X-Informatics Case Study: e-Commerce and Life Style Informatics: Recommender Systems I

February 4 2013

Geoffrey Fox
gcf@indiana.edu

http://www.infomall.org/X-InformaticsSpring2013/index.html

Associate Dean for Research and Graduate Studies, School of Informatics and Computing
Indiana University Bloomington

2013

Big Data Ecosystem in One Sentence

Use Clouds running Data Analytics Collaboratively processing Big Data to solve problems in X-Informatics (or e-X)

X = Astronomy, Biology, Biomedicine, Business, Chemistry, Climate, Crisis, Earth Science, Energy, Environment, Finance, Health, Intelligence, Lifestyle, Marketing, Medicine, Pathology, Policy, Radar, Security, Sensor, Social, Sustainability, Wealth and Wellness with more fields (physics) defined implicitly

Spans Industry and Science (research)

Education: Data Science see recent New York Times articles
http://datascience101.wordpress.com/2013/04/13/new-york-times-data-science-articles/

Recommender Systems in more detail

- A large number of online and offline commerce activities plus basic Internet site personalization relies on “recommender systems”
- Given real-time action by user, immediately suggest new actions (as in Amazon buy recommendations on web)
- Based on past actions of users (and others) suggest movies to look at, restaurants to eat at, events to go to, books and music to buy
- Based on mix of explicit user choice and grouping of internet sites, present customized Google News page
- Given sales statistics, decide on discounts at “real” supermarkets and placement of related (by analysis of buying habits) products
- Identify possible colleagues at Social Networking sites like LinkedIn
- Identify matches between employers and employees at sites like

Everything is an Optimization Problem?

- **Fit** Model to Data
  - Higgs + Background
- **Match** User to Jobs or Books or Other Users?
- **Classification** is optimizing assignment of members of an ontology (list of categories) to data
- **Keeping alive in a jungle** involves optimizing one distance from hungry lions
  - One's own neural net optimizes interpretation of pixels detected by your eye to identify the lions you are trying to avoid
- **Typically minimize some function (or maximize negative of function)**
  - Interesting feature of these problems is ingenious choice of function
  - Note Physics minimizes (free) energy
  - This usually involves thinking of people and/or items as points in a space (not always a traditional vector space)

Next Part of Course

- Comments from online site http://recommenderbook.net/teaching-material/slides on goals of recommender system and overall approaches
- Introduction to Kaggle site
- Comments from Netflix http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial on their approach