# X-Informatics Case Study: e-Commerce and Life Style Informatics: Recommender Systems II: Algorithms

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

X-informatics

How Wealth Informatics can help with your financial freedom?

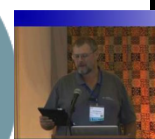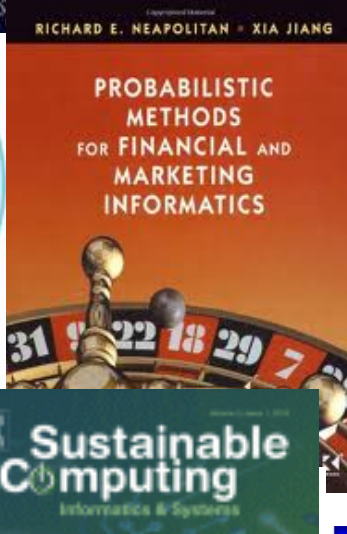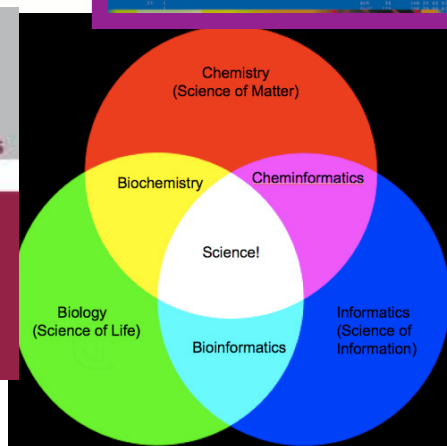Xinformatics

Earth Science INFORMATICS

Climate Informatics network

Biomedical Informatics
Computer Applications in Health Care and Biomedicine

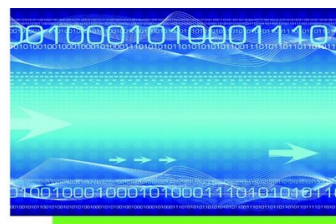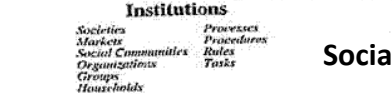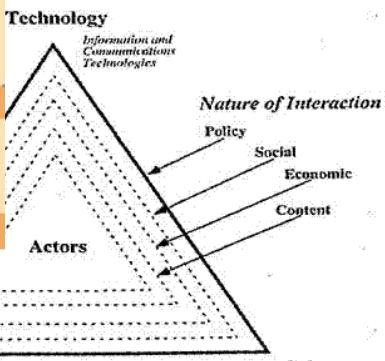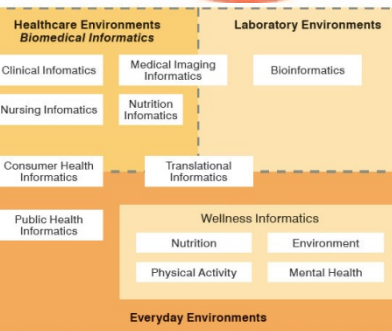Journal of Pathology Informatics

Paul Kantor · Gheorghe Muresan · Fred Roberts · Daniel D. Zeng · Fei-Yue Wang · Hsinchun Chen · Ralph C. Merkle (Eds.)

Intelligence and Security Informatics

Springer

AstroInformatics2012
Redmond, WA, September 10 - 14, 2012

Chemistry (Science of Matter)

Biochemistry

Cheminformatics

Science!

Biology (Science of Life)

Bioinformatics

Informatics (Science of Information)

RICHARD E. NEAPOLITAN · XIA JIANG

PROBABILISTIC METHODS FOR FINANCIAL AND MARKETING INFORMATICS

Research

NCICB Biomedical Informatics

Clinical Care

Bio-Informatics

Medical Informatics

Informatics

Health Management

Health Informatics

Health Information

Technology

Opportunities and Challenges in Crisis Informatics

Sustainable Computing
Informatics & Systems

USC Center For Energy Informatics

Home    Research    Publications    Smar

GEO Informatics
Knowledge for Surveying, Mapping & GIS Professionals

About the Center

Welcome to the Center For Energy Informatics (CEI) at USC, an Organized Research Unit (ORU) housed in the Viterbi School of Engineering. Energy Informatics is the application of info ene and

**Healthcare Environments**
**Biomedical Informatics**

| Clinical Infomatics | Medical Imaging Informatics | **Laboratory Environments** |
| Nursing Informatics | Nutrition Informatics | Bioinformatics |
| Consumer Health Informatics | Translational Informatics | |
| Public Health Informatics | | |

**Wellness Informatics**

| Nutrition | Environment |
| Physical Activity | Mental Health |

**Everyday Environments**

Technology

Information and Communications Technologies

Nature of Interaction

Policy
Social
Economic
Content

Actors

Institutions

Societies
Markets
Social Communities
Organizations
Groups
Households

Processes
Procedures
Rules
Tasks

Culture

Values
Norms
Talk
Discourse
Pop Culture
Artifacts

**Social Informatics**

Noelia Penelope Greer (Ed.)

**Business Informatics**
Information technology, Management,

policy informatics network

ASU School of Public Affairs
ARIZONA STATE UNIVERSITY

Lifestyle Informatics

| Applications of LI | Admission and registration |
| How is the training classified | VU Honours Programme |
| Occupation Pr | |
| Further study | |
| Student at the | |
| Watch the mov | |
| Studying Abro | |

Lifestyle Informatics: Let people li

The study Lifestyle Informatics is about s ... this bachelor including applied psycholog ... knowledge about language and informatic ... short better. Lifestyle Informatics: let peo ... Lifestyle Informatics

ENVIRONMENTAL INFORMATICS
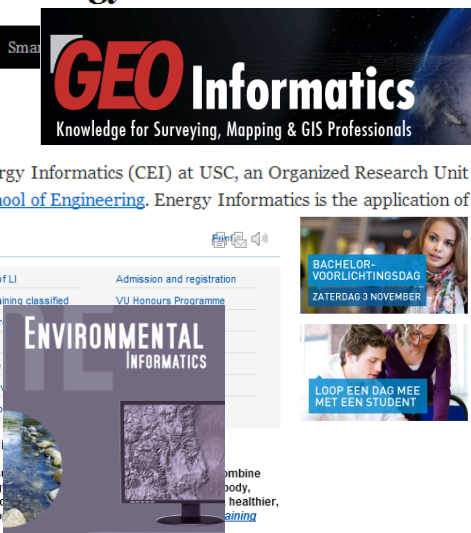
BACHELOR-VOORLICHTINGSDAG
ZATERDAG 3 NOVEMBER

LOOP EEN DAG MEE MET EEN STUDENT

# Overview of Problems

- Basic problem is personalized matching of items to people or perhaps collections of items to collections of people
- **People** to **products**: Online and Offline Commerce
- **People** to **People**: Social Networking
- **People** to **Jobs** or **Employers**: Job Sites
- **People+Queries** to the **Web**: Information Retrieval (search as in Bing/Google)

# http://en.wikipedia.org/wiki/Recommender_system

- When viewing a product on **Amazon.com**, the store will recommend additional items based on a matrix of what other shoppers bought along with the currently selected item.

- **Pandora** uses the properties of a song or artist (a subset of the 400 attributes provided by the Music Genome Project) in order to seed a "station" that plays music with similar properties. User feedback is used to refine the station's results, deemphasizing certain attributes when a user "dislikes" a particular song and emphasizing other attributes when a user "likes" a song. This is an example of a content-based recommender system.

- **Last.fm** creates a "station" of recommended songs by observing what bands and individual tracks that the user has listened to on a regular basis and comparing those against the listening behavior of other users. Last.fm will play tracks that do not appear in the user's library, but are often played by other users with similar interests. As this approach leverages the behavior of users, it is an example of a collaborative filtering technique.

- **Netflix** offers predictions of movies that a user might like to watch based on the user's previous ratings and watching habits (as compared to the behavior of other users), also taking into account the characteristics (such as the genre) of the film.

# The Google News personalization engine

# Google News portal (1)

- Aggregates news articles from several thousand sources
- Displays them to signed-in users in a personalized way
- Collaborative recommendation approach based on
  - the click history of the active user and
  - the history of the larger community
- Main challenges
  - Vast number of articles and users
  - Generate recommendation list in real time (at most one second)
  - Constant stream of new items
  - Immediately react to user interaction
- Significant efforts with respect to algorithms, engineering, and parallelization are required

# Google News portal (2)

- Pure memory-based approaches are not directly applicable and for model-based approaches, the problem of continuous model updates must be solved

- A combination of model- and memory-based techniques is used

- Model-based part: Two clustering techniques are used
  - Probabilistic Latent Semantic Indexing (PLSI) as proposed by (Hofmann 2004)
  - MinHash as a hashing method

- Memory-based part: Analyze story *co-visits* for dealing with new users

- Google's MapReduce technique is used for parallelization in order to make computation scalable

# Example: Markdown Optimization in Retail

Department stores and fashion retailers clear apparel inventory by reducing prices

"What prices do we set through the end of the season to meet inventory goals and maximize revenue?"

In practice

- Traditional pricing is based on business rules—not models
- Customers respond to prices, promotions at product-store level
- Weekly decisions are made at store level

gold star
**clearance**

Further reductions taken!
**60-80%**off
original prices.

§sas | THE POWER TO KNOW.

# Why an Analytical Solution?

- Huge number of decisions
  - 50,000 products/store x 2,000 stores = **100,000,000 decisions!**

- 1.5 - 3 terabytes of weekly data for two-year moving window
  - Number of units sold, price at point of sale, reference price
  - Starting and ending inventory in each store
  - Promotions, TV ads, circulars…

- Building blocks
  - Statistical model for demand
  - Price optimization

# Modeling Considerations

- Demand depends on
  - Seasonal effects, holidays
  - "Marketing mix" effects: price, promotion, inventory
  - Product life cycle effect : introduction and phase-out

- Pricing rules and customer behavior are complex
  - Magic price points (80% is lowest markdown)
  - Discontinuities in response (difference between $1.99 and $2.00)
  - Holiday effects are confounded with promotions

- Sales data are sparse for a product within a store

# Sparse, Noisy Data at Store-Product Level



Sales for Product 35649 at Store A
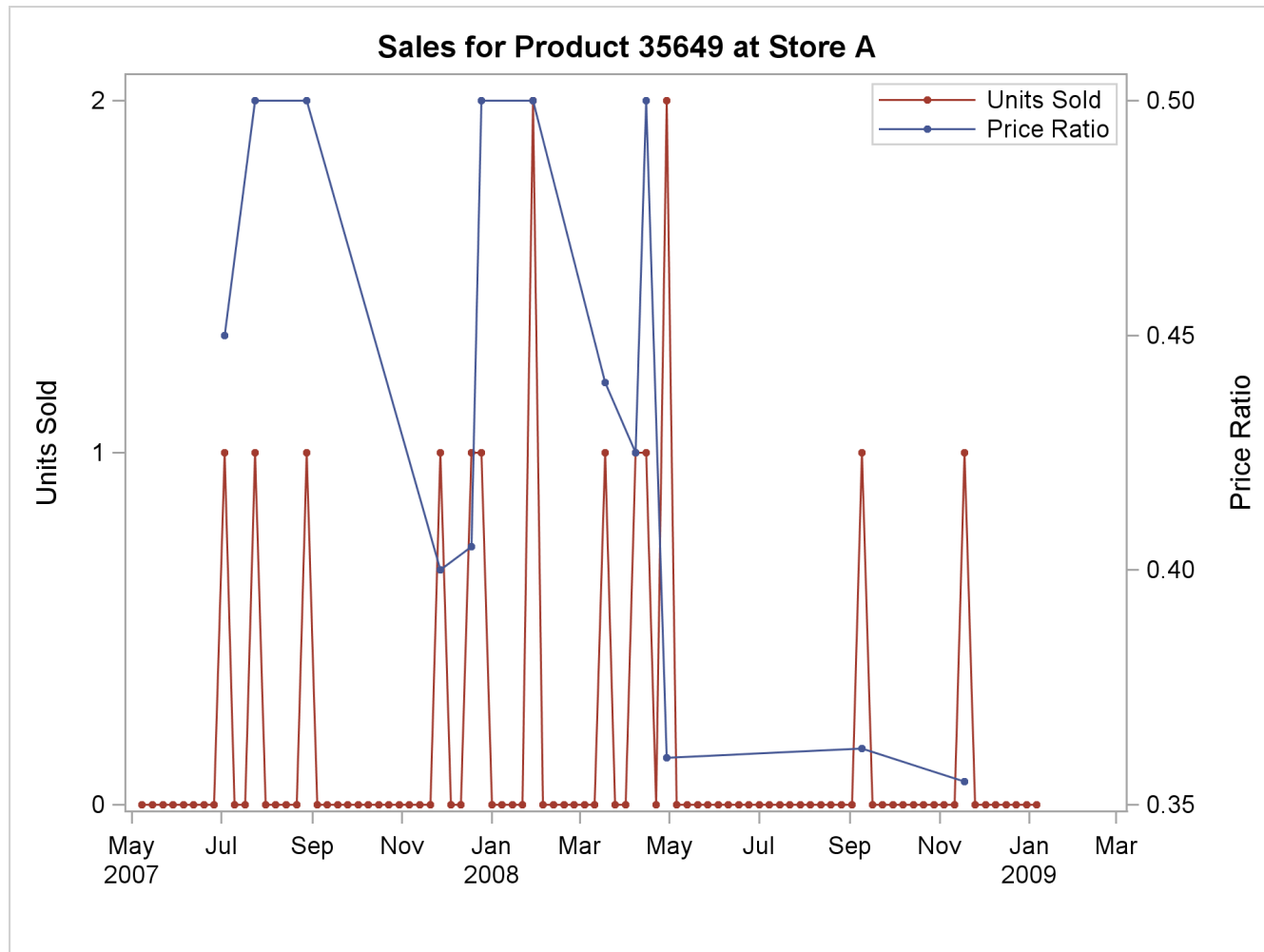
# What Can You Do With this Data?

- How do you estimate the components of demand?

- The key lies in analytical hierarchies for
    - Products
    - Store locations

# Data Aggregated at Region Level



Sales for Product 35649 in Region