KDD 2020 Highlights

Rajesh K. Gupta
Department of Computer Science and Engineering
University of California, San Diego
gupta@eng.ucsd.edu

Yan Liu
Computer Science Department
University of Southern California
yanliu.cs@usc.edu

1. INTRODUCTION

KDD made history in 2020. It was planned to be held in the San Diego Convention Center in an environment that would attract a large number of participants, companies to an attractive location. The program planners also made a concerted effort to broaden the intellectual scope of the forum as well as participation by diverse communities. In particular, the conference planned for offering complimentary full-time day-care to enable participation by women and parents with young children. The emergence of COVID-19 disease in March, however, caused considerable uncertainty leading to a final determination for a first-ever entirely virtual conference. This shift was dramatic at multiple levels since KDD is more than a meeting of presenters and their audience. It is also a place for demonstrations, chance encounters and a very engaging floor exhibition with its own ongoing events. Without exhibitors, their sponsors, and missing attendees, KDD won't be KDD. And it will also have a significant financial liability to the conference for the event contracts already in place. With significant cooperation from ACM and support from SIGKDD Executive Committee the team not only avoided financial impact but emerged with a net positive financial outcome even after reducing the registration fees by over 80%.

Beyond financials, the conference, while presenting numerous unprecedented challenges, emerged as a successful and memorable event with a number of heartwarming firsts and highest ever attendance and participation rates in the history of KDD. The challenges included inadequacy of the online platforms especially when it comes to providing virtual exhibitions integrated with technical talks on the same platform, technical support and a tremendous last-minute task of educating volunteers and presenters on the nuances of the online presentations and the platform chosen. While the team was able to pull together a strong program successfully, there is no question it required extraordinary (and in some cases, unnecessary) efforts by many involved. In a full report, we made several suggestions to streamline the Organizing Committee to improve these processes as well as suggestions for improved structuring of conference logistics versus conference content management by the volunteers to keep up with the growing scale, reach and intellectual diversity of the communities that KDD draws upon.

2. CONFERENCE HIGHLIGHTS

KDD 2020 made many firsts in KDD history, including

- First time close to 50% organizing committee and invited talk speakers were female and other underrepresented groups;

- First ever a full Diversity & Inclusion track with 20 amazing presenters and two mentoring panels that were very well attended;

- The conference attracted the largest number of paper submissions in both research track (1,353 submissions) and applied data science track (761 submissions);

- The conference attracted the largest number of attendees in its history with over 3,950 attendees drawn from over 86 countries despite tremendous challenges posed by time zone differences of attendees and presenters;

- Special sessions, panel discussions and a number of technical sessions covering various aspects of the data science behind COVID-19 disease and epidemiology.

Yet, perhaps the most salient feature of the conference was the tremendous engagement of the KDD volunteers in addressing the challenges posed by the dramatic shift in conference to virtual conference. The central theme of KDD 2020 is "data science for social good", "fairness, trust and inclusiveness in data science", and "data science for covid-19". Below is a review snapshot of the KDD-2020 program:

- Four keynote presentations by Emery Brown on data science in healthcare applications, Manuela Veloso on data science in financial applications, Yolanda Gill on explanation through meta-provenience and Alessandro Vespignani on spread modeling of COVID 19;

- 15 informative and innovative talks by industry leaders in our Applied Data Science Invited Speaker series featuring 9 women speakers;

- 217 Research presentations, 44 ADS presentations;

- 1 Research Panel and 1 ADS Panel;

- Three theme days including Deep Learning Day, Earth Day, and Health Day with the groundbreaking AI for COVID Health Day program;

- 31 lecture-style tutorials, 17 hands-on tutorials, 35 workshops, 2 late-breaking sessions, one on trustworthy data science, and the other is on COVID-19.

3. ATTENDEE STATISTICS

The 3,950 attendees came from 86 countries. Among them, 2,582 attendees were male, 1,093 were female, and the rest unidentified. The top 10 countries of attendees are US (2,385), China (297), Japan (147), Canada (107), South Korea (100), United Kingdom (85), India (76), Germany (61), Australia (49), and Brazil (38). As we can see the numbers of attendees are highly biased to the time zone for virtual conferences. In particular, Pacific Time zone was a significant disincentive for European participants. We believe a significantly higher participation potential exists for participation from China which was also affected by restrictions on content availability on a portion of our online platforms.