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **16**(3), 439–454 (2010). <https://doi.org/10.1109/TVCG.2009.84>
29. Endert, A., Fiaux, P., North, C.: Semantic interaction for sense-making: inferring analytical reasoning for model steering. *IEEE Trans. Vis. Comput. Graph.* **18**(12), 2879–2888 (2012). <https://doi.org/10.1109/TVCG.2012.260>
30. Endert, A., Fiaux, P., North, C.: Semantic interaction for visual text analytics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 473–482. ACM, New York (2012). <https://doi.org/10.1145/2207676.2207741>
31. Endert, A., Han, C., Maiti, D., House, L., Leman, S., North, C.: Observation-level interaction with statistical models for visual analytics. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 121–130 (2011). <https://doi.org/10.1109/VAST.2011.6102449>
32. Endert, A., Ribarsky, W., Turkay, C., Wong, B.W., Nabney, I., Blanco, I.D., Rossi, F.: The state of the art in integrating machine learning into visual analytics. In: Computer Graphics Forum (CGF) (2017). <https://doi.org/10.1111/cgf.13092>
33. Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C.D., Roberts, J.C.: Visual comparison for information visualization. *Inf. Vis.* **10**(4), 289–309 (2011). <https://doi.org/10.1177/1473871611416549>
34. Grabner, H., Bischof, H.: On-line boosting and vision. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 1, pp. 260–267. IEEE (2006)
35. Gschwandtner, T., Gärtner, J., Aigner, W., Miksch, S.: A Taxonomy of Dirty Time-Oriented Data, pp. 58–72. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-32498-7_5
36. Gleicher, M.: A framework for considering comprehensibility in modeling. *Big Data* **4**(2), 75–88 (2016). <https://doi.org/10.1089/big.2016.0007>
37. Hoi, S.C., Jin, R., Lyu, M.R.: Large-scale text categorization by batch mode active learning. In: World Wide Web. ACM, pp. 633–642 (2006). <https://doi.org/10.1145/1135777.1135870.3>
38. Heimerl, F., Koch, S., Bosch, H., Ertl, T.: Visual classifier training for text document retrieval. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **18**(12), 2839–2848 (2012)
39. Höferlin, B., Netzel, R., Höferlin, M., Weiskopf, D., Heidemann, G.: Inter-active learning of ad-hoc classifiers for video visual analytics. In: IEEE Visual Analytics Science and Technology (VAST). IEEE, pp. 23–32 (2012). <https://doi.org/10.1109/VAST.2012.6400492>
40. Janetzko, H., Sacha, D., Stein, M., Schreck, T., Keim, D.A., Deussen, O.: Feature-driven visual analytics of soccer data. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 13–22 (2014). <https://doi.org/10.1109/VAST.2014.7042477>
41. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual Analytics: Definition, Process, and Challenges, pp. 154–175. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-70956-5_7
42. Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N.H., Weaver, C., Lee, B., Brodbeck, D., Buono, P.: Research directions in data wrangling: visualizations and transformations for usable and credible data. *Inf. Vis.* **10**(4), 271–288 (2011). <https://doi.org/10.1177/1473871611415994>
43. Karpinski, M., Macintyre, A.: Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks. *J. Comput. Syst. Sci.* **54**(1), 169–176 (1997). <https://doi.org/10.1006/jcss.1997.1477>
44. Krause, J., Perer, A., Bertini, E.: Infuse: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **20**(12), 1614–1623 (2014). <https://doi.org/10.1109/TVCG.2014.2346482>
45. Lewis, J.M., Ackerman, M., de Sa, V.R.: Human cluster evaluation and formal quality measures: a comparative study. In: Annual

- Meeting of the Cognitive Science Society (CogSci), pp. 1870–1875 (2012)
46. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
 47. Losing, V., Hammer, B., Wersing, H.: Incremental on-line learning: a review and comparison of state of the art algorithms. *Neurocomputing* **275**, 1261–1274 (2017). <https://doi.org/10.1016/j.neucom.2017.06.084>
 48. Liu, T.-Y.: Learning to rank for information retrieval. *Found. Trends Inf. Retr.* **3**(3), 225–331 (2009). <https://doi.org/10.1561/15000000016>
 49. Liu, Z., Stasko, J.: Mental models, visual reasoning and interaction in information visualization: a top-down perspective. *IEEE Trans. Vis. Comput. Graph.* **16**(6), 999–1008 (2010). <https://doi.org/10.1109/TVCG.2010.177>
 50. Mamitsuka, N.A.H.: Query learning strategies using boosting and bagging. In: Shavlik, J.W. (ed.) *International Conference on Machine Learning (ICML)*, vol. 1, pp. 1–9. Morgan Kaufmann, Los Altos (1998)
 51. Möhrmann, J., Bernstein, S., Schlegel, T., Werner, G., Heide-mann, G.: Improving the usability of interfaces for the interactive semi-automatic labeling of large image data sets. In: Jacko, J.A. (ed.) *Human-Computer Interaction. Design and Development Approaches*, pp. 618–627. Springer, Berlin (2011)
 52. Mamani, G.M.H., Fatore, F.M., Nonato, L.G., Paulovich, F.V.: User-driven feature space transformation. *Comput. Graph. Forum (CGF)* **32**(3), 291–299 (2013). <https://doi.org/10.1111/cgf.12116>
 53. Mühlbacher, T., Piringer, H.: A partition-based framework for building and validating regression models. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **19**(12), 1962–1971 (2013). <https://doi.org/10.1109/TVCG.2013.125>
 54. Mühlbacher, T., Piringer, H., Gratzl, S., Sedlmair, M., Streit, M.: Opening the black box: strategies for increased user involvement in existing algorithm implementations. *IEEE Trans. Vis. Comput. Graph.* **20**(12), 1643–1652 (2014)
 55. Mitrović, D., Zeppelzauer, M., Breiteneder, C.: Features for content-based audio retrieval. *Adv. Comput.* **78**, 71–150 (2010)
 56. Norman, D.A.: *The Design of Everyday Things*, reprint, paperback edn. Basic Books, New York (2002)
 57. Olsson, F.: *A Literature Survey of Active Machine Learning in the Context of Natural Language Processing*, Technical report. Swedish Institute of Computer Science (2009)
 58. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
 59. Qi, G.-J., Hua, X.-S., Rui, Y., Tang, J., Zhang, H.-J.: Two-dimensional multilabel active learning with an efficient online adaptation model for image classification. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **31**(10), 1880–1897 (2009). <https://doi.org/10.1109/TPAMI.2008.218>
 60. Rauber, P.E., Fadel, S.G., Falcao, A.X., Telea, A.C.: Visualizing the hidden activity of artificial neural networks. *IEEE Trans. Vis. Comput. Graph.* **23**(1), 101–110 (2017)
 61. Riek, L.D., O'Scannon, M.F., Robinson, P.: Guess what? a game for affective annotation of video using crowd sourcing. In: *International Conference on Affective Computing and Intelligent Interaction*, pp. 277–285. Springer, Berlin (2011)
 62. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**(1), 157–173 (2008)
 63. Sedlmair, M., Aupetit, M.: Data-driven evaluation of visual quality measures. *Comput. Graph. Forum (CGF)* **34**(3), 201–210 (2015). <https://doi.org/10.1111/cgf.12632>
 64. Shurkhovetsky, G., Andrienko, N., Andrienko, G., Fuchs, G.: Data abstraction for visualizing large time series. *Comput. Graph. Forum (CGF)* (2017). <https://doi.org/10.1111/cgf.13237>
 65. Seifert, C., Aamir, A., Balagopalan, A., Jain, D., Sharma, A., Grottel, S., Gumhold, S.: *Visualizations of Deep Neural Networks in Computer Vision: A Survey*, pp. 123–144. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54024-5_6
 66. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. *Read. Inf. Retr.* **24**, 5 (1997). [https://doi.org/10.1002/\(SICI\)1097-4571\(199006\)41:4<288::AID-ASIS>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-4571(199006)41:4<288::AID-ASIS>3.0.CO;2-H)
 67. Sessler, D., Bernard, J., Kuijper, A., Kohlhammer, J.: Adopting Mental Similarity Notions of Categorical Data Objects to Algorithmic Similarity functions. (2014). Poster Paper. <http://www.vmv2014.gcc.tu-darmstadt.de/sites/program.html>
 68. Schreck, T., Bernard, J., Von Landesberger, T., Kohlhammer, J.: Visual cluster analysis of trajectory data with interactive kohonen maps. *Inf. Vis.* **8**(1), 14–29 (2009). <https://doi.org/10.1057/ivs.2008.29>
 69. Settles, B., Craven, M.: An analysis of active learning strategies for sequence labeling tasks. In: *Empirical Methods in Natural Language Processing, Computational Linguistics*, pp. 1070–1079 (2008)
 70. Settles, B., Craven, M., Ray, S.: Multiple-instance active learning. In: *Advances in Neural Information Processing Systems*, pp. 1289–1296 (2008)
 71. Settles, B.: *Active Learning Literature Survey*, Technical Report 1648. University of Wisconsin–Madison (2009)
 72. Settles, B.: Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP), Computational Linguistics*, pp. 1467–1478 (2011)
 73. Settles, B.: Active learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **6**(1), 1–114 (2012)
 74. Seifert, C., Granitzer, M.: User-based active learning. In: *IEEE International Conference on Data Mining Workshops*, pp. 418–425 (2010). <https://doi.org/10.1109/ICDMW.2010.181>
 75. Stasko, J., Görg, C., Liu, Z.: Jigsaw: supporting investigative analysis through interactive visualization. *Inf. Vis.* **7**(2), 118–132 (2008). <https://doi.org/10.1145/1466620.1466622>
 76. Sedlmair, M., Heinzl, C., Bruckner, S., Piringer, H., Möller, T.: Visual parameter space analysis: a conceptual framework. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **20**(12), 2161–2170 (2014). <https://doi.org/10.1109/TVCG.2014.2346321>
 77. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
 78. Sedlmair, M., Meyer, M., Munzner, T.: Design study methodology: reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **18**(12), 2431–2440 (2012). <https://doi.org/10.1109/TVCG.2012.213>
 79. Seung, H.S., Oppor, M., Sompolinsky, H.: Query by committee. In: *Workshop on Computational Learning Theory (COLT)*, pp. 287–294. ACM, New York (1992). <https://doi.org/10.1145/130385.130417>
 80. Stolper, C.D., Perer, A., Gotz, D.: Progressive visual analytics: user-driven visual exploration of in-progress analytics. *IEEE Trans. Vis. Comput. Graph.* **20**(12), 1653–1662 (2014)
 81. Sarkar, A., Spott, M., Blackwell, A.F., Jamnik, M.: Visual discovery and model-driven explanation of time series patterns. In: *Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, pp. 78–86 (2016). <https://doi.org/10.1109/VLHCC.2016.7739668>
 82. Seebacher, D., Stein, M., Janetzko, H., Keim, D.A.: Patent retrieval: a multi-modal visual analytics approach. In: *EuroVis Workshop on Visual Analytics (EuroVA), Eurographics*, pp. 013–017 (2016)

83. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G.P., Keim, D.A.: Knowledge generation model for visual analytics. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **20**(12), 1604–1613 (2014). <https://doi.org/10.1109/TVCG.2014.2346481>
84. Sacha, D., Sedlmair, M., Zhang, L., Lee, J.A., Weiskopf, D., North, S.C., Keim, D.A.: Human-centered machine learning through interactive visualization: review and open challenges. In: *Artificial Neural Networks, Computational Intelligence and Machine Learning* (2016)
85. Sacha, D., Sedlmair, M., Zhang, L., Lee, J.A., Peltonen, J., Weiskopf, D., North, S.C., Keim, D.A.: What you see is what you can change: human-centered machine learning by interactive visualization. *Neurocomputing* (2017). <https://doi.org/10.1016/j.neucom.2017.01.105>. ISSN = 0925-2312
86. Sacha, D., Zhang, L., Sedlmair, M., Lee, J.A., Peltonen, J., Weiskopf, D., North, S.C., Keim, D.A.: Visual interaction with dimensionality reduction: a structured literature analysis. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **23**(01), 241–250 (2016). <https://doi.org/10.1109/TVCG.2016.2598495>
87. Turkay, C., Kaya, E., Balcisoy, S., Hauser, H.: Designing progressive and interactive analytics processes for high-dimensional data analysis. *IEEE Trans. Vis. Comput. Graph. (TVCG)* **23**(1), 131–140 (2017)
88. Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J.: A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J. Sel. Top. Signal Process.* **5**(3), 606–617 (2011)
89. Von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: *Conference on Human Factors in Computing Systems (SIGCHI)*, pp. 319–326. ACM (2004)
90. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, Berlin (2013)
91. van der Corput, P., van Wijk, J.J.: Comparing personal image collections with picturevis. *Comput. Graph. Forum (CGF)* **36**(3), 295–304 (2017). <https://doi.org/10.1111/cgf.13188>
92. van den Elzen, S., van Wijk, J.J.: Baobabview: interactive construction and analysis of decision trees. In: *IEEE Visual Analytics Science and Technology (VAST)*, pp. 151–160 (2011). <https://doi.org/10.1109/VAST.2011.6102453>
93. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
94. Vendrig, J., Patras, I., Snoek, C., Worring, M., den Hartog, J., Raaijmakers, S., van Rest, J., van Leeuwen, D.A.: Trec feature extraction by active learning. In: *TREC* (2002)
95. Visentini, I., Snidaro, L., Foresti, G.L.: On-line boosted cascade for object detection. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4. IEEE (2008)
96. van Wijk, J.J.: The value of visualization. In: *VIS 05. IEEE Visualization, 2005*, pp. 79–86 (2005). <https://doi.org/10.1109/VISUAL.2005.1532781>
97. Wall, E., Das, S., Chawla, R., Kalidindi, B., Brown, E.T., Endert, A.: Podium: ranking data using mixed-initiative visual analytics. *IEEE Trans. Vis. Comput. Graph.* **24**(1), 288–297 (2018)
98. Wang, M., Hua, X.-S.: Active learning in multimedia annotation and retrieval: a survey. *CMTrans. Intell. Syst. Technol.* **2**(2), 10:1–10:21 (2011). <https://doi.org/10.1145/1899412.1899414>
99. Wu, Y., Kozintsev, I., Bouguet, J.-Y., Dulong, C.: Sampling strategies for active learning in personal photo retrieval. In: *IEEE International Conference on Multimedia and Expo. IEEE*, pp. 529–532 (2006). <https://doi.org/10.1109/ICME.2006.262442>
100. Wenskovich, J., North, C.: Observation-level interaction with clustering and dimension reduction algorithms. In: *Workshop on Human-In-the-Loop Data Analytics (HILDA)*. ACM, pp. 14:1–14:6 (2017). <https://doi.org/10.1145/3077257.3077259>
101. Wongsuphasawat, K., Smilkov, D., Wexler, J., Wilson, J., Mané, D., Fritz, D., Krishnan, D., Viégas, F.B., Wattenberg, M.: Visualizing dataflow graphs of deep learning models in tensorflow. *IEEE Trans. Vis. Comput. Graph.* **24**(1), 1–12 (2018). <https://doi.org/10.1109/TVCG.2017.2744878>
102. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 27, pp. 3320–3328. Curran Associates Inc, New York (2014)
103. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H.: Understanding Neural Networks Through Deep Visualization (2015). arXiv preprint [arXiv:1506.06579](https://arxiv.org/abs/1506.06579)
104. Yang, L., Jin, R.: Distance metric learning: a comprehensive survey. *Mich. State Univ.* **2**, 2 (2006)
105. Zeiler, M.D., Fergus, R.: Visualizing and Understanding Convolutional Networks, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
106. Zhu, Q., Keogh, E.J.: Using captchas to index cultural artifacts. In: *International Symposium on Advances in Intelligent Data Analysis IX*, pp. 245–257. Springer, Berlin (2010)



Jürgen Bernard is postdoc at TU Darmstadt, Germany. He was with Fraunhofer IGD when he received his Ph.D. degree in 2015. His research is in information visualization and visual analytics with an emphasis on data mining and machine learning. His data-centered focus is on time-oriented data and high-dimensional data. Many of his solutions combine techniques from cluster analysis, dimensionality reduction, or similarity search.



Matthias Zeppelzauer is a senior researcher at the Institute of Creative Media Technologies at St. Pölten University of Applied Sciences (Austria). He received his Ph.D. in Computer Science from Vienna University of Technology in 2011 with highest distinction. His research focuses on multimedia signal processing and machine learning. He is an active member of the research community and acts as a reviewer for numerous high-impact conferences and journals. Matthias received several performance scholarships from the Vienna University of Technology and was awarded by the Austrian Computer Society for outstanding achievements in the area of pattern recognition.



Michael Sedlmair received his Ph.D. degree in Computer Science from the Ludwig Maximilians University of Munich. After a postdoc at the University of British Columbia, he is now an assistant professor at the University of Vienna. His research interests include visualization design and evaluation, human factors in visualization and data analysis, and human–computer interaction.



Wolfgang Aigner is the Scientific Director of the Institute of Creative\Media/Technologies at St. Poelten University of Applied Sciences (Austria) and adjunct professor at TU Wien (Austria). He received his Ph.D. degree and habilitation in computer science from TU Wien in 2006 and 2013. His main research interests include visual analytics and information visualization with a particular focus on time-oriented data.