Administrative Details

Cornelia Caragea

Computer Science and Engineering
University of North Texas

June 9, 2016

Some Administrative Details

Instructor: Cornelia Caragea
My email: ccaragea@unt.edu
Prerequisites: basic knowledge on probability and statistics, data structures and algorithms
Class materials: will be made available at:
http://www.cse.unt.edu/~ccaragea/kdsin16.html

Required textbook:
Networks, Crowds, and Markets: Reasoning About a Highly Connected World
by David Easley and Jon Kleinberg

Available online at: http://www.cs.cornell.edu/home/kleinber/networks-book/
Required textbook:
Introduction to Information Retrieval
by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze
Cambridge University Press, 2008.

Online version available at: http://nlp.stanford.edu/IR-book/

Recommended Books for Machine Learning

- *Pattern Recognition and Machine Learning*, Christopher Bishop
- *Machine Learning: A Probabilistic Approach*, Kevin Murphy
- *Machine Learning*, Tom Mitchell
- *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Trevor Hastie, Robert Tibshirani, Jerome Friedman
  http://www-stat.stanford.edu/∼tibs/ElemStatLearn/

Syllabus

From Easley and Kleinberg textbook:

- Part I: Graph Theory and Social Networks (Chapters 1-5)
  - Graphs
  - Strong and Weak Ties
  - Positive and Negative Relationships
- Part IV: Information Networks and the Web (Chapters 13,14)
  - The Structure of the Web
  - Link Analysis and Web Search
- Part VI Network Dynamics: Structural Models (Chapters 19,20)
  - Cascading Behavior in Networks
  - The Small-World Phenomenon

Additional material published in journals and conferences.

Requirements

- Class attendance and participation! 15%
- Several homework assignments! 15%
- One exam! 20%
- A final project! 50%
  - Implementation project
  - Literature review project - one paper summary
  - Trying various machine learning and social network software on a benchmark dataset

Why Class Participation

- "We learn:
  10% of what we hear
  30% of what we see
  50% of what we see and hear
  70% of what we discuss
  80% of what we experience
  95% of what we teach others."
- "Teach me and I will forget;
  show me and I may remember;
  involve me and I will understand."
  Chinese Proverb.

Software Resources for Graphs

For visualization and basic network metrics:

- Gephi:
  - http://gephi.org/
  - Runs on Windows, Linux, Mac OS X.
  - Is open-source and free.
- Pajek:
  - http://vlado.fmf.uni-lj.si/pub/networks/pajek/
  - Runs on Windows.
  - Is free.
- NodeXL:
  - http://nodexl.codeplex.com/
  - Runs on Windows.
  - Is free, open-source template for Microsoft Excel 2007 and 2010.

Software Resources for Graphs

- Cytoscape:
  - http://www.cytoscape.org/
  - Runs on all platforms.
  - Is free, open-source, Java language based.
- NetworkX:
  - http://networkx.github.com/
  - Runs on all platforms.
  - Is free, open-source, Python language based.
- SoNIA:
  - http://www.stanford.edu/group/sonia/
  - Runs on all platforms.
  - Is free, open-source, Java language based.
- SNAP: Stanford Network Analysis Platform
  - http://snap.stanford.edu/snap/
  - Runs on all platforms.

Software Resources for Machine Learning

- Weka Java based ML package
- SVM Light
- LibSVM, LibLinear
- Mallet
- Torch, Theano, Caffe, Mocha: deep learning packages

Who Are You?

- Background?
- Research interests? List a few (research/CS/math/whatever) topics that interest you.
- Expectations from this course?