

# Industry Government Track Program Chairs' Welcome

The 2013 Industry Government track continued the recent focus on real applications of KDD technologies that provide measurable and documented benefits to Industry and Government. Of specific interest were papers likely to advance the general understanding of practical issues regarding the deployment and use of KDD technologies by Industry and Government. Papers were solicited in three areas: (1) deployed applications, (2) knowledge discoveries, and (3) emerging applications.

Papers that did not fit in these categories were unlikely to be accepted. In particular, papers that asserted the existence of a potential problem that could be solved by KDD technologies but provided no evidence of interest from potential users of a KDD solution and papers that used real data but did not attempt to deal with real-world complexities were rejected.

Papers that developed a novel algorithm and applied it to a real dataset and achieved better performance according to some metric but did not have any evidence of external validation or interest from potential users were also rejected from this track. Those papers are valuable for the KDD community but not a good fit for the Industry Government Track which is focused on real applications as opposed to applying an algorithm to a data set. As one PC member stated in a review "Using real data is not the same as solving real problems in a real operational setting." On the other hand, papers that discussed enabling infrastructure or issues of general interest regarding the real-world use of KDD technologies were viewed favorably, even when they did not discuss a specific case study or application. These infrastructure and issue papers were included in the "emerging" category.

Descriptions of the three areas and the review criteria were posted with the Call for Papers on the Conference web site and available to authors in advance of submission, and are reproduced below. Note that authors were required to identify in which of the three areas their paper should be evaluated and to address all the required elements for that area.

**Deployed:** Deployed KDD case studies describe deployed projects with measurable benefits that include KDD technology. These papers must clearly describe the industry or Government problem that is solved, the overall architecture of the deployed system, the data sources used, the reasons for the choices of particular KDD technologies, how KDD technologies solved the problem, the particular KDD process embodied by the deployed application, the use and payoff of the application, the costs to develop the application, the maintenance plan, and the number and types of users.

**Discovery:** Papers that describe discoveries of knowledge must clearly state what data sources and background knowledge were used, what data mining algorithms were tried, what overall KDD process was used, what the new discovered knowledge is, how the new knowledge was validated, and what the value to the industry or government is of such newly discovered knowledge. Claims regarding newly discovered knowledge must be externally validated. Examples of the form of such external validation might include acceptance in a peer-reviewed journal in the application domain of a paper describing the discovered knowledge, use of the discovered knowledge in a deployed application, or an investment decision that relies on the discovered knowledge. Note that back testing or improvements of internal metrics such as accuracy or area under the curve are not considered external validation. This category is intended for significant discoveries of domain knowledge that were enabled by KDD technologies and processes; it is not intended for new knowledge in the form of KDD research results.

**Emerging:** Emerging application and technology papers discuss prototype applications, tools for focused domains or tasks, useful techniques or methods, useful system architectures, scalability enablers, tool evaluations, or integration of KDD with other technologies. Such papers must clearly explain the requirements arising from the particular industry or government setting for which the application is being developed and from the particular databases on which the application is based. These papers must also identify how the emerging solution is using KDD technologies to address these requirements, the deployment plan, and the evaluation methodology and metrics for the emerging application. Pragmatic issues and considerations include important practical and research considerations, approaches, and architectures that enable successful applications. This category may include comparative evaluations of different KDD technologies for particular application problems or enabling infrastructures for large scale deployment of data mining or analytic techniques. Preferences will be given to papers whose insights may generalize to other domains or problems. Product advertisements will not be accepted.

The application areas cover a broad range, including security, advertising, sales, system monitoring, recruitment, health, and others, showing the wide range of fields to which KDD is being applied and adding value.

The track received 136 papers and accepted 34, of which 10 were considered deployed, 2 discovery, and 22 emerging, for an overall acceptance rate of 25%. All submitted papers received at least three reviews from separate PC members. For papers with divergent recommendations, a discussion among reviewers was held to reach consensus. We then performed a meta-review of all papers whose ratings and reviews were sufficiently positive that there was a possibility of acceptance, and made our final decisions based on paper quality and fit to the track's goals as evidenced by the paper's ranking on all of the review criteria.

We would like to thank all our reviewers, especially those who took on more papers than initially assigned to ensure that all papers received the necessary number of reviews.

We hope you find these papers interesting and insightful and look forward to continued improvements in the use of KDD techniques to provide real value to industry and government.

#### **Review Criteria:**

Content: does the paper contain all the required elements for its area?

Value/Payoff: what are the measureable benefits of the deployed application, the value of the newly discovered knowledge, or the generalizable insights from the emerging application?

Cost: what is the cost/effort of the reported project (to develop and maintain the application or to discover the new knowledge)?

External significance: is there clear evidence that the application problem being addressed and the KDD solution matters to an identified external set of users?

External validity: is the knowledge discovered/ used by the application externally validated?

Data and knowledge sources: does the paper thoroughly describe the data sources and background knowledge used?

Use of KDD technology: does the paper thoroughly describe the selection, use, appropriateness and effectiveness of the KDD technologies used?

KDD process: does the paper thoroughly describe the KDD process enabled by the application or used to discover knowledge?

Novelty/originality: does the work reported apply KDD technology to a new problem, develop or apply a new type of KDD technology to an existing problem, or apply KDD technology to achieve a better result than with other techniques?

Technical soundness: does the paper demonstrate an accurate understanding of KD research and technology? Does it identify relevant previous work? Does it appropriately match techniques to problem characteristics?

Evaluation: is the reported work carefully evaluated? Is it statistically sound? Are results compared to appropriate baselines? To other approaches and solutions? To previously reported work? Are its claims supported?

Deployment/use: is there clear evidence that the application is deployed or on a path to deployment, or that the discovered knowledge will be used?

Clarity: is the paper clearly written? Is it organized logically? Are there sufficient figures and tables to illustrate its main points?

Generality: does the paper contain insights or results that will generalize or apply to other domains or applications?

Accessibility: is the paper accessible to non-specialists in the application domain.

**Rayid Ghani and Ted E. Senator**  
Industry Government Track Program Chairs