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
Literature

• Programming Collective Intelligence
  • Seagaran, 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  • Quite old and some examples are outdated, but it gives a good overview of the field and the algorithms are very well explained.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)

• … 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

MyMoodle

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:
  \[ \text{Score}_F = \left( \text{Score}_{A1} + \text{Score}_{A2} + \text{Score}_{A3} + \text{Score}_{A4} + 3.5 \times \text{Score}_{\text{Project}} \right) / 7.5 \]

• The final score is rounded to nearest integer

• Grades are converted to scores using the following conversion:

<table>
<thead>
<tr>
<th>Grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Grading

• Example:
  – You receive the following grades:
  – Your final grade is then:
    \[ (3 + 5 + 4 + 5 + 3.5 \times 4) / 7.5 = 4.13 \approx 4 \]
  – 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