

*Summer 2007 IPAM Research in Industrial Projects for Students***Symantec Project Proposal: A Reputation Problem**

Industrial Partner: Symantec Research Labs
Industrial Advisers: Gerry Egan & Darren Shou {gerry_egan, darren_shou}@symantec.com
Faculty Mentor: Kendall Giles (kendallgiles@gmail.com)

1 Background

How can we provide Internet users a safer online experience? Part of the solution will involve reputation. A reputation system collects, distributes, and aggregates feedback about participants' past behavior. These systems help people decide whom to trust, encourage trustworthy behavior, and deter participation by those who are dishonest even though few producers or consumers of the system know one another.

2 The Problem

Symantec is building a web reputation system that will aggregate data from a wide variety of sources. Data will consist of positive attributes (e.g. site has an SSL certificate and has been in business > 5yrs etc) and negative attributes (e.g. found X malicious downloads out of a total Y scanned downloads on site, detected Z pop up windows, detected A tracking cookies etc) about a web property. Propose and implement an approach to providing a simplified reputation score that can easily be tested and tuned. Possibly consider machine learning to improve the scoring system over time based on real-world experiences.

3 Candidate Starting Points & Directions

Consider the algorithms that rank webpage results returned from a search engine query. Each page is given a score that nominatively indicates webpage relevance. These ranking algorithms often are based on just quantitative attributes of the site, such as the number of links to and from a particular page. In online social networks, however, such as those based on product recommendations and purchases, there are more factors that a site visitor might consider before conducting a transaction at that site. Users of Ebay, for example, can view the reputation of a seller before purchasing a product. That reputation is based on the social and subjective assessments of previous users who have conducted transactions with that seller. By combining quantitative site attributes with feedback from users, we seek to provide an efficient and relevant system to enhance a person's online experiences. An existing database of website attributes will be used to construct and test a variety of webpage reputation scoring algorithms. An example starting approach might be to consider a linear regression classification system. In addition, effective means of presenting and communicating to a site visitor that page's reputation should be considered. For example, Jean Camp's social network NetTrust system demonstrates how such trust information can be communicated to a site visitor. Candidate algorithms and visualization approaches will be implemented and tested.

4 Research Tasks and Deliverables

- Task 1: Assess relevant/available features to be used as inputs, such as pre-defined website statistics and user input
- Task 2: Create website reputation score that is a function of selected input features
- Task 3: Design algorithms for score computation, score update, classification labeling
- Task 4: Prototype the most promising approach to visualize the reputation of a site to a user based on provided data sets

- Deliverable 1: A technical report detailing approaches taken, experimental results and conclusions
- Deliverable 2: Presentation to Symantec Research Labs and separately to IPAM RIPS at Project's day
- Deliverable 3: The prototype source/software

5 References

- Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer, 2001.
- L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.
- Alla Genkina, Jean Camp, "Re-Embedding Existing Social Networks into Online Experiences to Aid in Trust Assessment", (April 1, 2005).

Available at SSRN: <http://ssrn.com/abstract=707139> or DOI: 10.2139/ssrn.707139

- Ye Wang, Henry Emurian, "An Overview of Online Trust: Concepts, Elements, and Implications", *Computers in Human Behavior* 21 (2005) 105-125.
- Matthew Richardson, Rakesh Agrawal, Pedro Domingos, "Trust Management for the Semantic Web", *Proceedings of the Second International Semantic Web Conference*, 2003.