

Evaluation of a graduate level data mining course with industry participants

Peter Christen

Department of Computer Science,
Faculty of Engineering and Information Technology,
ANU College of Engineering and Computer Science,
The Australian National University

Contact: peter.christen@anu.edu.au

Course Web site: <http://cs.anu.edu.au/student/comp8400/>

Data mining education

- Increasing demand from industry for graduates with data mining / analytics knowledge
- Data mining is multi-disciplinary (statistics, machine learning, databases, algorithms and data structures, visualisation, business aspects, etc.)
- Data mining courses are now being taught by many universities, as well as private providers
 - Hard to teach all aspects of data mining (either some topics in depth, or breath of many topics)
 - With different emphasis
 - At different levels (under- or post-graduate)
 - Assuming different pre-requisites (prior knowledge)

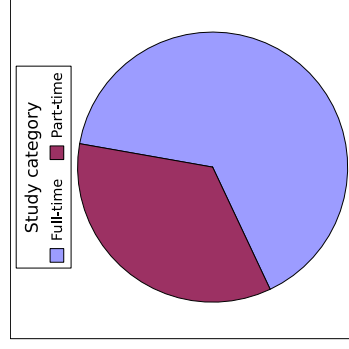
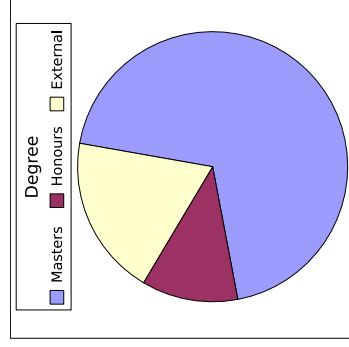
Outline

- Data mining education
- Data mining course evaluations
- Student background
- Course overview
 - Course modules
 - Tutorial papers
- Course evaluation
 - Based on an end-of-semester questionnaire
 - Look at difference between Master students and external participants
- Outlook

Data mining course evaluations

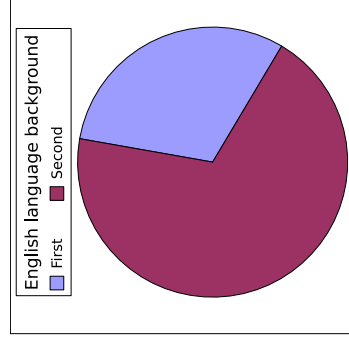
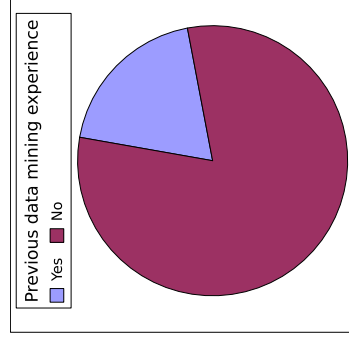
- There is only a small number of evaluations of data mining courses in the literature
 - They mainly describe course structure and content
 - Courses for undergraduates and Masters
 - Courses for CS and non-CS students
 - Most only cover core algorithms used in data mining
 - No thorough analysis of student feedback
 - Only individual courses, not over several years
- There is an ACM SIGKDD curriculum committee and a curriculum proposal available at:
<http://www.sigkdd.org/curriculum.php>

Student background (1)



- 27 students enrolled (1 dropped out)
- Only 1 female student
- 18 Masters, 3 CS honours, 5 external participants

Student background (2)



- 5 students had previous data mining experience (including 3 of the 5 external participants)
- Almost 70% of students had English as a second language

Course overview

- Nineteen one-hour lectures (mostly based on *Han and Kamber* text book)
- Four one-hour tutorials (research paper discussions)
- Four one-hour practical laboratories (using *Rattle* data mining tool)
- Two assignments (essay, algorithm implementation or practical data mining project, each worth 15% of final mark)
- One 5-minute student presentation (on a data mining research paper of their choice, worth 20%)
- Final written examination (not open book but 1 A4 cheat sheet, 3 hours, worth 50%)

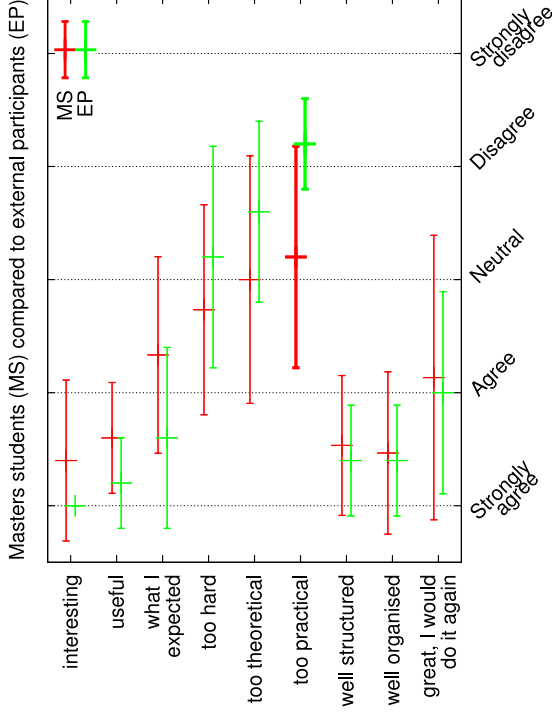
Course modules

Topic	Hours
1 Course introduction and data mining overview	1
2 Data mining process and data issues	2
3 Data pre-processing and data integration	2
4 Mining frequent patterns and association	2
5 Cluster analysis	2
6 Classification and prediction	4
7 Mining time series and data streams	1
8 Privacy-preserving data mining	1
9 Web and text data mining	2
10 End-to-end data mining, trends and social impacts	2

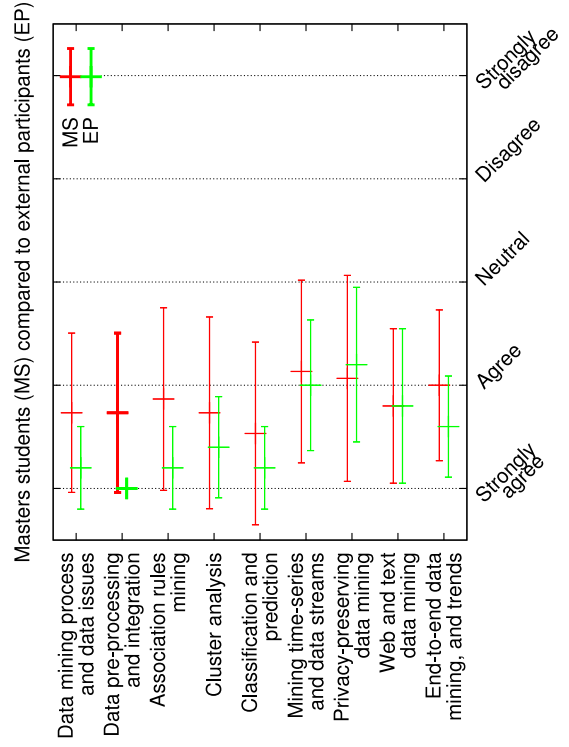
Tutorial papers

Tutorial	Papers discussed
1	<ul style="list-style-type: none"> - Data cleaning: Problems and current approaches - Methods for evaluating and creating data quality
2	<ul style="list-style-type: none"> - Fast algorithms for mining association rules - Selecting the right interestingness measure for association patterns
3	<ul style="list-style-type: none"> - On comparing classifiers: Pitfalls to avoid and a recommended approach - Classifier technology and the illusion of progress
4	<ul style="list-style-type: none"> - State-of-the-art in privacy preserving data mining - Names: A new frontier in text mining

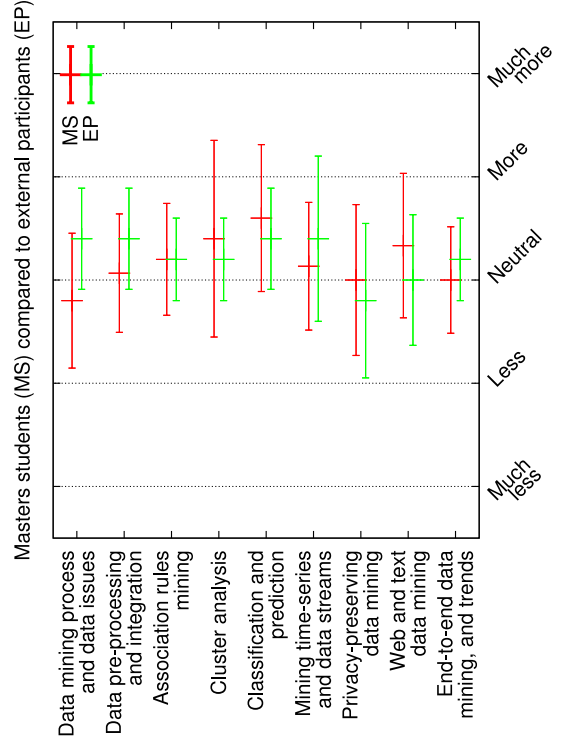
In general the course has been ...



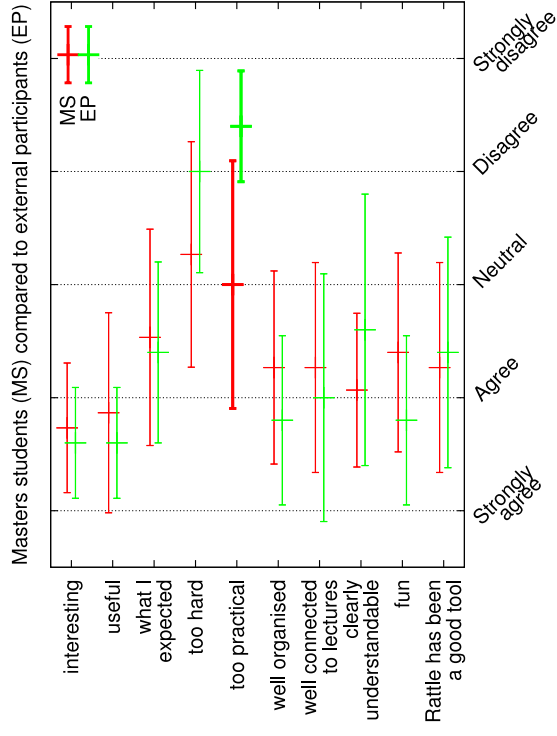
Interestingness of topics



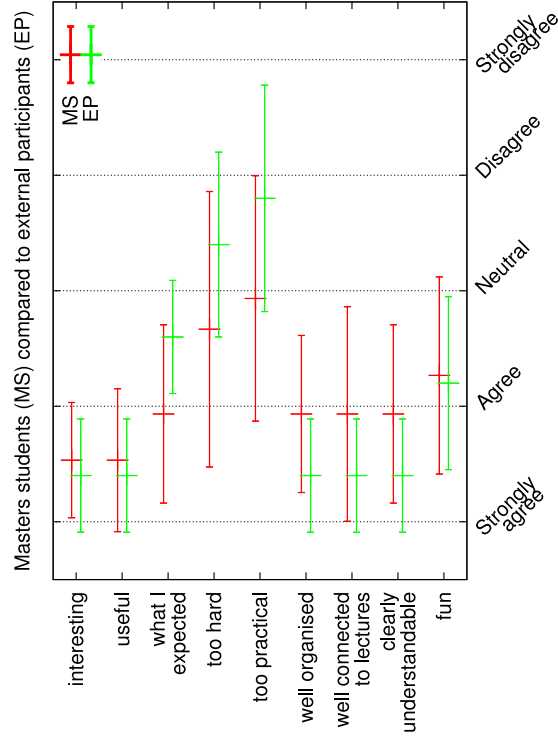
Coverage of topics



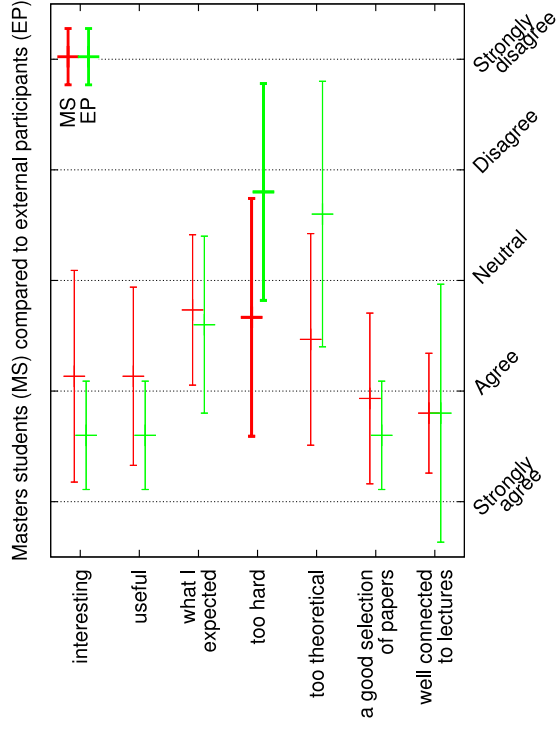
Practical laboratory sessions were ...



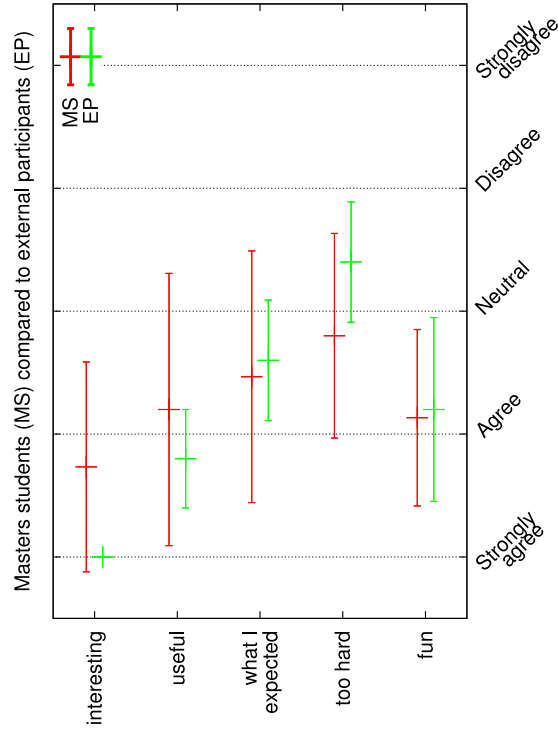
The assignments were ...



The tutorials were ...



The paper presentation was ...



Outlook

- Overall, the course has been well received by both Master students and external participants
- Excellent contributions from external participants, especially during tutorial discussions
- Future improvements include
 - Increase number of lectures so that topics can be covered in more detail
 - Provide list of specific questions and additional background material with tutorial papers
 - Longer student presentations (5 minutes is too short)
 - Allow more freedom in selecting data sets for assignments