How "Applied" is Fifteen Years of VAST conference?

Lei Shi[1]*, Lei Xia[1], Zipeng Liu[1], Ye Sun[1], Huijie Guo[1], Klaus Mueller[2]
[1] Beihang University, Beijing, China
[2] Stony Brook University, New York, United States

Abstract

Visual analytics (VA) science and technology emerge as a promising methodology in visualization and data science in the new century. Application-driven research continues to contribute significantly to the development of VA, as well as in a broader scope of VIS. However, existing studies on the trend and impact of VA/VIS application research stay at a commentary and subjective level, using methods such as panel discussions and expert interviews. On the contrary, this work presents a first study on VA application research using data-driven methodology with cutting-edge machine learning algorithms, achieving both objective and scalable goals. Experiment results demonstrate the validity of our method with high F1 scores up to 0.89 for the inference of VA application papers on both the expert-labeled benchmark dataset and two external validation data sources. Inference results on 15 years of VAST conference papers also narrate interesting patterns in VA application research's origin, trend, and constitution.

Index Terms: Human-centered computing—Visualization

1 Introduction

Visual analytics (VA) is coined as "the science combining the automated analytical method with interactive visualization" [31] [23]. The annual VAST conference [1] (2006∼2020) has been the premier venue to present the latest VA techniques. Among the 15 years of VAST conference, there is an essential type of research form, namely application papers, that develop and/or apply VA techniques to specific application domains. Representative domains include finance [10], urban computing [8], sports [32], and many others.

The visualization community (VIS) has long debated the pros and cons of application papers. Lorensen, in his famous essay "On the death of visualization", called for more application research in VIS to draw external customers [25]. Weber et al., on a panel of VIS'16, echoed the proposal of Lorensen (apply or die!) and discussed various issues focusing on how to appropriately assess the value of application papers [33]. However, in the latest VIS'22 paper, Wu et al. mentioned the growing concerns about the technical rigorousness and contribution validity of VAST application papers [34]. Despite the controversial opinions, the development of VA application research follows an upward spiral curve (Fig. 2), and the community generally agrees on many of its positive aspects. First, application papers tell success stories of VA/VIS techniques and promote the image of the visualization community in areas outside. Second, many funding opportunities come from application-driven missions and institutions, e.g., DOE/NIH and private sectors. Third but not least, VA applications do have the potential to drive fundamental VIS research if appropriately reflected [27].

Although existing debates [25] [33] [34] have covered a broad spectrum of the merits of VA application research, there is still a moderate gap in the quantitative understanding of its development. Lorensen's essay was composed in a commentary style [25], and the report from the VIS'16 panel summarized viewpoints made by a group of VIS researchers [33]. Both could be affected by subjective, personal biases. Wu et al.'s work [34], the only systematic study on VA application research till now, adopted an interview-based methodology that still depends on participants' subjective feedback. This work roots in a data-driven, quantitative, and objective investigation of VAST paper content, where cutting-edge analytics methods can be applied to mine facts and patterns of the VA application research. Specially, we adopt machine learning (ML) based classifiers to infer VAST application papers for its scalability of extension to the whole VIS community (InfoVis, SciVis, etc., merged into VIS in 2021).

Accomplishing the work here requires overcoming multifaceted challenges. In the community, comprehensive datasets on VIS literature as well as their periphery information, have been well established, e.g., VisPubData [21], VIS30K [11], VisImages [16]. However, none of these datasets is built to understand the value of VA applications. The work by Wu et al. exemplified the concept of VA system/application research [34], but only with a short list of 32 papers at VIS'21. No formal criterion for VA application research is given. The VIS submission website also provides keyword categories corresponding to VAST application papers [22], but they need to be more accurate and comprehensive, according to our investigation. In this work, we make the following contribution to tackle the challenges above and present a first step toward understanding VA application research from a data science perspective.

- We propose an objective at best criterion to define VAST application papers, which is validated to be consistent by two visualization experts independently. The expert study also establishes a benchmark dataset for application research covering one-third of VAST papers of all time.
- An end-to-end machine learning pipeline is developed, which infers VAST application papers on our benchmark dataset with an F1 score of 0.89, using cross-validated evaluations. Experiments on external VAST application datasets demonstrate similarly high performance.
- The result applying our inference model reveals interesting patterns on the history, trend, and current position of VA application research. First, surprisingly VA research indeed starts from applications and steps into stable development in the latest decade. Second, VAST application papers do not differ much from other VAST technique papers in title, abstract, and keywords. As we expected, they do diverge from SciVis research and are closer to InfoVis research.

2 Background and Related Work

2.1 Application Research in Visualization and VA

Application research has long been promoted in the history of the visualization community. Lorensen, in his short essay [25], first gave a retrospect of the birth and peak of visualization research as the founding generation. He cautioned about a growing risk at that age – visualization research was moving away from its customers because of the need for application domain knowledge. Later on, the call has been partially met by more effort in major events, such as the IEEE Workshop on Visualization in Practice [20], IEEE VIS Application Spotlights [2], and mostly the birth of VAST symposium/conference series from 2006 [1]. VAST has been commonly believed to be more "applied" than parallel events such as SciVis and InfoVis.

*leishi@buaa.edu.cn

As the VIS community comes to maturity, the understanding of the value of application research also deepens. In the panel of VIS'16 [33], a group of researchers discussed the pros and cons of advocating more application papers. Notably, the evaluation of application papers in VIS [12] remains an open challenge. These papers are often attacked because of the unclear contribution to VIS, lack of general knowledge learned, and weak reproducibility [17]. In the latest interview study of Wu et al. [34], similar criticisms were collected from review comments of VA application/system papers. Defenses to these systems summarized from the author's feedback and suggestions were made to promote VA system research, e.g., constructing knowledge bases to derive general knowledge [15]. Despite abundant literature discussing VIS/VA application research, no data-driven study is based on existing VIS paper collections.

2.2 Bibliographic Data Sources for VIS Community

As a community active for reflection [27], high-quality publication datasets become vital. Although general-purpose academic datasets such as Microsoft Academic Graph (MAG) [3], DBLP [6], AMiner [5] have been well provisioned, extracting academic data corpus for the visualization community is still challenging due to the issue of data quality and the wide variety of paper formats. Fortunately, Isenberg et al. have created and continued to maintain VisPubData [21], an open dataset containing all-time publications presented at the VIS conference series, up to the latest year. Both paper metadata and citations within the conference are included.

Other focused bibliographic data sources are also available in the community. VIS30K [11] collects the full set of figures and tables from VIS papers, by applying cutting-edge deep learning algorithms. VisImages [16] also prepares an image data set for visualization papers. Moreover, they extract visualization bounding boxes out of these images and conduct case studies on modeling and analysis of these visualizations. Liu et al. summarize and classify open visualization resources for community usage [24].

Upon publication datasets, many insightful analytics have been conducted. Isenberg et al. derive a set of visualization topics by mining keywords associated with each paper [22]. Hao et al. investigate the field of study, collaboration patterns, and citation flows among VIS venues and countries, based on a later version of MAG [19]. Shin et al. [30] and Shi et al. [29] study influence profiles of academic entities and research papers. None of the existing bibliometric studies on visualization literature focuses on the identification and analysis of application research as the work here.

3 Machine Learning on VAST Dataset

3.1 Data Source and Pre-processing

We start by collecting all published VAST papers from 2006 to 2020 as the notion of VAST ends after the transition into a new publication model in 2021 [4]. From the VisPubData dataset [21], we find 571 VAST papers, which are verified with the publisher website IEEE Xplore [7] using their DOIs. We obtain the metadata (including the title, year, abstract, and keywords) from VisPubData and the full text of these papers from IEEE Xplore.

To leverage the structural information of an academic paper, we extract the section titles from the full text using ParsCit [14] and the figure captions using the VIS30K dataset [11]. To leverage the topical information, we apply the BERTopic algorithm [18] on the title, abstract, and keywords of each paper, which forms 14 topics. From this topic model, we assign a topic vector of length 14 to each paper as its vector representation for downstream modeling.

3.2 Criterion for VAST Application Papers

There is no trivial way to define an application paper, although there are relevant efforts at the VAST conference. First, the submission process of a VAST paper normally requires input on paper types (technique & algorithm, system, application & design study, empirical study, theory & model, etc.), but there is no exclusive choice on whether or not a paper does apply research. Moreover, the choice of paper is a subjective judgment that could be incongruous among authors. Second, there are usually application tracks in the conference program, but it is, in fact, a tedious and error-prone job to collect this data and align it with our source dataset.

In this work, we propose a more objective criterion to define a VAST application paper compared to the status quo, i.e., the paper should present *visual analytics technique customized for one application domain only*. With this definition, the extension of VA techniques in an application paper from one domain to another is not straightforward or not heavily discussed. On the opposite side, VAST non-application papers normally present VA techniques or theories that are not limited or customized to one single application domain. A special note: the papers about VA for machine/deep learning are considered non-application papers unless they are only customized to a single domain. The proposed criterion is confirmed to be unambiguous by two VA experts in a collaborative study, and it is also validated via data-driven quantitative experiments. We note that while this application domain can be defined by a standard field of study hierarchy (e.g., those in MAG), throughout this work, we stick to the informal domain tag literally mentioned in the original paper. According to our goal to identify VAST application papers, mapping informal domains to standard hierarchy, though not impossible, is beyond the scope of this work.

3.3 Expert study

Two VA experts, who have published multiple papers in vis-related venues, participate in the expert study. 190 out of 571 VAST papers (33.3%) are sampled for the study, as shown in the bottom box of Fig. 1. The papers in more recent years are sampled with larger weights as we conjecture that the chance of application papers increases over time. The two experts manually and independently classify the 190 papers into application or non-application by the proposed criterion, given the full PDF files. To ensure a consistent classification standard, they begin with the same set of 30 papers as the first round. The results show that they disagree on two papers (6.7%) but can reach a consensus after a quick discussion. In the second round, the remaining sampled papers are divided into two random groups of 80 papers, each handled by one expert exclusively.

The expert study sets up a labeled dataset of 63 VAST application papers and 127 others. Initially, we speculate that the classification of VAST application papers can be achieved using primitive textual features: title, abstract, keywords, image captions, and section titles. However, the model trained on the labeled dataset with these vectorized features leads to a rather poor performance of $F1 < 0.4$ (10-fold cross-validation) in our best trial. The incompetence could be ascribed to the lack of differentiability of these features.

To overcome this problem, we step back from optimizing the machine learning models and reflect on the proposed criterion for VAST application papers. The core idea is to identify the specific domain each application paper focuses on. This domain entry, in case extractable from the paper content, could be a key hidden feature to identify application papers. Therefore, in the next round of expert study, each researcher labels at least one application domain from each VAST application paper in his/her group. Note that as the annotation task requires heavy cognitive effort from the expert, we develop an online paper annotation system to help accomplish the task. As shown in the attached video demo, the system supports the entire annotation workflow, including the exportation and comparison of annotation results. Finally, 119 application domain labels from 63 VAST application papers are identified.

After the study, upon careful examination and discussion, we find that all domain labels appear in the *title+abstract+keywords (TAK)* content of each paper. We hypothesize that it would suffice to only

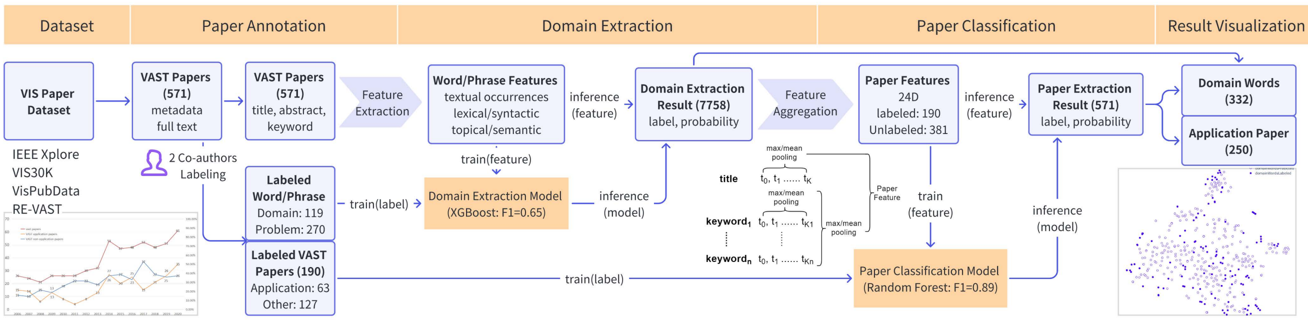

Figure 1: The pipeline of our end-to-end machine learning approach.

Table 1: The list of features used for the classification of VAST paper words/phrases into application domain and non-domain entries.

Category	Description	#Ft	Sig.
textual occurrences	# of occurrences in title	1	3.0e-03
	# of occurrences in keyword	1	5.1e-07
	# of occurrences in abstract	7	≤0.0159
lexical syntactic	part of speech type	3	≤0.0253
	dependency parsing type	13	≤0.0455
topical semantic	word/phrase topic probability	14	≤1.9e-08
	paper topic probability	4	≤0.0126
	(word, paper)-BERT-similarity	1	5.0e-05
	(word, paper)-topic-similarity	2	≤0.0405

look at the TAK content for paper classification. We then try to verify it by skimming the full-text content of the labeled papers, especially the result sections. Only 1 out of 190 papers disagrees: it has a case study deployed in another domain, which is absent from the TAK content. In fact, as VA researchers, we would have highlighted the cross-domain generalizability of the proposed techniques (if true) in the TAK to increase visibility. The TAK-dominant pattern is critical for our data-driven approach, as otherwise, it would require modeling the full text, a much more difficult task.

3.4 Model-based Paper Type Inference

We design an end-to-end machine learning pipeline to infer VAST application papers. As shown in Fig. 1, two models are trained and applied: one to extract domain words/phrases from each VAST paper, and another to classify VAST application papers using the extracted domain information. To build the first model, three types of features are computed for each word/phrase in the paper (Table 1): textual occurrences, lexical/syntactic attributes, and their topical/semantic information. To evaluate these features, Mann-Whitney U [26] and Brunner-Munzel [9] test are conducted between the labeled domain words/phrases and others. Finally, 46 significant features are selected. In the second model for paper classification, input features include the inferred probability for each word/phrase to be a domain and the derived binary domain label, duplicated twice in unigram/bigram settings. As shown in Fig. 1, these four groups of features are aggregated on the title and keyword sections, respectively, using a combination of average and max pooling strategy. Finally, 24 features are extracted for paper classification.

To further improve the inference performance, we exploit two optimization strategies. First, we discover that there is at least one domain word/phrase in the title+keywords content of each VAST application paper, which leads to a focused optimization strategy, that is, extracting domain entries from merely the title and keywords. It reduces the data imbalance ratio from 96.4 to 20.2, greatly boosting the domain extraction performance. The imbalance problem lies in the difficulty of labeling all words/phrases closely related to the

Table 2: Performance of the proposed machine learning pipeline, with alternatives on learning algorithm, word embedding, and the use of unlabeled dataset.

Model	Domain Extraction F1	Domain Extraction Precision	Domain Extraction Recall	Paper Classification F1	Paper Classification Precision	Paper Classification Recall
MLP	.42±.10	.60±.12	.35±.13	.50±.00	.34±.00	**.99**±.00
SVM	.61±.01	**.82**±.06	.49±.01	.65±.00	.79±.14	.58±.09
XGB	**.65**±.02	.70±.07	**.62**±.06	.84±.01	**.94**±.05	.76±.03
RF	.62±.01	.66±.07	.59±.05	**.89**±.00	.91±.02	.87±.01
Semi-	.62±.01	.67±.06	.58±.04	.56±.02	.70±.03	.47±.02
SciBert	.65±.02	.72±.04	.59±.02	.83±.01	.80±.02	.85±.02

domain of an application paper. This issue is resolved by detecting all synonyms of each labeled domain in the TAK content of the paper. Eventually, 96.8% of our labeled VAST application papers detect at least one domain word/phrase in title+keywords. Second, we apply the classical self-training semi-supervised learning method [28] to combine useful information from 381 unlabeled VAST papers.

We evaluate the proposed end-to-end machine learning pipeline on the labeled VAST dataset. We use stratified 10-fold cross-validation to accommodate the imbalanced data. Table 2 summarizes the inference performance using different embedding methods and learning models. For the inference of the VAST application paper class (minority), the optimal F1 score as high as 0.89 can be achieved with a random forest (RF) classifier and the standard BERT pre-training embeddings. The semi-supervised learning does not improve the overall performance, probably because the unlabeled VAST papers can not transfer useful information without intensive processing. For the domain extraction task, the best performance is achieved using the XGBoost classifier [13], without semi-supervised learning. Note that due to the high data imbalance rate between the labeled non-domain and domain entries (20.2), an F1 score of 0.65 on the minority class can be deemed excellent in practice.

We further collect two external validation datasets to avoid data cherry-picking as our labeled VAST dataset is manually selected and prepared. The first dataset is provided by the work of Wu et al. [34], who identify a list of 32 VIS'21 papers on VA systems and applications. The second dataset is obtained by mining the Visualization Paper Submission and Keyword Dataset shared by Isenberg [22]. We extract all the 34 accepted VAST papers with either primary or top-3 secondary keywords belonging to the high-level keyword category of VA applications. As mentioned in Sect. 3.2, the classification by submission keywords has inherent drawbacks, e.g., a low-level category called Situational Awareness could hardly be regarded as VA application by our definition. The two researchers in the expert study run a pass on these 34 candidate papers, and only 14 of them are identified as VAST application papers by the proposed criterion.

We perform the same machine learning pipeline with the best settings to the two external datasets, which achieves the recall metrics of 0.84 and 0.71 respectively. Note that we can only report the

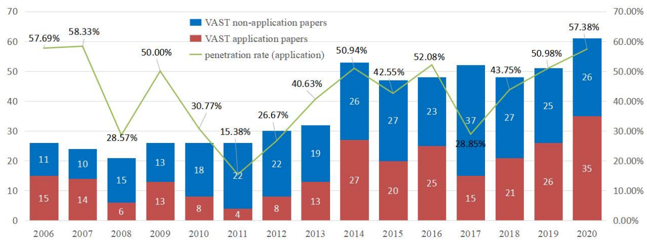

Figure 2: The dynamics of the number of VAST (non-)application papers and their penetration rates.

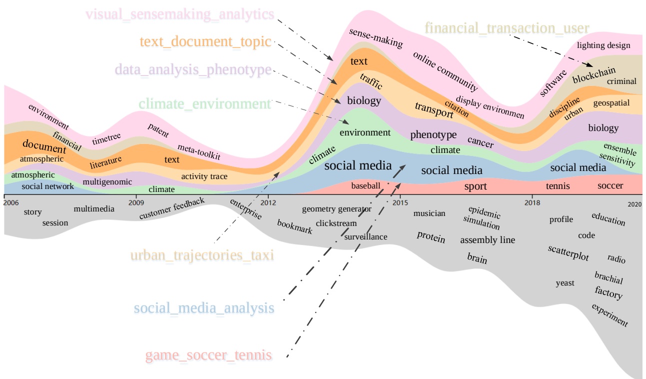

Figure 3: The stream graph showing 8 topics of 250 inferred VAST application papers, with the bottom grey topic a misc./background one. Keywords in each stream indicate time-sensitive topic content.

recalls as there are only positive VAST application papers available. In the less performant second dataset, 3 out of 4 misclassified papers, published before 2011, do not include any author-provided keywords in the metadata. Drilling down to more misclassification cases in our labeled dataset reveals that the inference of VAST application papers can not be completely precise and accurate. For example, some paper titles include rare words as specific visual analytics methods, which can be falsely predicted as application domains. They can be hard to avoid unless there are ways to integrate external knowledge about all application domains.

4 Results and Visualization

We apply the derived machine learning models to 571 VAST papers of all time, which identifies 250 VAST application papers, including the labeled ones. Fig. 2 summarizes the dynamics of the number and percentage of application papers in VAST annually. The number of VAST application papers peaks in 2020 and 2014, the same years with the peak of the total number of VAST papers. There is a surge of application papers in 2014, coming right after 2013 when all VAST papers are invited to publish at TVCG. Surprisingly, the highest penetration rates of application papers appear in both the early years of the VAST conference (2006, 2007) and the last year (2020). This implies that the VA research starts from an application-driven objective, contradicting our previous conjecture, and it also culminates with applications in the last year. From a birds-eye view, the popularity of VA application research follows an upward spiral curve, although with downturns in 2008/2011/2017, quite similar to the evolution of the community's viewpoint on the subject.

To explore the ups and downs of topics among the identified VAST application papers, we apply BERTopic [18] on their TAK content, which produces 8 topics, with one large topic containing all the application papers on infrequent topics/domains. Fig. 3 visualizes these topics in a stream graph where labeled/extracted domain words are placed to the corresponding topic stream in their most frequent time periods. Several interesting patterns can be observed, including the emergence of sport and blockchain applications recently, the rise of social media VA from 2012, and the relatively fading trend on

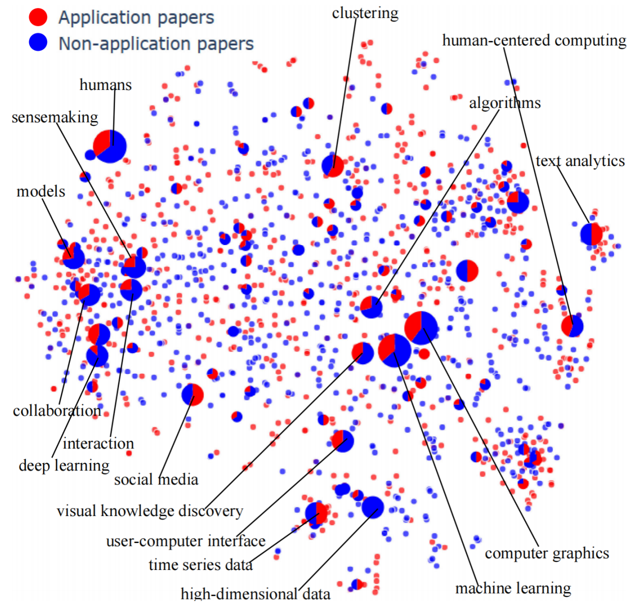

Figure 4: The projection of keyword distributions of VAST application and non-application papers.

text-related applications. We also notice the surge of a long tail of domains ever since 2010, as indicated by the grey topic beneath the horizon, which may imply the proliferation of the VA technique into new and more diversified application areas.

We analyze the differences between VAST app./non-app. papers on their author-provided keywords. Fig. 4 depicts the keyword distribution of the two paper classes using BERT and tSNE. Stop words in VA, such as "visual analytics/analysis", "(data/information) visualization", have been removed. No notable differences are observed. That is, the two classes intertwine in the projection. VAST application papers occupy a larger share than the average of 0.44 in keyword ratio on "clustering", "social media", and "text analytics"; while technique/theory papers lead in all other frequent keywords, especially "models", "interaction", "humans", and "deep learning".

5 Conclusion and Discussion

This work quantifies the portion of application research in 15 years of VAST conference, based on a high-quality paper dataset. We define an objective at best criterion for VAST application papers and build a labeled benchmark data through collaborative expert study. An end-to-end domain word and paper type inference pipeline is proposed, which combines elaborate feature extraction and engineering with state-of-the-art machine learning algorithms. An impressive F1 score of 0.89 is achieved in classifying the minor application class from labeled VAST papers. On 250 inferred VAST application papers, interesting or even surprising patterns on the origin and development of VA application research are identified and illustrated via multiple forms of visualization. We note that though the history of the VAST conference is not too long for full-scale labeling, the suite of techniques proposed here could be potentially extended to the entire VIS venues or even beyond.

There are several limitations in this work. First, each entry of labeled domain words/phrases needs to be adjacent in paper content due to the constraint of our preliminary annotation system. This may affect data quality though annotators have been asked to identify the most appropriate adjacent words as the domain entries. Second, the machine learning models are inherently stochastic, and thus the outcomes may fluctuate, though the result variance is shown to be small. More advanced models like GPT could reduce non-determinism. We plan to address these limitations in future work.

Acknowledgments

This work was supported by National Key R&D Program of China (2021YFB3500700), NSFC Grant 62172026, National Social Science Fund of China 22&ZD153, the Fundamental Research Funds for the Central Universities, and SKLSDE.

References

[1] IEEE Conference on Visual Analytics Science and Technology, 2006–2020. `http://vis.computer.org/visweek2010/vast/`.

[2] IEEE VIS Application Spotlights, 2019. `https://ieeevis.org/year/2019/info/application-spotlights`.

[3] Microsoft academic graph. https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/, 2021.

[4] New area model for ieee vis, 2021. `https://ieeevis.org/year/2021/info/call-participation/call-for-participation`.

[5] Aminer. https://www.aminer.org, 2023.

[6] Dblp. https://dblp.org/, 2023.

[7] IEEE Xplore, 2023. `https://ieeexplore.ieee.org`.

[8] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology*, pp. 51–58, 2008.

[9] E. Brunner and U. Munzel. The nonparametric behrens-fisher problem: asymptotic theory and a small-sample approximation. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 42(1):17–25, 2000.

[10] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Wirevis: Visualization of categorical, time-varying data from financial transactions. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology*, pp. 155–162, 2007.

[11] J. Chen, M. Ling, R. Li, P. Isenberg, T. Isenberg, M. Sedlmair, T. Möller, R. S. Laramee, H.-W. Shen, K. Wünsche, et al. Vis30k: A collection of figures and tables from ieee visualization conference publications. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3826–3833, 2021.

[12] M. Chen and D. S. Ebert. An ontological framework for supporting the design and evaluation of visual analytics systems. In *Computer Graphics Forum*, vol. 38, pp. 131–144, 2019.

[13] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *KDD'16*, pp. 785–794, 2016.

[14] I. G. Councill, C. L. Giles, and M.-Y. Kan. Parscit: an open-source crf reference string parsing package. In *LREC'08*, vol. 8, pp. 661–667, 2008.

[15] D. Deng, A. Wu, H. Li, J. Lan, Y. Wang, H. Qu, and Y. Wu. Kb4va: A knowledge base of visualization designs for visual analytics. *arXiv preprint arXiv:2211.02567*, 2022.

[16] D. Deng, Y. Wu, X. Shu, J. Wu, S. Fu, W. Cui, and Y. Wu. Visimages: A fine-grained expert-annotated visualization dataset. *IEEE Transactions on Visualization & Computer Graphics*, (01):1–1, 2022.

[17] J.-D. Fekete and J. Freire. Exploring reproducibility in visualization. *IEEE Computer Graphics and Applications*, 40(5):108–119, 2020.

[18] M. Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[19] H. Hao, Y. Cui, Z. Wang, and Y.-S. Kim. Thirty-two years of ieee vis: Authors, fields of study and citations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1016–1025, 2022.

[20] B. Hentschel, M. Meyer, H. Hagen, and R. Maciejewski. Applied visualization. *IEEE Computer Graphics and Applications*, 38(3):30–32, 2018.

[21] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. D. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A metadata collection about ieee visualization (vis) publications. *IEEE transactions on visualization and computer graphics*, 23(9):2199–2206, 2016.

[22] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller. Visualization as seen through its research paper keywords. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):771–780, 2016.

[23] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In *Information Visualization*, pp. 154–175. Springer, 2008.

[24] X. Liu, M. Alharbi, J. Best, J. Chen, A. Diehl, E. Firat, D. Rees, Q. Wang, and R. S. Laramee. Visualization resources: a starting point. In *International Conference Information Visualisation (IV)*, pp. 160–169, 2021.

[25] B. Lorensen. On the death of visualization. In *NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges*, vol. 1, 2004.

[26] P. E. McKnight and J. Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.

[27] M. Meyer and J. Dykes. Reflection on reflection in applied visualization research. *IEEE computer graphics and applications*, 38(6):9–16, 2018.

[28] C. Rosenberg, M. Hebert, and H. Schneiderman. Semi-supervised self-training of object detection models. 2005.

[29] L. Shi, H. Tong, J. Tang, and C. Lin. Vegas: Visual influence graph summarization on citation networks. *IEEE Transactions on Knowledge and Data Engineering*, 27(12):3417–3431, 2015.

[30] M. Shin, A. Soen, B. T. Readshaw, S. M. Blackburn, M. Whitelaw, and L. Xie. Influence flowers of academic entities. In *2019 IEEE conference on visual analytics science and technology (VAST)*, pp. 1–10, 2019.

[31] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

[32] J. Wang, K. Zhao, D. Deng, A. Cao, X. Xie, Z. Zhou, H. Zhang, and Y. Wu. Tac-simur: Tactic-based simulative visual analytics of table tennis. *IEEE transactions on visualization and computer graphics*, 26(1):407–417, 2019.

[33] G. H. Weber, S. Carpendale, D. Ebert, B. Fisher, H. Hagen, B. Shneiderman, and A. Ynnerman. Apply or die: On the role and assessment of application papers in visualization. *IEEE computer graphics and applications*, 37(3):96–104, 2017.

[34] A. Wu, D. Deng, F. Cheng, Y. Wu, S. Liu, and H. Qu. In defence of visual analytics systems: Replies to critics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1026–1036, 2022.