

Web Intelligence

Course code: 2DV515

Dr. Johan Hagelbäck



johan.hagelback@lnu.se



<http://aiguy.org>



What is web intelligence?

- Today, we have huge amounts of data generated by lots of users available on the web:
 - Movie and song recommendations
 - Wikipedia
 - Web page recommendations
 - Product recommendations
 - ...
- The basic idea is:
 - How can we learn from and draw conclusions about data to create new products or services?



Example: Netflix

- Netflix recommends movies and TV shows that you might like based on what you previously have looked at
- In 2006, Netflix announced a prize of \$1 million to the first team that improved the accuracy of their recommendation system by 10%
- In 2007, the best team managed to improved the system by 7%
- In September 2009, the prize was awarded to the *BellKor's Pragmatic Chaos* team
- Clearly, there are huge benefits for better recommendation systems



Example: Google

- Before Google started, search engines only looked at the contents of a web page when presenting search results to users
- The founders of Google, Larry Page and Sergei Brin, took a completely new approach
- They ranked the search results based on contents and also on how popular a web site is
- This greatly increased the relevance of search results, and Google quickly became the number one search engine



Web intelligence

- These examples are exactly what web intelligence is
- We use data available on the web to create new products or services
- Web intelligence covers many fields:
 - Artificial Intelligence
 - Machine learning
 - Visualization
 - Information technology
 - Statistics
 - ...



The course



Lectures

- The course consists of 7 pre-recorded lectures:
 1. Introduction (this)
 2. Recommendation systems
 3. Clustering
 4. Search Engines
 5. What is Machine Learning?
 6. Data and Learning
 7. Naïve Bayes algorithm



Lectures

- Watch the lectures at your own pace
- Read the literature to get a deeper understanding of things
- Don't hesitate to ask questions about the contents, preferably on Slack

The screenshot shows a course page for 'WEB Intelligence'. The sidebar on the left contains the following sections:

- Course info**
 - About the course
 - Welcome letter
 - Forum
 - Deadlines and Submissions
 - Recommended Reading
 - Course Management
 - Syllabus
- Lectures**
 - L01 - Introduction
 - L02 - Recommendation Systems**
 - L03 - Clustering
 - L04 - Search Engines
 - L05 - What is Machine Learning?
 - L06 - Data and Learning
 - L07 - Naive Bayes
- Assignments**
 - A1 - Recommendation System
 - A2 - Clustering
 - A3 - Search Engine
 - A4 - Machine Learning
 - Datasets

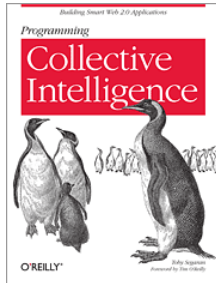
The main content area for 'L02 - Recommendation Systems' includes:

- Contents**
 - Introduction to Recommendation Systems
 - Similarity measures
 - Find matching users
 - Find recommended products
- Recommended reading**
 - Programming Collective Intelligence**
 - Chapter 2
 - Mining of Massive Datasets**
 - Chapter 3
 - Chapter 9
- Lecture slides**

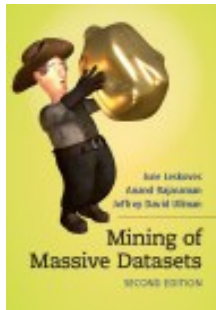
Available for download [here](#).
- Recording**

Is available [here](#).

Literature



- Programming Collective Intelligence
- Segaran, Toby (2008)
- Quite old and some examples are outdated, but it gives a good overview of the field and the algorithms are very well explained.



- Mining of Massive Datasets
- Leskovec, Jure & Rajaraman, Anand & Ullman, Jeffrey David (2014)
- Free e-book. Explains concepts in an easy way with lots of examples.

- ... and some online resources



Submissions

- There are four assignments and a project in the course
- For each assignment you have to:
 - Submit a link to the source code at GitHub (or similar)
 - Do an oral examination (physically or online)
- The assignments have goals for different grades – you choose what grade you aim for
- Submit at MyMoodle and sign up for an examination slot



Course pages

The screenshot shows a MyMoodle course page. At the top, there's a navigation bar with 'MyMoodle', 'Svenska (sv)', 'Mina senaste kurser', and 'Den här kursen'. A sidebar on the left lists categories: 'Datavetenskap', 'Computer Science', 'Medieteknik', and 'Media Technology'. The main content area has a header 'Öppna alla' and 'Fäll ihop alla'. Below that is a 'Resources' section with 'Course contents' and 'Oral examination form'. The 'Submissions' section lists 'Assignment A1', 'Assignment A2', and 'Assignment A3'. The page is annotated with 'Ämne 1' and 'Ämne 2'.

MyMoodle

WEB INTELLIGENCE

The screenshot shows a Coursepress course page. The main heading is 'L02 – Recommendation Systems', updated on 2018-10-05 by Johan Hagelbäck. The 'Contents' section lists: 'Introduction to Recommendation Systems', 'Similarity measures', 'Find matching users', and 'Find recommended products'. The 'Recommended reading' section includes 'Programming Collective Intelligence' (Chapter 2) and 'Mining of Massive Datasets' (Chapter 3 and Chapter 9). The 'Lecture slides' section is available for download. The 'Recording' section is also available. A sidebar on the left lists course info, welcome letter, forum, deadlines, recommended reading, course management, and syllabus. Below that is a 'Lectures' list from L01 to L07, and an 'Assignments' list from A1 to A4 and datasets.

Coursepress



Assignments

- Assignment 1:
 - Recommendation system for a small movie ratings data set
 - Implemented as a web application using a RESTful web service
- Assignment 2:
 - Clustering a dataset of blogs
 - Implemented as a web application using a RESTful web service
- Assignment 3:
 - Search engine for Wikipedia pages
 - Implemented as a web application using a RESTful web service
- Assignment 4:
 - Naïve Bayes machine learning algorithm for classification tasks

You can choose any programming language you like,
and any REST framework



Project

- There are four different projects you can choose from:
 - Recommendation System
 - Clustering
 - Web scraping and Search Engine
 - Machine Learning
- The projects build upon what you have done in the assignments
- Choose the project you find most interesting!



Grading

- The final grade on the course is calculated as follows:
$$\text{Score}_F = (\text{Score}_{A1} + \text{Score}_{A2} + \text{Score}_{A3} + \text{Score}_{A4} + 3.5 * \text{Score}_{\text{Project}}) / 7.5$$
- The final score is rounded to nearest integer
- Grades are converted to scores using the following conversion:

Grade	A	B	C	D	E
Score	5	4	3	2	1



Grading

- Example:
 - You receive the following grades:
A1: C, A2: A, A3: B, A4:A, Project: B
 - Your final grade is then:
 $(3 + 5 + 4 + 5 + 3.5 * 4) / 7.5 = 4.13 \approx 4$
 - According to the table, a score of 4 is a **B**



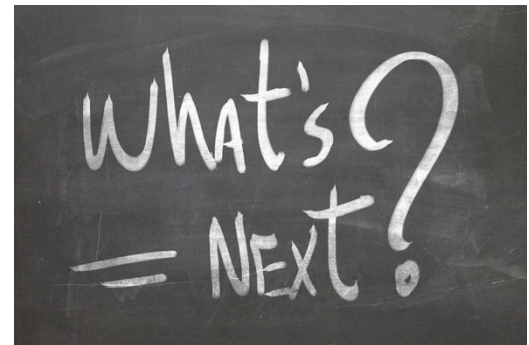
Communication

- Questions, discussions, etc. shall be on the Slack channel [#2DV515](#)
- Also use Slack for private messages with the course manager
- We will also use Slack for online oral examinations



What to do next?

- Watch lecture 2, read recommended reading and start working on the first assignment
- Keep track of the deadlines so you don't miss any submission
- Post questions and join discussions on Slack
- ... continue on the other lectures, assignments and the project!
- Give me a shout if you need help!



Web Intelligence

Course code: 2DV515

Dr. Johan Hagelbäck



johan.hagelback@lnu.se



<http://aiguy.org>

