ChairVisE: An analytic lens for conference submission data

Animesh Prasad, Saumya Ahuja, Shenhao Jiang, Bimlesh Wadhwa, Min-Yen Kan
National University of Singapore, Singapore
Georgia Institute of Technology, USA
{animesh|saumya|bimlesh|kanmy}@comp.nus.edu.sg
jiangshenhao@gatech.edu

ABSTRACT

We introduce an analytic and interactive visualization framework for conference submission metadata. Our system, ChairVisE, is a workflow tool that imports submission metadata, allows the crossing and linking between key submission metadata objects, and the integration of both sensitive data in the anonymized form and user-contributed metadata. It supports common statistics that are often reported by general and program chairs and caters to several common submission systems. We demonstrate ChairVisE through two use cases of different scales, on the JCDL and ACL 2018 conferences.

CCS CONCEPTS

• **Information systems → Data analytics**; • **Applied computing → Document metadata**;

KEYWORDS

conference meta-data, visualization, analytics

ACM Reference Format:
Animesh Prasad, Saumya Ahuja, Shenhao Jiang, Bimlesh Wadhwa, Min-Yen Kan. 2019. ChairVisE: An analytic lens for conference submission data. In *Proceedings of ACM/IEEE Joint Conference on Digital Library (JCDL'19).* ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

Data mining on published articles is inherently lossy due to two reasons: 1) discarded metadata, and 2) confidential information. First, final publications are mainly rendered in portable document format (PDF), discarding much of the rich metadata useful for analytics. Second, a vast amount of scientific information remains uncaptured, as rejected submissions' titles and contents are not published, and confidential reviews are individually confidential and sensitive. While program chairs do need to be sensitive to individual reviews and scores, aggregate data can mask individually sensitive values, yet allow the mining of useful insights. Capturing such meta-data at scale and its open-source publishing would significantly forward the agenda of Science of Science analytics.

As a separate use case, program chairs of conferences often open a conference with summary highlights about the submission statistics and acceptance analytics, sometimes also cumulating with a preface in the proceedings itself. These tasks require program chairs or their staff to gather submission statistics. Inspired chairs may also seek to find unique aspects of the journey in the conference organization and emerging trends to visualize and discuss with its community. Our observation is that data analytics automation can solve the routine tasks and free chairs to work on curating meaningful analytics to share with their community.

We address both challenges in the scope of the scientific conference workflow. We propose an implemented framework, ChairVisE [1] (pronounced "Chair-Wise") that enables the aggregate analysis of such data by building upon the conference workflow.

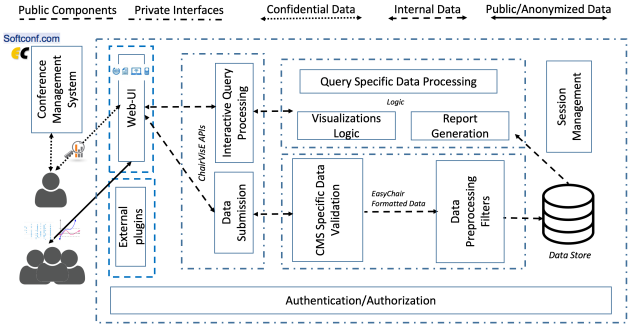


Figure 1: ChairVisE's modular architecture. Arrows indicate information flow, *cf.* Sec. 3.

2 RELATED WORK

To our knowledge, no scholarly work has tackled the specific problem of conference submission analytics at the point of conference submission and organization, which is what our work proposes. In contrast, the existing work primarily uses public, post-publication data — DBLP, Microsoft Academic Graph, Semantic Scholar Open Research Corpus — among many, and exploit the relationship between various entities and attributes — paper, author, institution, year, venue, geographic location and others. These relationships often visualized in individual or aggregate forms [1]. VOSviewer[2], Sci2[2], Histcite[3] and BibExcel[4] are non-exhaustive examples of such tools which can import published data and perform analytics.

3 CHAIRVISE FRAMEWORK

Fig. 1 illustrates the data model and accompanying workflow of ChairVisE, which supports two specific user roles: the chair and the public. The chair is the administrative, privileged user with

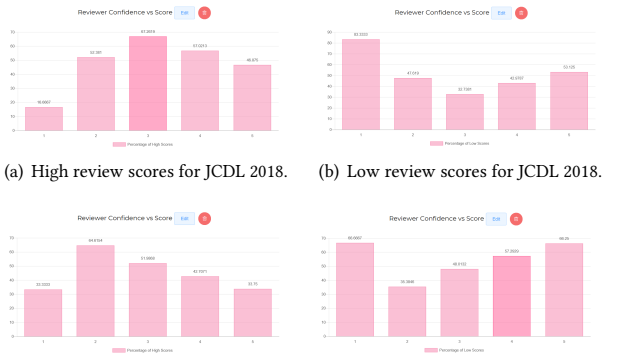
JCDL'19, June 2019, Urbana-Champaign, IL, USA

[1] https://github.com/WING-NUS/ChairVisE
[2] https://sci2.cns.iu.edu/
[3] http://interest.science.thomsonreuters.com/forms/HistCite/
[4] https://homepage.univie.ac.at/juan.gorraiz/bibexcel/

access to confidential data whereas public only have view of data approved by the chair.

Let us examine the framework in action from a running example.

- **Upload.** A logged-in conference chair user creates a conference project. After exporting the conference data from the conference management system (CMS), the privileged user upload the exported metadata to ChairVisE.
- ChairVisE's data model is built around three first-class data objects: *authors*, *submissions* and *reviews*. The default upload format takes these three fields as separate tables (files) with foreign key constraints that help join values together – *e.g.* `author.csv`, `submission.csv`, and `review.csv`. This default is common to the supported mapping for EasyChair CMS and is mapped (with metadata crosswalking) to the SoftConf CMS. ChairVisE can support other CMSs with relatively little effort in developing similar crosswalks between formats to the common data format.
- The system validates standard columns in each uploaded file for the expected data types. ChairVisE cooperatively allows users to modify crosswalks dynamically via a graphical interface, to proactively assist with malformed or custom metadata imports.
- The uploaded data is stored to a private table within ChairVisE's SQL database. The chair can apply various forms of anonymization, column extraction or other programmatic filters (inclusive of changing crosswalks) both pre- and post-database store.
- **Visualize.** The chair then selects one or many predefined visualizations from a drop-down list. The inventory of possible visualizations depend on the type of data represented in the import and are currently implemented in Javascript using Vue.js. These may differ per CMS, as certain CMS capture different fields associated with each of the three first-class data objects. For example, *submissions* data often captures the *submission time*, but some CMS also additionally store fields for the *modification time* and a flag for whether a submission was *withdrawn*.
- The system retrieves the relevant stored data, and renders the visualization to the chair's web browser. The chair gets immediate feedback, and can modify the SQL query engine for advance analytics. Once the underlying source data driving the visualization and its accompanying rendering parameters are satisfactory, the chair can save the visualization's state.
- ChairVisE also implements an arbitrary HTML text block as a visualization, to serve as caption to a visualization, or as an independent intervening textual discussion. Chairs repeat this process of selecting a visualization and fine-tuning it, to curate and form a narrative (similar in spirit to a conference blog post).
- **Share.** The chair can then print the report as a PDF (for preface preparation), or publish it for access, through generated, secret URLs granting either edit (for collaborating chairs) or view-only (for public) access. The URLs save session state such that the user (or one with knowledge of the edit URL) can authenticate and continue modification of the conference analytics.
- **Public Consumption.** Public users can view the conference analytics through the view-only URL. Certain views are interactive, meaning that the public can also post-apply certain filters to the views made available by the chair to drill down on analytics of interest to them. Public users can create their own version of the conference project, but the analytics available to the public is limited to non-private data, as designated by chair users.



(a) High review scores for JCDL 2018. (b) Low review scores for JCDL 2018.

(c) High review scores for ACL 2018. (d) Low review scores for ACL 2018.

Figure 2: Sample ChairVisE multivariate confidence (x–axis) versus percentage (y–axis) of reviewers with high (left) or low (right) score visualization.

4 SAMPLE VISUALIZATIONS

ChairVisE supports many univariate and multivariate visualizations from multiple CMSs. We showcase an interesting example by ChairVisE, using conferences' data – JCDL 2018 (medium-scale, EasyChair based) and ACL 2017 (large-scale, Softconf based) – for which the authors had access the original, privileged CMS data.

In Fig. 2, we visualize the percentage of reviews per score bracket (on the y–axis) and reviewer confidence (on the x–axis). Figs. 2(a) and 2(b) shows the statistics for JCDL 2018, where confidence is in $[0, 5]$ and review scores in $[-3, 3]$ (higher the better). We take $[-3, 0)$ as low and $(0, 3]$ as high scores. Unsurprisingly, the general trend is that reviewers with either high or low confidence tend to give lower scores. This trend holds even with the relatively larger scale ACL 2018. In Figs. 2(c) and 2(d) for ACL 2018, where confidence is in $[0, 5]$ and review scores in $[1, 5]$ (higher is better), we take $[1, 3)$ as low and $(3, 5]$ as high. Interestingly, the effect is not as pronounced as in JCDL where the maximal difference is as large as 50% between certain confidence intervals. This pattern reaffirms the process of triage practiced by small- and mid-sized conferences rather than review score threshold based blanket cut-offs for acceptance. However, this also suggests that reviewers' confidence should also be taken into account, rather than just the margin from the clean accept threshold. Further, this calls for more studies on exploring the underlying issues causing such phenomena – lack of expertise, overly diverse conference scope, sub-optimal reviewer assignment, reviewers' personal biases – among many.

5 CONCLUSION

We propose ChairVisE, a framework to facilitate the analysis of rich conference meta-data and thereby better assisting the understanding, transparency and analytics of the research community for itself. It supports common statistics as well as complex patterns that are often reported by general and program chairs and caters to several common submission systems.

REFERENCES

[1] Jiaying Liu, Tao Tang, Wei Wang, Bo Xu, Xiangjie Kong, and Feng Xia. 2018. A Survey of Scholarly Data Visualization. *IEEE Access* 6 (2018), 19205–19221. https://doi.org/10.1109/ACCESS.2018.2815030

[2] Nees van Eck and Ludo Waltman. 2009. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 2 (2009), 523–538.