Class Presentation Assgnment: Part 1
Topic Selection

Due date: Wednsday, October 7, in-class

Class Presentation

This is one of the two quarter-long assignments in the course (the second one is the team project).

During the course of the quarter you will select a topic of interest in the area of data mining, knowledge discovery in data, web mining and/or information retrieval, study the available literature on the topic and prepare a number of deliverables, including an in-class presentation.

The assignment is broken into a number of stages, each with one or more deliverables:

Stage 1. Topic Selection: Find a topic of interest, create a short description of what you want to study and present to class.

Stage 2. Preparation: Do a literature search, create a reading list, read it.

Stage 3. Delivery: Create an overview of your topic (in a form of lecture notes) and prepare and deliver an in-class presentation.

Collaboration notes

This is an individual assignment in a sense that each student is expected to submit *individual deliverables* at each stage of the assignment. In particular, **every student in the class** will prepare *the final written deliverable (lecture notes)* and will personally deliver *the in-class presentation.*

At the same time, the following forms of collaboration are allowed:

- **Joint work on the same topic.** If you believe that there is enough of interesting material in a specific topic for more than one presentation, multiple students can work on the topic. In this case, topic description, reading lists/bibliography and similar topic-specific deliverables can be joint (one list pert topic, rather than per student), while the final deliverables — individual. You are expected to present different things in class, but beyond that, it is up to you and your peers on what to cover.

  Please note, that on each stage I will expect each student to do about the same amount of work. So, for a group of two people, reading lists are expected to be longer, for example, than for a topic covered by an individual student.

- **Peer-to-peer communication.** Feel free to talk to your peers (as well as to me) about your work on this assignment. At the end of the quarter, you are expected to know very well the contents of your in-class presentation, however, as you are preparing it, feel free to talk to others in the class, if you need help understanding a article, or simply could use someone as a sounding board.

- **Content coordination.** I would like each student to create a *lecture notes-style* deliverable for the in-class presentation (so that I could use them in future instruction). It would be great if in-class presentations and lecture notes contained as little overlap in content as possible. You are allowed to coordinate with each other, and to link to the work others in order to ensure lack or the bare minimum of repetition in the content of your deliverables.

0.1 In-class presentation overview.

Presentation length. Based on your work throughout the quarter, each of you will prepare an in-class presentation. We have alotted six class periods (possibly seven) for these presentations, at the end of the quarter (tentatively starting November 9). We can fit three presentations into a single class period, so, each presentation should be about 35 minutes in length.

Presentation style. Your in-class presentations should be styled as **lectures** or **mini-lectures**, rather than research reports. The objective of each presentation is to introduce the class to the new technical content (problems, techniques, methods, algorithms, etc). You can use any means you find useful to deliver the presentation (board, slides, handouts, etc.).

Scheduling of presentations. The schedule of presentations will be worked out by me in consultation with everyone. My goal is to schedule presentations on related topics and issues close to each other. Some of you **will have to** present on November 9, while some will present much later (December 2).

Stage 1: Topic Selection

The first stage of your assignment is to select a topic of interest and prepare a short description of what you want to eventually present to class.

Please note that while you may (and in fact are encouraged to) use the textbook as the source of inspiration for your topic, you are expected to work on and present the material that is not covered in the textbook. For the most part, this entails reading academic papers.

Resources. The textbook can give you an idea about various types of interesting problems that can be covered. While the textbook is not complete in the sense of coverage of the field (e.g., text mining is mostly omitted), it is a good starting point. Beyond that, you can do a preliminary literature search. Data mining is the topic of multiple conferences and journals. Among those, I recommend as starting points the following:

- ACM Computing Surveys. This is a general purpose survey journal, but in the past 10-15 years, it has published a number of surveys in the areas of our itnerest. These surveys are excellent starting points for your work as well.
- ACM SIGKDD. ACM has a SIG (special interest group) devoted to data mining and similar topics: SIGKDD. SIGKDD sponsors an annual conference, you may find it useful to look in the proceedings of recent years for state-of-the-art assessment of your chosen subfield of data mining.
- Other Conferences. IEEE ICDM, International Conference on Data Mining; CIKM, Conference on Information and Knowledge Management; ACM SIGMOD Conference (it is a database conference with data mining tracks), AAAI Conference (Conference of the American Association for AI), IJCAI (International Joint Conference on AI)... There are many more conferences with Data Mining tracks on them.
- Academic Journals. ACM Transactions on Databases (TODS), ACM Transactions on Information Systems (TOIS), IEEE Transactions on Knowledge and Data Engineering (TKDE), Data Mining and Knowledge Discovers (Springer), Statistical Analysis and Data Mining, ACM Transaction on the Web (TWeb), International Journal of Web Based Communities (IJWBC), International Journal of Web Engineering and Technology, International Journal of Web Information Systems, Transactions on Machine Learning and Data Mining, Journal of Artificial Intelligence Research (JAIR), Information Systems and more...

  Note, that many of the journals above are devoted to broader topics, and KDD/Data mining form only part of their pool of articles. You main work with these sources will come later in the quarter, when you are looking for the actual papers to read.

Deliverables

There are two deliverables for this stage.

- **Written description.** Write a paragraph (or two) on the topic you have selected. Outline why it is interesting and try to specify what exactly you will be presenting in class (the latter is not mandatory - you may choose a topic now, but decide what you will be talking about later in the quarter). Include a couple of references to give me an idea of what you are thinking about (and also to assure me that there is literature on this matter).

  All written descriptions shall be places on the course wiki. (Special area for this will be created).

- **A 2-min informal presentation in class.** On the due day of this stage each student will announce their choice of the topic and try to explain briefly why they are interested in it.