應用機器學習技術

Machine Learning Techniques for Applications

- **Course Objectives**

  Although it is possible to learn a variety of machine learning and data mining techniques from lectures or books, how to accurately and effectively apply the techniques to the real-world data is a completely different story. In many cases, data minders have to suffer a painful process of trail and error because of lack of experience. Therefore, it is quite important to learn how to deal with the practical issues on data.

- **Course Description**

  In this course, we aim to increase students' experiences by handling some real-world problems proposed as the past or ongoing competitions in machine learning or data mining society. In particular, our goal is to participate in ACM KDDCup2012, which is the most prestigious data mining competition. The course will be run in an interactive manner, and students must discuss with the instructors and other classmates about their findings and the problems they encounter every week.

- **Class Schedule**

  1. Before KDDCup 2012 Registration Opens (March 5): overview of machine learning and data mining techniques by instructors
     1) Association Rules and Classification (one week)
     2) Support Vector Machines and AdaBoost (one week)
     3) Feature Selection and Evaluation (one week)

  2. Before KDDCup 2012 Begins (March 12): literature survey or work on other similar competition datasets

  3. Before KDDCup 2012 Ends (June 29): work on KDDCup 2012 until end of competition and possibly paper writing

later (if we win)

- **Teaching Approach**

  The course will be run in an interactive way, in which students must discuss with the instructors and other classmates about their findings as well as the problems they encountered every week. Students need to implement different kinds of intelligent systems for the competition and run extensive experiments to verify them. Students will compete with the other students in the class as well as other teams all over the world in KDDCup. Students will have WEEKLY presentation about their progress in the previous week.

- **Teaching Assistant Tasks**

  Organize student's presentations.

- **Course requirements/Grading standards**

  1. Depend on student's weekly performance (judged by their efforts, novelty, and presentation), and weighted by how much they contribute to the overall competition results.

  2. No exams.

- **References**
  - Ian H. Witten and Eibe Frank. Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations. Third edition, Morgan Kaufmann, 2011.
  - Tom M. Mitchell. Machine Learning. McGraw-Hill, 1997.
  - Ethem Alpaydin. Introduction to Machine Learning, second edition, The MIT Press, 2010.
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second edition, Springer, 2009.
  - Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.