Visual Informatics 1 (2017) 48–56

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Towards better analysis of machine learning models: A visual analytics perspective

Shixia Liu *, Xiting Wang, Mengchen Liu, Jun Zhu

Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Available online 18 January 2017

Keywords:
Interactive model analysis
Interactive visualization
Machine learning
Understanding
Diagnosis
Refinement

ABSTRACT

Interactive model analysis, the process of understanding, diagnosing, and refining a machine learning model with the help of interactive visualization, is very important for users to efficiently solve real-world artificial intelligence and data mining problems. Dramatic advances in big data analytics have led to a wide variety of interactive model analysis tasks. In this paper, we present a comprehensive analysis and interpretation of this rapidly developing area. Specifically, we classify the relevant work into three categories: understanding, diagnosis, and refinement. Each category is exemplified by recent influential work. Possible future research opportunities are also explored and discussed.

1. Introduction

Machine learning has been successfully applied to a wide variety of fields ranging from information retrieval, data mining, and speech recognition, to computer graphics, visualization, and human–computer interaction. However, most users often treat a machine learning model as a black box because of its incomprehensible functions and unclear working mechanism (Fekete, 2013; Liu et al., 2017; Mühlbacher et al., 2014). Without a clear understanding of how and why a model works, the development of high-performance models typically relies on a time-consuming trial-and-error process. As a result, academic researchers and industrial practitioners are facing challenges that demand more transparent and explainable systems for better understanding and analyzing machine learning models, especially their inner working mechanisms.

To tackle the aforementioned challenges, there are some initial efforts on interactive model analysis. These efforts have shown that interactive visualization plays a critical role in understanding and analyzing a variety of machine learning models. Recently, DARPA I2O released Explainable Artificial Intelligence (XAI) (DAR, 2016) to encourage research on this topic. The main goal of XAI is to create a suite of machine learning techniques that produce explainable models to enable users to understand, trust, and manage the emerging generation of Artificial Intelligence (AI) systems.

In this paper, we first provide an overview of interactive model analysis. Then we summarize recent interactive model analysis techniques based on their target tasks (such as understanding how a classifier works) (Heimerl et al., 2012). Research opportunities and future directions are discussed for developing new interactive model analysis techniques and systems.

2. Scope and overview

We are focused on research and application problems within the context of machine learning. Fig. 1 illustrates a typical machine learning pipeline, from which we first obtain data. Then we extract features that are usable as input to a machine learning model. Next, the model is trained, tested, and gradually refined based on the evaluation results and experience of machine learning experts, a process that is both time consuming and uncertain in building a reliable model. In addition to an explosion of research on better understanding of learning results (Cui et al., 2011, 2014; Dou et al., 2013; Dou and Liu, 2016; Liu et al., 2012, 2014a, 2016c; Wang et al., 2013, 2016), researchers have paid increasing attention to leveraging interactive visualizations to better understand and iteratively improve a machine learning model. The main goal of such research is to reduce human effort when training a reliable and accurate model. We refer to the aforementioned iterative and progressive process as interactive model analysis.

Fig. 2 illustrates the basic idea of interactive model analysis, where machine learning models are seamlessly integrated with state-of-the-art interactive visualization techniques capable of translating models into understandable and useful explanations for an expert. The strategy is to pursue a variety of visual analytics

* Corresponding author.
E-mail address: shixia@tsinghua.edu.cn (S. Liu).
Peer review under responsibility of Zhejiang University and Zhejiang University Press.

http://dx.doi.org/10.1016/j.visinf.2017.01.006

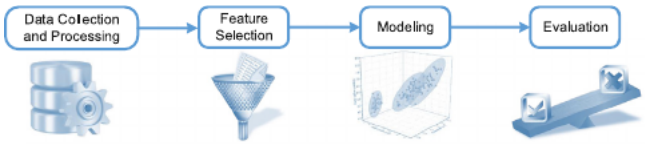

Fig. 1. A pipeline of machine learning.

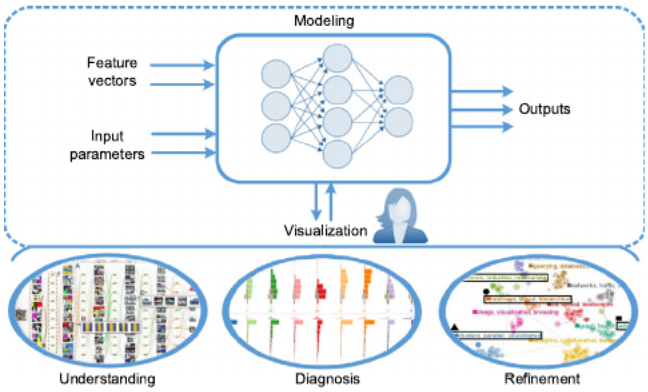

Fig. 2. An overview of interactive model analysis.

techniques in order to help experts understand, diagnose, and refine a machine learning model. Accordingly, interactive model analysis aims to create a suite of visual analytics techniques that

- understand why machine learning models behave the way they do and why they differ from each other (*understanding*);
- diagnose a training process that fails to converge or does not achieve an acceptable performance (*diagnosis*);
- guide experts to improve the performance and robustness of machine learning models (*refinement*).

3. Discussion and analysis of existing work

Most recent efforts in interactive model analysis aim to help machine learning experts understand how the model works, such as the interactions between each component in the model. More recently, there have been some initial attempts to diagnose a training process that failed to converge or did not achieve the desired performance, or to refine the learning model for better performance.

3.1. Understanding

Many techniques have been developed to help experts better understand classification models (Paiva et al., 2015; Turner, 2016; Tzeng and Ma, 2005) and regression models (Zahavy et al., 2016). Among all models, neural networks have received the most attention. They have been widely used and achieved state-of-the-art results in many machine learning tasks, such as image classification and video classification (LeCun et al., 2015). To better understand the working mechanism of neural networks, researchers have developed various visualization approaches, which can be classified into two categories: point-based and network-based.

Point-based techniques (Zahavy et al., 2016; Rauber et al., 2017) reveal the relationships between neural network components, such as neurons or learned representations, by using scatterplots. Each learned representation is a high-dimensional vector whose entries are the output values of neurons in one hidden layer. Typically, each component is represented by a point. Components with similar roles are placed adjacent to each other by using dimension reduction techniques such as Principal Component Analysis (PCA) (Wold et al., 1987) and t-SNE (Maaten and Hinton, 2008). Point-based techniques facilitate the confirmation of hypothesis on neural networks and the identification of previously unknown relationships between neural network components (Rauber et al., 2017).

Fig. 3 shows a point-based visualization developed by Rauber et al. (2017). In this figure, each point denotes the learned representation of a test sample. The color of each point encodes the class label of each test sample. As shown in the figure, after training, the visual separation between classes is significantly improved. This observation provides evidence for the hypothesis that neural networks learn to detect representations that are useful for class discrimination. Fig. 3(b) also helps with the understanding of misclassified samples, which are marked by triangle glyphs. The figure illustrates that many misclassified samples are visual outliers whose neighbors have different classes. Also, many outliers correspond to test samples that are difficult for even humans to classify. For example, an image of digit 3 is misclassified because it is very similar to some images of digit 5.

Although point-based techniques are useful for presenting the relationships between a large number of neural network components, they cannot reveal the topological information of the

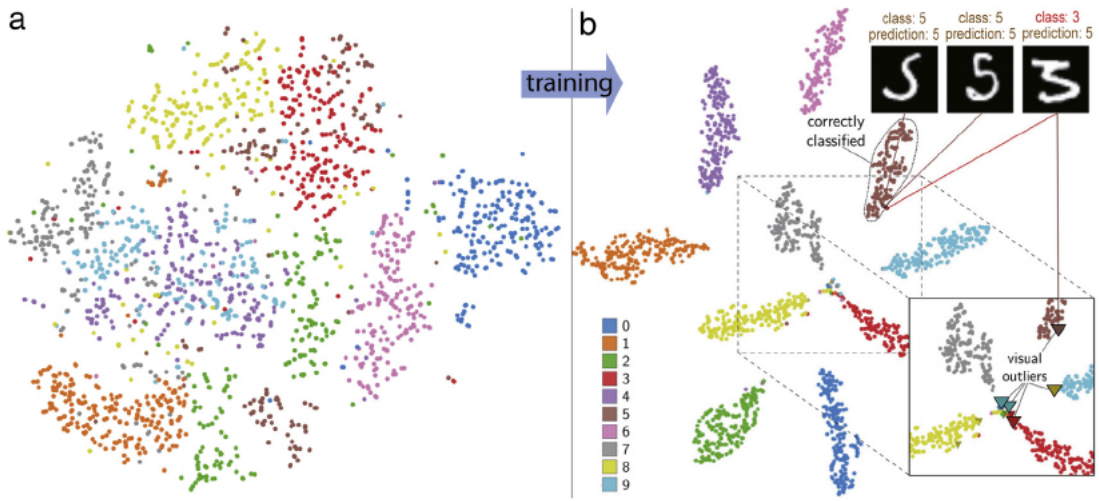

Fig. 3. Comparison of test sample representations (a) before and (b) after training (Rauber et al., 2016).

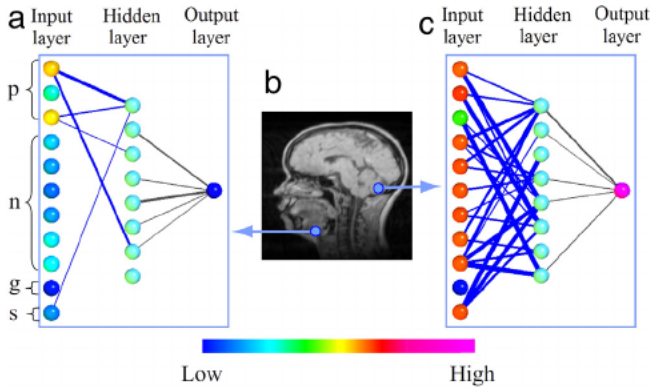

Fig. 4. Topology of a neural network trained to classify brain and non-brain materials (Tzeng and Ma, 2005).

networks. As a result, they fail to provide a comprehensive understanding of the roles of different neurons in different layers and the interactions between them. Network-based techniques (Harley, 2015; Streeter et al., 2001; Craven and Shavlik, 1992) solve this problem by displaying the network topology. These techniques usually represent a neural network as a directed acyclic graph (DAG) and encode important information from the network by the size, color, and glyphs of the nodes or edges in the DAG.

Fig. 4 shows the visualization generated by a pioneer network based technique (Tzeng and Ma, 2005). This figure presents a neural network trained to classify whether a voxel within the head belongs to the brain or not. Here, each voxel is represented by its scalar value s, gradient magnitude g, scalar values of its neighbors n, and its position p. The width of each edge encodes the importance of the corresponding connection. The nodes in the input and output layers are colored based on their output values. The color of the node in the output layer indicates that the neural network is able to correctly classify the voxel on the left to non-brain materials (low output value) and the voxel on the right to brain materials (high output value). The network topologies in Fig. 4(a) and Fig. 4(c) demonstrate that the voxel on the left is classified to non-brain materials mainly because of its position, while the voxel on the right needs all inputs except for the gradient magnitude g to be correctly classified to brain materials.

The aforementioned technique can effectively visualize neural networks with several dozens of neurons. However, as the number of neurons and connections increase, the visualization may become cluttered and difficult to understand (Tzeng and Ma, 2005). To solve this problem, Liu et al. (2017) developed CNNVis, a visual analytics system that helps machine learning experts understand and diagnose deep convolutional neural networks (CNNs) with thousands of neurons and millions of connections (Fig. 5). To display large CNNs, the layers and neurons are clustered. A representative layer (neuron) is selected for each layer (neuron) cluster. To effectively display many connections, a biclustering-based algorithm is used to bundle the edges and reduce visual clutter. Moreover, CNNVis supports the analysis of multiple facets of each neuron. To this end, CNNVis visualizes the learned features of each neuron cluster by using a hierarchical rectangle packing algorithm. A matrix reordering algorithm was also developed to reveal the activation patterns of neurons.

3.2. Diagnosis

Researchers have developed visual analytics techniques that diagnose model performance for binary classifiers (Amershi et al., 2015), multi-class classifiers (Liu et al., 2017; Alsallakh et al., 2014; Ren et al., 2017), and topic models (Chuang et al., 2013). The goal of these techniques is to help machine learning experts understand why a training process did not achieve a desirable performance so that they can make better choices (e.g., select better features) to improve the model performance. To this end, current techniques utilize the prediction score distributions of the model (i.e., sample-class probability) to evaluate the error severity and study how the

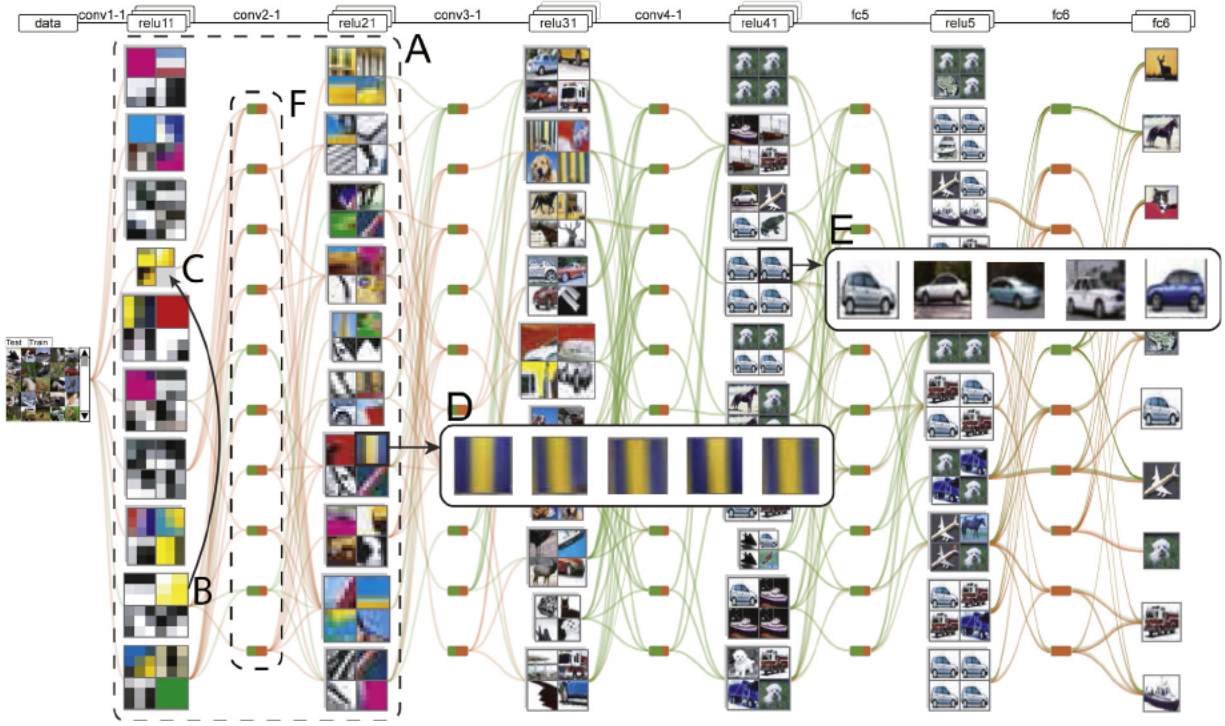

Fig. 5. CNNVis, a visual analytics approach to understanding and diagnosing deep convolutional neural networks (CNNs) (Liu et al., 2016b) with a large number of neurons and connections.

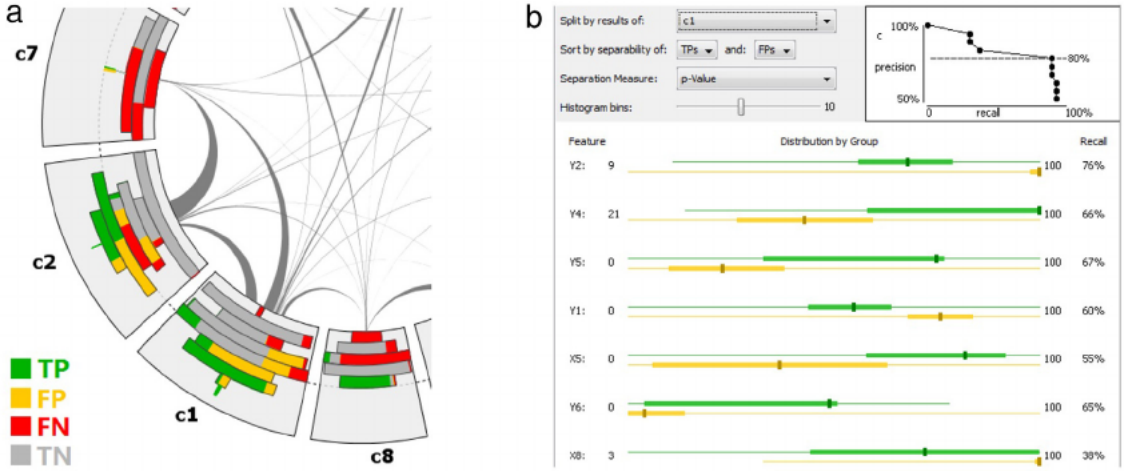

Fig. 6. A visual analytics tool that helps machine learning experts diagnose model performance with (a) a confusion wheel and (b) a feature analysis view (Alsallakh et al., 2014).

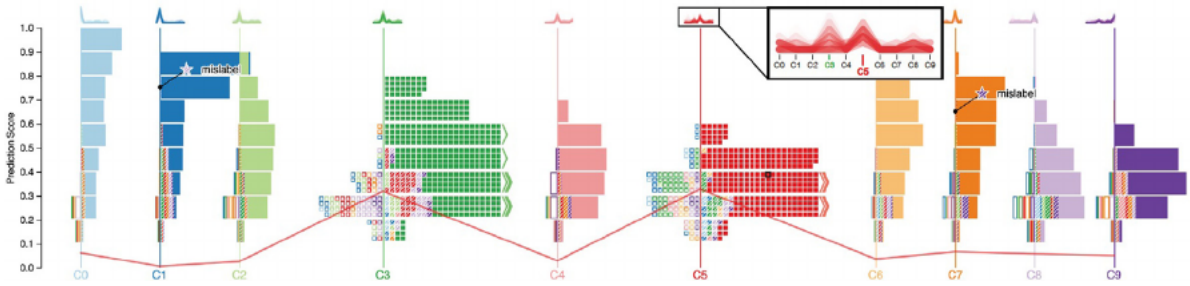

Fig. 7. Squares, a visual analytics tool that supports performance diagnosis of multi-class classifiers within a single visualization to reduce the cognitive load of users during analysis (Ren et al., 2016).

score distributions correlate with misclassification and selected features.

One typical example is the model performance diagnosis tool developed by Alsallakh et al. (2014). This tool consists of a confusion wheel (Fig. 6(a)) and a feature analysis view (Fig. 6(b)). The confusion wheel depicts the prediction score distributions by using histograms. For each class c_i, bins that correspond to samples with low (high) prediction scores of c_i are placed close to the inner (outer) ring. The chords in the confusion wheel visualize the number of samples that belong to class c_i misclassified to class c_j (between-class confusion). This view enables users to quickly identify the samples that are misclassified with a low probability (e.g., the false-negative samples (FNs) in c_7). These samples are easier to improve compared with other samples. The feature analysis view illustrates how two groups of samples (e.g., true-positive samples and false-positive samples) can be separated by using certain features. This view helps users to make better choices in terms of feature selection.

Although the aforementioned technique provides valuable guidance for performance improvement, the confusion wheel can introduce distortion by displaying histograms in a radial display. Researchers also point out that multiple coordinated visualizations may add complexity to the diagnosis process (Ren et al., 2017). To eliminate the distortion and reduce the cognitive load of users, Ren et al. proposed Squares (Ren et al., 2017), a visual analytics tool that supports performance diagnosis within a single visualization. As shown in Fig. 7, Squares is able to show prediction score distributions at multiple levels of detail. The classes, when expanded to show the lowest level of detail (e.g., c_3 and c_5), are displayed as boxes. Each box represents a (training or test) sample. The color of the box encodes the class label of the corresponding sample and the texture represents whether a sample is classified correctly (solid fill) or not (striped fill). The classes with the least number of details (e.g., c_0 and c_1) are displayed as stacks. Squares also allows machine learning experts to explore between-class confusion (see polylines in Fig. 7) within the same visualization.

More recently, there have been some initial efforts on diagnosing deep learning models (Liu et al., 2017; Zahavy et al., 2016). One example is CNNVis (Liu et al., 2016b) (Fig. 5). By revealing multiple facets of the neurons, the interactions between neurons, and relative weight changes between layers, CNNVis allows machine learning experts to debug a training process that fails to converge or does not achieve an acceptable performance. It also helps to find potential directions to prevent the training process from getting stuck or improve the model performance. Another example is the method developed by Zahavy et al. (2016), which employs t-SNE to disclose relationships between learned representations and uses saliency maps to help users analyze influential features. Case studies on three ATARI games demonstrate the ability of this method to find problems that pertain to game modeling, initial and terminal state modeling, and score over-fitting.

3.3. Refinement

After they gain an understanding of how machine learning models behave and why they do not achieve a desirable performance, machine learning experts usually wish to refine the model by incorporating the knowledge learned. To facilitate this process, researchers have developed visual analytics systems that provide interaction capabilities for improving the performance of supervised (Paiva et al., 2015) or unsupervised models (Wang et al., 2016; Liu et al., 2016).

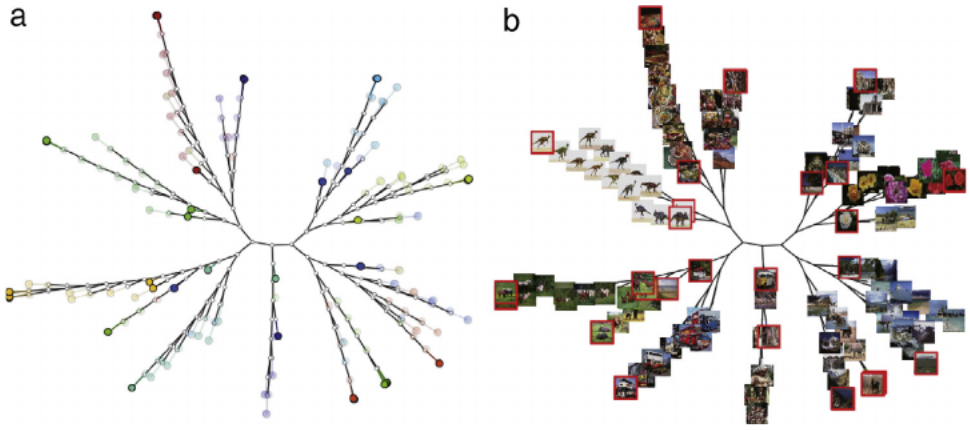

Fig. 8. Interactive training sample selection that enables classifier refinement (Paiva et al., 2015). Candidate samples are represented by (a) circles and (b) images.

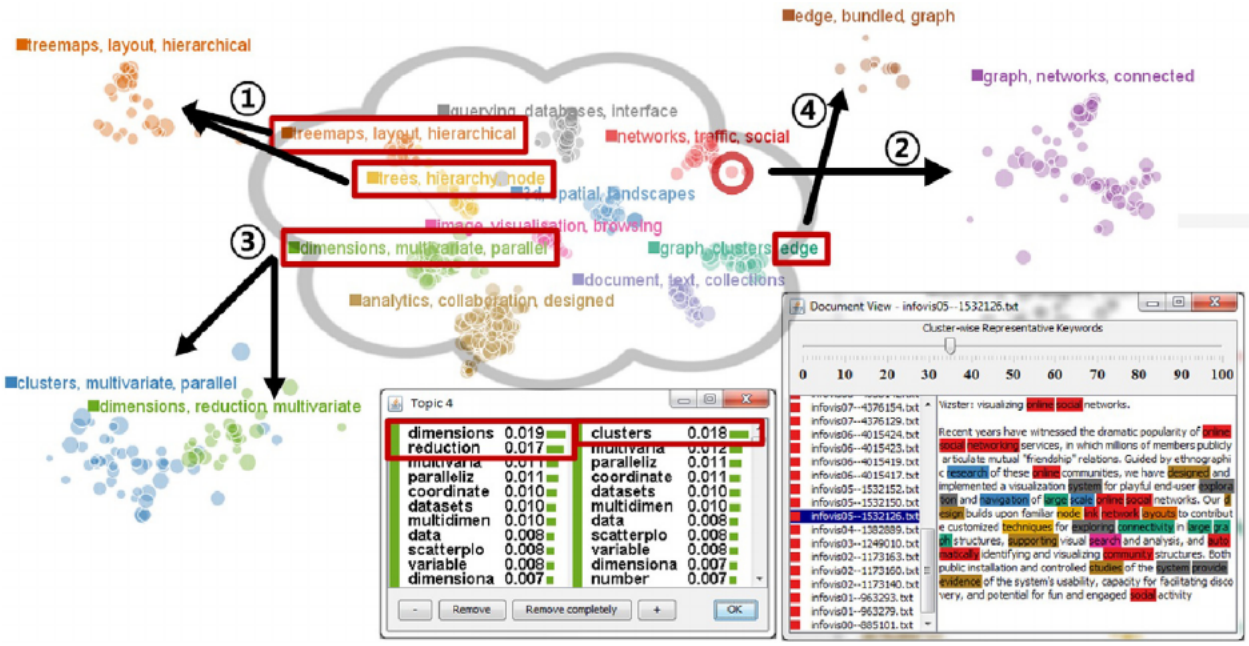

Fig. 9. UTOPIAN (Choo et al., 2013), a visual analytics system for interactive refinement of topic models.

Current techniques for refining supervised models mainly focus on multi-class classifiers (Paiva et al., 2015; Alsallakh et al., 2014). These techniques allow users to insert their knowledge by controlling factors that significantly affect classification results. Commonly considered factors include training samples, features, types of classifiers, and parameters used in training. For example, the technique developed by Paiva et al. (2015) allows users to interactively select training samples, modify their labels, incrementally update the model, and rebuild the model by using new classes. Fig. 8 shows how this technique supports informed training sample selection. Here, each sample is displayed as a point in Fig. 8(a) and an image in Fig. 8(b). These samples are organized by using Neighbor Joining trees (Paiva et al., 2011). After observing the trees, the user carefully selected 43 samples from the core of the tree and the end of the branches. Training with these samples generates a classifier with an accuracy of 97.43%.

The techniques for refining unsupervised models usually incorporate user knowledge into the model in a semi-supervised manner (Tzeng and Ma, 2004; Choo et al., 2013; Liu et al., 2014b). A typical example in this field is UTOPIAN (Choo et al., 2013), a visual analytics system for refining topic model results. In UTOPIAN, the topics are initially learned using Nonnegative Matrix Factorization (NMF) (Lee and Seung, 1999) and the learned topics are displayed using a scatterplot visualization. As shown in Fig. 9, UTOPIAN allows users to interactively (1) merge topics, (2) create topics based on exemplar documents, (3) split topics, and (4) create topics based on keywords. Moreover, UTOPIAN also supports topic keyword refinement. All these interactions are centered around a semi-supervised formulation of NMF that enables an easy incorporation of user knowledge and an incremental update of the topic model.

There are also some refinement tools that aim to help business professionals who are not familiar with complex machine learning models. For example, Wang et al. developed a visual analytics system, TopicPanorama (Wang et al., 2016; Liu et al., 2014b), to help business professionals analyze and refine a full picture of relevant topics discussed in multiple textual sources. The full picture is generated by matching the topic graphs extracted from different sources with a scalable algorithm to learn correlated topic models (Chen et al., 2013). TopicPanorama allows users to identify potentially incorrect matches by examining the uncertainties

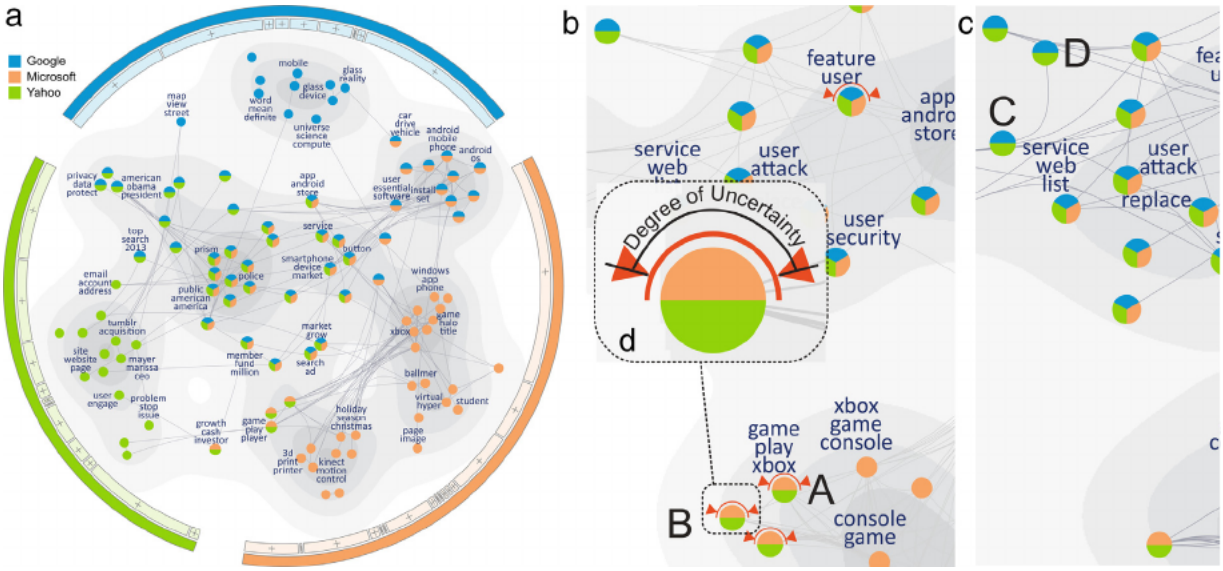

Fig. 10. TopicPanorama (Wang et al., 2016), a visual analytics system for analyzing a full picture of relevant topics from multiple sources: (a) Panorama visualization, (b) a matching result with two incorrect matches **A** and **B**, (c) the updated matching result with corrected matches **C** and **D**, and (d) an uncertainty glyph.

of the matches. Moreover, by incorporating metric learning and feature selection into the graph matching model, TopicPanorama allows users to incrementally improve and refine the matching model.

Fig. 10(a) shows a full picture of the topics related to three IT companies: Google, Microsoft, and Yahoo. Here, the topic nodes of different companies (sources) are represented with different colors and the common topics are encoded in a pie chart. A public relations manager cared about game related topics, so she enabled the uncertainty glyphs (Fig. 10(d)) to examine potential incorrect matches. After some exploration, she identified two incorrect matches, **A** and **B**, that match Microsoft Xbox games to Yahoo sport games (Fig. 10(b)). After she unmatched **B**, she found **A** was changed to **C** and **B** was changed to **D**, which correctly matched Google sport games to Yahoo sport games (Fig. 10(c)).

Another example is MutualRanker (Liu et al., 2016), a visual analytics tool to retrieve salient posts, users, and hashtags. To effectively retrieve salient posts, users and hashtags, they built a mutual reinforcement graph (MRG) model (Wei et al., 2008) that jointly considers the content quality of posts, the social influence of users, and the popularity of hashtags. They also analyzed the uncertainty in the results. Based on the retrieved data and the uncertainty, they developed a composite visualization that visually illustrates the posts, users, hashtags, their relationships, and the uncertainty in the results. With this visualization, business professionals are able to easily detect the most uncertain results and interactively refine the MRG model. To efficiently refine the model, they developed a random-walk-based Monte Carlo sampling method that can locally update the model based on user interactions. A typical use case of MutualRanker is shown in Fig. 11, where an expert found that the cluster "nationalparks" shared the uncertainty propagated from the "shutdown", "democrats", and "republicans" cluster. This indicates there is high uncertainty in the ranking scores of the hashtags in the "nationalparks" cluster. According to his domain knowledge, the expert increased the ranking scores of "#nationalparks" in that cluster and the ranking scores of other relevant hashtags were automatically updated.

4. Research opportunities

We regard existing methods as an initial step and there are many research opportunities to be further explored and pursued, which will be discussed in the following subsection in terms of technical challenges and future research.

4.1. Creating explainable models

Although machine learning models are widely used in many applications, they often fail to explain their decisions and actions to users. Without a clear understanding, it may be hard for users to incorporate their knowledge into the learning process and achieve a better learning performance (e.g., prediction accuracy). As a result, it is desirable to develop more explainable machine learning models, which have the ability to explain their rationale and convey an understanding of how they behave in the learning process. The key challenge here is to design an explanation mechanism that is tightly integrated into the machine learning model.

Accordingly, one interesting future work is to discover which part(s) in the model structure explains its different functions and play a major role in the performance improvement or decline of each iteration. Another interesting venue for future work is to better illustrate the rationale behind the model and the decisions made. Recently, there have been some initial efforts in this direction (Letham et al., 2015; Lake et al., 2015). For example, Lake et al. (2015) developed a probabilistic program induction algorithm. They built simple stochastic programs to represent concepts, building them compositionally from parts, subparts, and spatial relations. They also demonstrated that their algorithm achieved human-level performance on a one-shot classification task, while outperforming recent deep learning approaches. However, for the tasks that have abundant training data, such as object and speech recognition, the less explainable deep learning still outperforms the algorithm. Thus, there is still a long way to go for researchers to develop more explainable models for these tasks.

4.2. Analysis of online training process

Most of the existing methods focus on analyzing the final results (Ren et al., 2017) or one snapshot (Liu et al., 2017) of the model in the interactive training process. In many cases, only analyzing the results or a single snapshot is not enough to understand why a training process did not achieve a desirable performance. Thus, it is necessary to analyze the online training process.

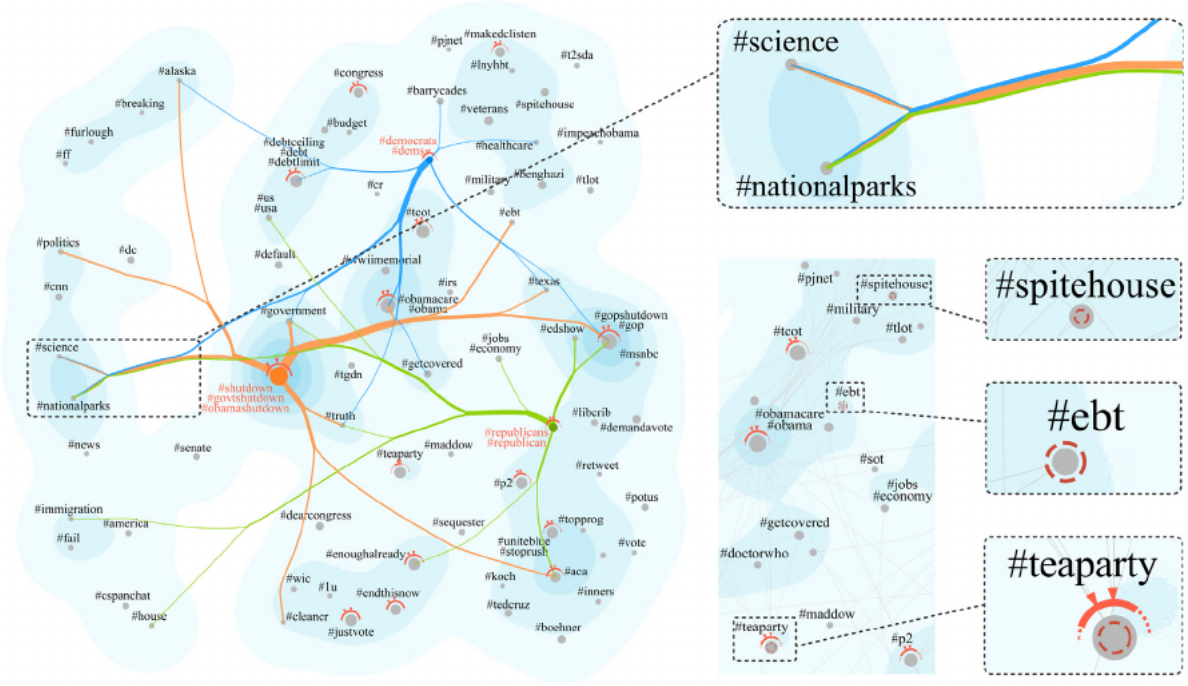

Fig. 11. MutualRanker (Liu et al., 2016a), a visual analytics toolkit to retrieve salient posts, users, and hashtags. MutualRanker enables interactive refinement of uncertain results.

One challenge in analyzing the online training process is the difficulty of selecting and comparing representative snapshots from a large number of snapshots. When comparing different snapshots, one possible solution is to adopt progressive visual analytics (Stolper et al., 2014) to shorten the period of time between user interactions and the execution of the model. The basic idea of progressive visual analytics is to produce meaningful partial results during the training process and integrating these partial results into an interactive visualization, which allows users to immediately explore the partial results.

Another challenge is automatically and accurately detecting anomalies in the training process. Currently, the training process is sometimes too long (e.g., more than one week for an expert to supervise the whole training process of a large deep neural network (Krizhevsky et al., 2012)). In these scenarios, it is necessary to automatically detect anomalies and timely notify the expert. Automatic and accurate identification of anomalies is still a challenging research topic (Tam et al., 2017). Thus, it is desirable to employ an interactive visualization, which can better combine the human ability to detect anomalies and the power of machines to process large amounts of data, which has been initially studied in some recent work (Cao et al., 2016; Zhao et al., 2014).

4.3. Mixed initiative guidance

To improve the performance of machine learning models and better incorporate the knowledge of experts, researchers have developed a set of guidance techniques. Such efforts have arisen from two main research communities: machine learning and information visualization. From the machine learning community, researchers have developed a wide array of techniques for system initiated guidance (Settles, 2012; Cohn et al., 1994, 1996; McCallumzy and Nigamy, 1998), where the system plays a more active role, for example, by making suggestions about appropriate views or next steps in the iterative and progressive analysis process. From the information visualization community, researchers have designed a number of techniques for user initiative guidance (Wang et al., 2016; Liu et al., 2016; Choo et al., 2013; Liu et al., 2014b; Pezzotti et al., 2016), where the user is the active participant in improving and refining the performance and learning results.

In many tasks, it is preferable to combine system imitative guidance and user initiative guidance as mixed initiative guidance to maximize the value of both. Accordingly, mixed initiative guidance is defined as a type of visual reasoning or feedback process in which the human analyst and the machine learning system can both actively foster the guidance to improve the machine learning model. Although mixed initiative guidance is very useful, supporting it is technically demanding. There are two major challenges that we need to address.

First, it is not easy to seamlessly integrate system initiative guidance and user initiative guidance in one unified framework. System initiative guidance is usually based on the learning process and the evaluation of the results, while user initiative guidance is typically based on the experience and domain knowledge of the expert. Accordingly, we need to study how to define an efficient working mechanism to integrate them and support smooth communication between them. For example, one interesting research problem is how to reveal the provenance of system initiative guidance to illustrate why a suggestion is made by the system. Then, based on this, the expert can better understand the rationale behind the suggestion and provide his/her feedback accordingly. Another potential research problem is to effectively extract appropriate and sufficient user/system data to create a unified model for both the user and the system.

Second, there may be several conflicts between system initiative guidance and user initiative guidance in real-world applications. For example, for a given training sample, the system and the user may have different opinions on which class it belongs to. As a result, how to solve these conflicts is an interesting research problem that needs further exploration.

4.4. Uncertainty

While visual analytics is very useful in helping machine learning experts gain insights into the working mechanisms of models and devise ways to improve model performance, it may also introduce

uncertainties into the analysis process. It has been shown that uncertainty awareness positively influences human trust building and decision making (Sacha et al., 2016). Thus, it is important to quantify and analyze uncertainties (Correa et al., 2009; Wu et al., 2010) in interactive model analysis, which is challenging for a number of reasons.

First, uncertainties may originate from each stage of the interactive model analysis process (e.g., training, visualization, refinement) and increase, decrease, split, and merge during the whole process (Wu et al., 2012). Thus, it is difficult to effectively quantify the uncertainties. One interesting direction for future research is to develop visual analytics techniques that effectively measure and quantify the uncertainty in data processing, model building, and visualization (Sacha et al., 2016) and help experts quickly identify the potential issues in a machine learning model of interest.

Second, it is challenging to model different types of uncertainties as well as their interactions by using a unified framework. During the interactive model analysis process, there are uncertainties that originate from the machine side (e.g., imperfect machine learning models) and uncertainties that originate from the human side (e.g., incorrect expert feedback). These two kinds of uncertainties will interact with and influence each other. For example, if the system presents misleading information to the experts, they may return incorrect feedback that results in problematic modification of the model. Another example is that allowing experts to view and refine results of many test samples may encourage overfitting (Ren et al., 2017). Accordingly, an interesting research problem is how to model different types of uncertainties and their interactions with a unified model.

References

Alsallakh, B., Hanbury, A., Hauser, H., Miksch, S., Rauber, A., 2014. Visual methods for analyzing probabilistic classification data. IEEE TVCG 20 (12), 1703–1712. http://dx.doi.org/10.1109/TVCG.2014.2346660.

Amershi, S., Chickering, M., Drucker, S.M., Lee, B., Simard, P., Suh, J. 2015. Modeltracker: Redesigning performance analysis tools for machine learning. In: CHI, pp. 337–346. http://dx.doi.org/10.1145/2702123.2702509.

Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y.R., Lin, C.Y., 2016. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. IEEE TVCG 22 (1), 280–289.

Chen, J., Zhu, J., Wang, Z., Zheng, X., Zhang, B. 2013. Scalable inference for logistic-normal topic models. In: NIPS.

Choo, J., Lee, C., Reddy, C.K., Park, H., 2013. Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. IEEE TVCG 19 (12) 1992–2001. http://dx.doi.org/10.1109/TVCG.2013.212.

Chuang, J., Gupta, S., Manning, C.D., Heer, J. 2013. Topic model diagnostics: Assessing domain relevance via topical alignment. In: ICML, pp. 612–620.

Cohn, D., Atlas, L., Ladner, R., 1994. Improving generalization with active learning. Mach. Learn. 15 (2), 201–221.

Cohn, D.A., Ghahramani, Z., Jordan, M.I., 1996. Active learning with statistical models. JAIR 4 (1), 129–145.

Correa, C.D., Chan, Y.H., Ma, K.L. 2009. A framework for uncertainty-aware visual analytics. In: VAST, pp. 51–58.

Craven, M.W., Shavlik, J.W., 1992. Visualizing learning and computation in artificial neural networks. Int. J. Artif. Intell. Tools 1 (03), 399–425.

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., Qu, H., Tong, X., 2011. TextFlow: Towards better understanding of evolving topics in text. IEEE TVCG 17 (12), 2412–2421.

Cui, W., Liu, S., Wu, Z., Wei, H., 2014. How hierarchical topics evolve in large text corpora. IEEE TVCG 20 (12), 2281–2290.

DAR, 2016. Explainable Artificial Intelligence (XAI). http://www.darpa.mil/program/explainable-artificial-intelligence.

Dou, W., Liu, S., 2016. Topic-and time-oriented visual text analysis. IEEE Comput. Graph. Appl. 36 (4), 8–13.

Dou, W., Yu, L., Wang, X., Ma, Z., Ribarsky, W., 2013. HierarchicalTopics: Visually exploring large text collections using topic hierarchies. IEEE TVCG 19 (12), 2002–2011.

Fekete, J.-D., 2013. Visual analytics infrastructures: From data management to exploration. Computer 46 (7), 22–29.

Harley, A.W., 2015. An interactive node-link visualization of convolutional neural networks. In: International Symposium on Visual Computing. Springer, pp. 867–877.

Heimerl, F., Koch, S., Bosch, H., Ertl, T., 2012. Visual classifier training for text document retrieval. IEEE TVCG 18 (12), 2839–2848.

Krizhevsky, A., Sutskever, I., Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105.

Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B., 2015. Human-level concept learning through probabilistic program induction. Science 350 (6266), 1332–1338.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755), 788–791.

Letham, B., Rudin, C., McCormick, T.H., Madigan, D., 2015. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. Ann. Appl. Stat. 9 (3), 1350–1371.

Liu, S., Cui, W., Wu, Y., Liu, M., 2014a. A survey on information visualization: Recent advances and challenges. The Visual Computer 30 (12), 1373–1393.

Liu, M., Liu, S., Zhu, X., Liao, Q., Wei, F., Pan, S., 2016. An uncertainty-aware approach for exploratory microblog retrieval. IEEE TVCG 22 (1), 250–259. http://dx.doi.org/10.1109/TVCG.2015.2467554.

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J.J.H., Liu, S., 2017. Towards better analysis of deep convolutional neural networks. IEEE TVCG 23 (1), 91–100. http://dx.doi.org/10.1109/IEEETVCG.2016.2598831.

Liu, S., Wang, X., Chen, J., Zhu, J., Guo, B. 2014b. TopicPanorama: A full picture of relevant topics. In: VAST, pp. 183–192. http://dx.doi.org/10.1109/VAST.2014.7042494.

Liu, S., Yin, J., Wang, X., Cui, W., Cao, K., Pei, J., 2016c. Online visual analytics of text streams. IEEE TVCG 22 (11), 2451–2466. http://dx.doi.org/10.1109/IEEETVCG.2015.2509990.

Liu, S., Zhou, M.X., Pan, S., Song, Y., Qian, W., Cai, W., Lian, X., 2012. TIARA: Interactive, topic-based visual text summarization and analysis. ACM TIST 3 (2), 25:1–25:28.

Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (Nov), 2579–2605.

McCallumzy, A.K., Nigamy, K. 1998. Employing EM and pool-based active learning for text classification. In: ICML, pp. 359–367.

Mühlbacher, T., Piringer, H., Gratzl, S., Sedlmair, M., Streit, M., 2014. Opening the black box: Strategies for increased user involvement in existing algorithm implementations. IEEE TVCG 20 (12), 1643–1652. http://dx.doi.org/10.1109/IEEETVCG.2014.2346578.

Paiva, J.G., Florian, L., Pedrini, H., Telles, G., Minghim, R., 2011. Improved similarity trees and their application to visual data classification. IEEE TVCG 17 (12), 2459–2468.

Paiva, J.G.S., Schwartz, W.R., Pedrini, H., Minghim, R., 2015. An approach to supporting incremental visual data classification. IEEE TVCG 21 (1), 4–17. http://dx.doi.org/10.1109/TVCG.2014.2331979.

Pezzotti, N., Lelieveldt, B., Maaten, L.V.D., Hollt, T., Eisemann, E., Vilanova, A., 2016. Approximated and user steerable tSNE for progressive visual analytics. IEEE TVCG PP (99), 1–14. http://dx.doi.org/10.1109/TVCG.2016.2570755.

Rauber, P.E., Fadel, S., Falcao, A., Telea, A., 2017. Visualizing the hidden activity of artificial neural networks. IEEE TVCG 23 (1), 101–110. http://dx.doi.org/10.1109/TVCG.2016.2598838.

Ren, D., Amershi, S., Lee, B., Suh, J., Williams, J.D., 2017. Squares: Supporting interactive performance analysis for multiclass classifiers. IEEE TVCG 23 (1), 61–70. http://dx.doi.org/10.1109/TVCG.2016.2598828.

Sacha, D., Senaratne, H., Kwon, B.C., Ellis, G., Keim, D.A., 2016. The role of uncertainty, awareness, and trust in visual analytics. IEEE TVCG 22 (1), 240–249.

Settles, B., 2012. Active Learning. In: Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool.

Stolper, C.D., Perer, A., Gotz, D., 2014. Progressive visual analytics: User-driven visual exploration of in-progress analytics. IEEE TVCG 20 (12), 1653–1662. http://dx.doi.org/10.1109/TVCG.2014.2346574.

Streeter, M.J., Ward, M.O., Alvarez, S.A., 2001. Nvis: An interactive visualization tool for neural networks.

Tam, G.L., Kothari, V., Chen, M., 2017. An analysis of machine- and human-analytics in classification. IEEE TVCG 23 (1), 71–80. http://dx.doi.org/10.1109/TVCG.2016.2598829.

Turner, R. 2016. A model explanation system: Latest updates and extensions. In: ICML Workshop.

Tzeng, F.Y., Ma, K.L. 2004. Acluster-space visual interface for arbitrary dimensional classification of volume data. In: Sixth Joint Eurographics - IEEE TCVG Conference on Visualization, pp. 17–24. http://dx.doi.org/10.2312/VisSym/VisSym04/017-024.

Tzeng, F.Y., Ma, K.L. 2005. Opening the black box - data driven visualization of neural networks. In: IEEE Visualization, pp. 383–390. http://dx.doi.org/10.1109/VISUAL.2005.1532820.

Wang, X., Liu, S., Liu, J., Chen, J., Zhu, J., Guo, B., 2016. TopicPanorama: A full picture of relevant topics. IEEE TVCG 22 (12), 2508–2521. http://dx.doi.org/10.1109/TVCG.2016.2515592.

Wang, X., Liu, S., Song, Y., Guo, B. 2013. Mining evolutionary multi-branch trees from text streams. In: KDD, pp. 722–730.

Wei, F., Li, W., Lu, Q., He, Y. 2008. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In: SIGIR, pp. 283–290.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometr. Intell. Lab. Syst. 2 (1), 37–52.

Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., Qu, H., 2010. OpinionSeer: Interactive visualization of hotel customer feedback. IEEE TVCG 16 (6), 1109–1118.

Wu, Y., Yuan, G.X., Ma, K.L., 2012. Visualizing flow of uncertainty through analytical processes. IEEE TVCG 18 (12), 2526–2535.

Zahavy, T., Ben-Zrihem, N., Mannor, S. 2016. Graying the black box: Understanding dqns. In: ICML pp. 1899–1908.

Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y.R., Collins, C., 2014. #FluxFlow: Visual analysis of anomalous information spreading on social media. IEEE TVCG 20 (12), 1773–1782.