

# VAST 2009 Challenge: An Insider Threat

Georges Grinstein<sup>1</sup>

University of Massachusetts Lowell

Catherine Plaisant

University of Maryland

Jean Scholtz and Mark Whiting

Pacific Northwest National Laboratory

## ABSTRACT

The 4<sup>th</sup> VAST Challenge centered on a cyber analytics scenario and offered three mini-challenges with datasets of badge and network traffic data, a social network including geospatial information, and security video. Teams could also enter the Grand challenge which combined all three datasets. In this paper we summarize the dataset, the overall scenario and the questions asked in the challenges. We describe the judging process and new infrastructure developed to manage the submissions and compute accuracy measures in the social network mini challenge. We received 49 entries from 30 teams, and gave 23 different awards to a total of 16 teams.

**Keywords:** visual analytics, human information interaction, sense making, evaluation, metrics, contest.

**Index Terms:** H.5.2 [Information Interfaces & Presentations]: User Interfaces – Evaluation/methodology

## 1 BACKGROUND

As in previous years [1,2,3,4] the objective of the VAST 2009 Challenge [5] is to help researchers move visual analytics discoveries and applications into practice via an innovative evaluation forum. These contests and challenges also help us develop and test metrics and evaluation methods for visual analysis environments, and make realistic tasks and datasets with ground truth available to our community.

## 2 VAST 2009 CHALLENGE

The VAST 2009 Challenge provided three mini-challenges and one Grand Challenge. This structure continues as we found that participation has greatly increased from the single grand challenge offered in 2006 and 2007 even though this year the 49 entries were down from the 73 entries in 2008. This was anticipated as we offered one less mini challenge and one of the mini challenges required processing video data. We presented awards identifying excellent work rather than determining winners and for this year we gave multiple awards to participants rather than limit them to only one award. Teams receiving awards were invited to submit a two page paper to the VAST proceedings. Teams not receiving awards were invited to submit a two page paper to the VAST

compendium. All participating teams were invited to attend a participant workshop held just prior to the VAST symposium.

The VAST 2009 Challenge scenario concerned a fictitious, cyber security event. An employee leaked important information from an embassy to a criminal organization. Participants were asked to discover the identity of the employee and the structure of the criminal organization. Participants were provided with the following three data sets, one for each mini-challenge:

- badge and network traffic within the embassy
- social network data (including geospatial information) about the criminal organization
- video data from cameras located near the embassy

The National Visualization and Analytics Center (NVAC) Threat Stream Generator project team at Pacific Northwest National Laboratory developed the data sets. Each set was embedded with non-trivially discoverable ground truth [6].

Each mini-challenge consisted of a data set, instructions and a number of questions to be answered. Participants could enter one or more of the mini-challenges. The Grand Challenge required participants to pull together information from all three data sets and write a debrief summarizing the situation. Teams were asked to provide a video and a clear process description of how the system was utilized to arrive at their conclusions as we have no access to the systems.

### 2.1 VAST 2009 Challenge Entries

We had 5 Grand Challenge (GC) entries and 44 mini-challenge entries. The breakdown of entries into the mini-challenges was:

- 22 for badge and network traffic
- 17 for social network
- 5 for video

Twenty eight different organizations from thirteen countries submitted entries. Eighteen were student teams.

### 2.2 Judging

Based on the participation numbers in 2008, we revised our judging procedure. We recruited members of the visual analytic research community and professional analysts to do a first round of review of the mini-challenges. Each judge was given access to the solutions and reviewed 4-5 entries online. They were asked to judge the entries based on the process descriptions submitted by the team which included screen shots and videos. They were asked to give ratings for the usefulness, efficiency and intuitiveness of the analytic process used, the visualizations and the interactions. The ground truth embedded in the datasets enabled us to provide quantitative evaluations using a number of measures of accuracy. Judges were also asked to comment on the novelty of the submission.

Based on the first round of reviews, an evaluation committee consisting of the chairs and professional analysts conducted a second round of review focusing on the best candidates from the

---

<sup>1</sup>grinstein@cs.uml.edu; plaisant@cs.umd.edu;  
jean.scholtz@pnl.gov; mark.a.whiting@pnl.gov

initial round of reviews. All review comments, including the numeric ratings from the judges, were given to the teams.

Awards for the VAST 2009 Challenges were not predefined; however the final award categories ended up being similar to the ones given in 2008: analytic techniques, visualizations, interactions and analytic debriefs. We gave 23 awards to 16 teams. These awards and the associated entries are described in the 16 two page summaries included in these proceedings. Two page summaries from other participating teams can be found in the VAST 2009 Compendium.

We made the solutions and all entries material available to the participating teams once the judging was completed. We will survey the teams to review the benefits of viewing other entries immediately.

Solutions and materials are made public at the time of the symposium in the Visual Analytics benchmark repository [7] which now includes several benchmarks, examples of uses and relevant papers.

### 3 AUTOMATED SUBMISSION, EVALUATION AND REVIEW SYSTEM

This year we developed an automated submission, evaluation and review system and web site to handle a larger volume of submissions. Once a team registered they could download the answer templates, view the specifics of how the accuracy measures were calculated and formally enter the challenge by uploading their entry. Entries were automatically checked to ensure that they were complete and accurately formatted.

Teams participating in mini-challenge 2 (the social network mini-challenge), were able to obtain accuracy scores for the social network they identified from the data. Teams were allowed to submit and receive an accuracy measure three times prior to submitting their final entry.

Scoring was automatic upon entering the challenge. Scoring for mini-challenge 1 and mini-challenge 2 used the simple measure of percent correct minus a percent incorrect. The scores ranged from -1 to 1 with a score of 0 occurring when a contestant submitted an answer that included the entire dataset. For mini-challenge 3 the score was based on the number of correct events.

On the administration portion of the website we were able to view all the teams that entered, view any entry, assign the two categories of reviewers, view reviews, triage the entries, keep track of awards and control who sees what through login.

### 4 PARTICIPANT DISCUSSION SESSION

We decided against holding an interactive session during VisWeek 2009. In the past, a small number of teams from the grand challenge entries were invited to work with an analyst in analyzing a small, but similar dataset. This event used a lot of our resources and benefited only a small number of teams.

Instead, we decided to enlarge the participant workshop held in 2008 from one evening session to a full day. The 2009 participant workshop include an overview of the entries, talks by analysts and an opportunity for participants to demonstrate their work to others. This gives teams an opportunity to interact with other teams and hopefully foster partnerships.

We anticipate that information about what worked and what did not will help the visual analytic research community make even more progress in the coming years. The workshop involves the participants in planning for the coming years, including identifying data sets of interest and discussing metrics and feedback.

## 5 SUMMARY

Accuracy becomes more difficult to judge as our tasks become more realistic and the data sets become more complex. For some of the mini-challenges, accuracy was less important than the supporting evidence provided within analyses. As the size of datasets increase, accuracy will also become more difficult to assess as it will be impractical to explore all possibilities.

As in previous years, teams were asked to provide an analytic product in the Grand Challenge. An analysis of the situation differs from just reporting the facts. While most of our teams do not have access to analysts, we are pleased with the progress that teams have made on using excellent analytical techniques and constructing analytic debriefs. Using their own tool to actually perform analysis has helped many teams see where refinements are needed.

The majority of the teams described tools developed specifically for the challenge. Others, however, used toolkits or adapted tools from other domains and described how these could be easily customized or modified to complete the mini-challenges or grand challenge.

We gave six awards for outstanding visualizations this year, including some for representation of uncertainty. We encourage readers to review the papers in these proceedings to see examples of novel visualizations developed for the VAST 2009 Challenge.

## 5 THE PATH FORWARD

The format of mini challenges and a combined grand challenge will be followed in 2010. Interested participants are advised to explore the Challenge repository [7].

### ACKNOWLEDGMENTS

We thank the National Visualization and Analytics Center<sup>TM</sup> (NVAC<sup>TM</sup>) located at the Pacific Northwest National Laboratory in Richland, WA. The Pacific Northwest National Laboratory is managed for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830. Members of the committee were also supported in part by the National Science Foundation (IIS-0713087, IIS-0712770, IIS-071319) and the National Institute of Standards and Technology. The participant workshop was supported in part by NSF (0925482).

We also wish to thank Jereme Haack, Carrie Varley and Cindy Henderson of Pacific Northwest National Laboratory; Andrew Canfield from Mercyhurst College, the students of the University of Massachusetts Lowell who worked to develop and manage the submission site, process the data and various evaluation computations, especially Laura Costello, Shawn Konecni, Heather Byrne and Adem Albayrak; and Swetha Reddy of the University of Maryland who developed and maintained the Challenge website and repository.

### REFERENCES

- [1] VAST 2006 Contest: [www.cs.umd.edu/hcil/VASTcontest06](http://www.cs.umd.edu/hcil/VASTcontest06)
- [2] VAST 2007 Contest: [www.cs.umd.edu/hcil/VASTcontest07](http://www.cs.umd.edu/hcil/VASTcontest07)
- [3] VAST 2008 Challenge: [www.cs.umd.edu/hcil/VASTchallenge08](http://www.cs.umd.edu/hcil/VASTchallenge08)
- [4] Costello, L., Grinstein, G., Plaisant, C. and Scholtz, J., Advancing User-Centered Evaluation of Visual Analytic Environments through Contests, *Information Visualization* (to appear).
- [5] VAST 2009 Challenge: [www.cs.umd.edu/hcil/VASTchallenge09](http://www.cs.umd.edu/hcil/VASTchallenge09)
- [6] Whiting, M., Haack, J., and Varley, C. 2008. Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software. In *Proc. of BELIV'08*, ACM, New York, NY.
- [7] Visual Analytics benchmark repository: [www.cs.umd.edu/hcil/VArepository](http://www.cs.umd.edu/hcil/VArepository)