CS 286r Final Projects

Fall 2008

1 Project Guidelines

In this handout, we give you all the requirements for your course project, that will constitute 40% of your final grade. Your course project may be done individually or in pairs. If you choose to work in pairs, we expect the quality of your project to reflect the fact that two people have been working on the project. Your course project will consist of the following components:

1. A 2-page Project Proposal **due November 17, 2008 in class**

   In your project proposal, you should provide a high-level description of your project. What problem are you trying to solve? What results do you hope to get? Why is your problem important? How does your project relate to the papers we have been reading in class? Your goal should be to convince us that your ideas are well enough focused, and that you know what the first few steps will be. It is not sufficient to simply take one of the ideas below and turn that in as your project proposal. In addition to this, please make it clear to the teaching staff, what help you need from us (pointers to papers, access to data, etc.) if any.

2. A 2-page Project Update **due December 10, 2008 in class**

   In your project update, you are to detail any and all progress you have made on your project since you turned in the project proposal. In addition to this, you should clearly state what additional results you hope to get. Have the goals of your project changed at all after obtaining some preliminary results? We are requiring you turn in an update so that you do not save all the work to the last minute.

3. A 20-minute Project Presentation **during reading period**

   You are to prepare a 15-minute presentation of your course project and you will have 5 minutes to answer questions. You are required to attend all other students' presentations. The hope is that you can all get some good feedback on your projects from your peers.

4. A Final Report **due January 12, 2009 at 12 noon**

   Your project report is the most important part of your project. Your project report should read like a self contained research paper. What problem are you solving? Motivate your problem. What is the related work? What results did you obtain? What is your experimental methodology? (Of course, this question only pertains if you do an experimental project). What questions would you like to address in future work? Please note that we have set the due data as late as possible. No extension can be further granted.

If you choose to do a literature review of an area, it is not sufficient to simply take one paper and summarize it. We expect an in depth literature review. Your report should address: What is known so far? What are the open questions? What are the challenges in this area? In addition, at the end of a literature review, it is important to grasp what are some possible doable research questions that you could tackle. Your report should have some discussion about what you think the next step in this field is. *Literature review projects are considered individual projects. If you pick a literature review idea from our list of example project ideas, please get our approval before you work on it. We do not allow more than two individuals work on the same literature review.*

Below we have given you a number of project ideas to get you started in thinking about a project. You are in no way limited to these project ideas, in fact, you are encouraged to find your own project. Since the project is something that you will spend a great deal of time on, we hope that you pick something that you like and something that you think you have intuition for. We give you some guidelines for picking your own project:

1. Theoretical Project: Find a theoretical paper that we read or a theoretical paper related to the themes of this class. Are there too many assumptions made in the model? What happens when you try to relax one of the assumptions or alter the model a little? Where do the proof techniques in the paper break down for this new model? After understanding why the proof techniques in the paper do not hold for your model, you may be able to propose new methods for obtaining similar results.

2. Experimental Project: Find an experimental paper that we read or a experimental paper related to the themes of this. Are there things that you wish the authors did?

Remember to have fun with your project!

2 Example Project Ideas

2.1 Prediction Markets

- Empirical study of prediction markets: Data of IEM and Intrade are available online. Inkling Markets are pretty open at allowing users to set up their own markets especially for academic use. If you have some hypothesis to test for prediction markets, think about a way to examine it empirically. (Note that if you choose to set up markets to test your hypothesis, only running the markets is not enough. You have to perform the data analysis.)

- Discounted LMSR: Dimitrov and Sami [9] proposed discounted LMSR to discourage bluffing in LMSR. But discounting may bring other undesirable properties, such as arbitrage opportunities. Analyze under what situations these undesirable properties arise and how much impact they have on information aggregation. Propose a variation of the mechanism that can reduce these undesired properties. (Note that discounted LMSR essentially changes the liquidity parameter b in the market. Thus, the proposed mechanism may help to deal with the challenge of how to adjust the liquidity parameter after the market is in operation.)

- Outcome manipulation: We read papers on manipulation of prices in prediction markets by bluffing. This assumes that traders only care about profit in the market. A more important concern arises when traders have stakes outside of the markets, e.g. when market prices

affect decisions and traders have different values on different decision outcomes. Do an extensive literature review around this problem. There aren't many papers directly related to this problem. The novelty of your work is to find connections of literatures that are seemingly indirectly related. You can also try to propose a model for this problem, and provide justifications for your model and some initial analytical results.

- Combinatorial prediction markets: Think about a combinatorial setting that's different than what we have covered in class. Design a reasonably expressive bidding language. Examine whether a call market or a market maker can efficiently operate the market for the bidding language, or provide some approximation matching/pricing algorithm for it. For example, consider tournament betting with a call market.

- Market simulation: Theoretical papers regarding prediction markets make strong assumptions about agent rationality and information structure, which often are too good to be true. Use simulation to study how different rationality levels and different information structure affect price accuracy. (You may build your simulation based on existing market simulation software, such as the Santa Fe Artificial Stock Market, if you believe that can help you.)

- Distributed computation: Consider an extension of Feigenbaum et. al. [10] Think about the incomplete information case where some of the bits are not known by any market participants. Use a similar model to show the computation power of markets.

- "Predicting" prediction markets and event detection: If trading does not occur without reason, then it should be the result of outside information, which people condition upon to derive new expectations and trade based on these new beliefs. Assuming that you have a corpus of news from the daily political and economic press, can you parse them to try and forecast price movements in prediction market securities? Can you use news and prediction markets to detect important events?

- Information diffusion: Alter the traditional market model such that information is not broadcast to all the traders, but instead revealed in random to just a few of them. Traders do not know everybody else and do not know what information is revealed to whom. Consider whether recipient agents always have an incentive to propagate this information; give examples. If yes, think about whether they want to do it truthfully. (Assuming truthfulness, you may also try to quantify the impact of this diffusion process to price convergence speeds experimentally or theoretically.) If no, suggest ways that may help alleviate the problem. For example, how easy is it to have a policing service "punishing" people who provably have withheld or transmitted false information? Finally, if no information is transmitted but the original information recipient just traded on his signal, what might be the response of this "neighbors" in the graph? Consider that always conditioning on the first agent's observed action may make them susceptible to manipulation and also reveal unwanted information to yet their other neighbors; on the other hand, not conditioning means no information diffusion and ultimately no information aggregation.

- Common priors assumption: Many problems discussed in class assume common priors. Pick a problem. Examine the effects of either (1) removing the common prior assumption altogether, or (2) stipulating that priors of all agents are sufficiently close ($D(p_i||p_j) < \epsilon$). It is expected that most of the proofs break down, a negative result. Can you derive any positive results?

2.2 Social Choice

- Computational aspects of voting: Perform an extensive literature review on computational results of voting protocols. It includes but is not limited to complexity of winner determination, worst-case complexity of manipulation, average-case complexity of manipulation, complexity of communication, and situation for restricted preference domains, and situation for incorporating randomness. The novelty of the literature review is to provide new understanding of existing results, identify open problems and directions, and maybe provide alternative proofs for some of the existing results.

- Approximation algorithms for manipulation: Find approximation algorithms for manipulating voting rules that are provably hard to manipulate.

- False-name-proofness and strategy-proofness: Characterize the relation of false-name-proofness and strategy-proofness, and of group false-name-proofness and group strategy-proofness. You may need to review existing results on voting rules that satisfy these properties, and build your results on them.

- False-name-proofness with restricted preference: We have seen many "negative" results when requiring voting rules to be false-name-proof or group false-name-proof. With restricted preferences, there may be some positive results. Examine false-name-proof and/or group false-name-proof voting rules (with or without costly voting) for single-peaked preference or dichotomous preference.

- Evaluation of false-name-proofness: False-name-proof voting rules have to randomize over alternatives in determining the winner. This means that sometimes such voting rules may elect an alternative that few voters prefer. Assuming that voting are costly, simulate optimal false-name-proof voting rules introduced by Wagman and Conitzer [17] and standard voting rules that are not false-name-proof to evaluate their efficiency.

- PageRank and recommendation systems: The page rank axioms paper [2] that we read is the first of a stream of research on ranking systems. Later work provides some impossibility results [1], quantifies incentive compatibility of ranking systems [4, 5], extends the approach to personalized ranking [3] and recommendation systems [6]. Perform an extensive review on axiomatic analysis of ranking systems and their application to recommendation systems. Identify open research questions as well as doable problems that you can tackle.

2.3 Peer Production

- Novel peer production: Find a commonly used peer production system not discussed in class. Perform an experimental analysis of the system. Were you able to find any inefficiencies in the system? Propose methods for overcoming any inefficiencies.

- Cooperation and evolution: Consider the problem of cooperation, free-riding and punishment in a peer production system. In these systems, a common goal is accomplished through individual contributions. What is the reason why these individuals choose to participate in the first place? And, once they decide to take part in the process, what guides them into collaborating with each other to achieve a joint goal? Discuss evolutionary arguments in

favor of cooperation and build a model and a simulator to show that either (1) cooperation is favored by natural selection or (2) a cooperation "gene," stochastically expressed, leads a species to higher evolutionary fitness.

- Reputation manipulation: In the "Novelty, Popularity and Attention" paper [18], Wu and Huberman analyze the performance of a number of ranking rules for Digg.com. What possible ways can users manipulate the rankings for Digg.com? Can you propose novel methods to overcome possible manipulations? We mention the following manipulation: Users of Digg.com can send stories of interest to their friends and ask them to "digg" them. Since "digg-ing" is not costly, an individual may feel compelled to digg a story even if he/she is not interested in the story. Digg.com also has an underlying social network where users can establish who their friends are on the website. Can you propose a method to deal with this sort of attack? Keep in mind that it is likely that those who are friends, may also have common interests. How can you deal with these two opposing factors?

- User behavior in peer production systems: Yang et. al. [19] identify how user participation changes over time and what a "winning strategy is" in Taskcn. Consider Yahoo! Answers or the ESP Game, can you gather data and perform an empirical analysis to support or reject the hypothesis of how users play over time?

- Experts in peer production systems: Zhang et. al. [20] identify "experts" in online forums. Can you modify their method and identify "experts" in Wikipedia or in online QA forums? Can you propose methods to take advantage of this information?

2.4 Human Computation

- Applications of human computation: We have seen many applications of human computation paradigm in class. In addition to the papers studied in class, several other papers [15, 14, 13, 16] describe systems that employ the human computation paradigm to solve real world problems. Can you come up with a problem that you think that would be appropriately solved using human computation? Design an initial system to solve your problem. This design should have consideration for methods of cheating and attacks by bots.

- Model of human computation: Very little work has been done in theory to explain how people behave in human computation systems and to predict outcomes of systems. Can you propose a model for a human computation system and use it to analyze the game play?

2.5 Reputation Systems

- Peer-prediction: Given the peer-production method for eliciting feedback, can you characterize when collusion may be beneficial for raters? How much impact does collusion have on feedback elicitation?

- Sybilproofness of reputation systems: In the "Sybilproof Reputation Mechanisms" paper [7], it is assumed that agents can split their incoming links among sybils. This is not possible in many situations, such as web links and many peer to peer systems. Relax this assumption, and examine how the results change.

- Manipulability under sybil strategies: Cheng and Friedman [8] studied manipulability of PageRank under sybil strategies. Pick a different reputation mechanism, and study its manipulability under sybil strategies. You can either perform an analytical analysis or use simulation to get insights.

- Reputation system strain: Enumerate all the strategies reputation systems use to negate attacks by malicious users. Although good strategies exist for particular types of attacks, how do these systems fare towards a combination of attackers? Does the usability of the system decline with increased efforts to defend against attackers? Study this question using a simulator, where the proportions of the various attackers are an externally defined parameter.

2.6 Scrip Systems

- Simulation of scrip systems: Friedman et. al. [11, 12] provide some theoretical results on behavior of scrip systems. Their results are based on a set of assumptions. Relax some of the assumptions and simulate the system to examine how far it deviates from Friedman's results.

References

[1] A. Altman and M. Tennenholtz. On the axiomatic foundations of ranking systems. In *International Joint Conference on Artificial Intelligence*, 2005.

[2] A. Altman and M. Tennenholtz. Ranking systems: The pagerank axioms. In *ACM conference on Electronic commerce*, 2005.

[3] A. Altman and M. Tennenholtz. An axiomatic approach to personalized ranking systems. In *International Joint Conference on Artificial Intelligence*, 2006.

[4] A. Altman and M. Tennenholtz. Quantifying incentive compatibility of ranking systems. In *AAAI*, 2006.

[5] A. Altman and M. Tennenholtz. Incentive comaptible ranking systems. In *International Conference on Autonomous Agents and Multiagent Systems*, 2007.

[6] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz. Trust-based recommendation systems: an axiomatic approach. In *WWW*, 2006.

[7] A. Cheng and E. Friedman. Sybilproof reputation mechanisms. In *ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems*, 2005.

[8] A. Cheng and E. Friedman. Manipulability of pagerank under sybil strategies. In *Workshop of Networked Systems (NetEcon)*, 2006.

[9] S. Dimitrov and R. Sami. Non-myopic strategies in prediction markets. In *ACM Conference on Electronic Commerce*, 2008.

[10] J. Feigenbaum, L. Fortnow, D. M. Pennock, and R. Sami. Computation in a distributed information market. *Theoretical Computer Science*, 343(1-2):114–132, 2005.

[11] E. Friedman, J. Halpern, and I. Kash. Mefficiency and nash equilibria in a scrip system for p2p networks. In *ACM Conference on Electronic Commerce*, 2006.

[12] E. Friedman, J. Halpern, and I. Kash. Optimizing scrip systems: Efficiency, crashes, hoarders, and altruists. In *ACM Conference on Electronic Commerce*, 2007.

[13] C.-J. Ho, T.-H. Chang, and J. Y. jen Hsu. Photoslap: A multi-player online game for semantic annotation. In *AAAI*, 2007.

[14] L. von Ahn, S. Ginosar, and M. Blum. Improving accessibility of the web with a computer game. In *ACM conference on Human Factors in Computing Systems*, 2006.

[15] L. von Ahn, M. Kedia, and M. Blum. Verbosity: A game for collecting common-sense knowledge. In *ACM conference on Human Factors in Computing Systems*, 2006.

[16] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum. reCAPTCHA: Human-based character recognition via web security measures. *Science*, pages 1465–1468, 2008.

[17] L. Wagman and V. Conitzer. Optimal false-name-proof voting rules with costly voting. In *AAAI*, 2008.

[18] F. Wu and B. Huberman. Popularity, novelty, and attention. In *ACM conference on Electronic commerce*, 2008.

[19] J. Yang, L. Adamic, and M. Ackerman. Crowdsourcing and knowledge sharing: Strategic user behavior on taskcn. In *ACM conference on Electronic commerce*, 2008.

[20] J. Zhang, M. Ackerman, and L. Adamic. Expertise networks in online communities: Structure and algorithms. In *WWW*, 2007.