Evaluation and Assessment of Recommenders Using Monte Carlo Simulation

July 17, 2012

Renato A. C. Capuruço and Prof. Dr. Luiz F. Capretz

PhD candidate/SE, r.capu@uwo.ca

University of Western Ontario,

Department of Electrical and Computer Engineering

London, Ontario, CANADA

Western Engineering

Outline

- Introduction
- Uncertainty on SRS
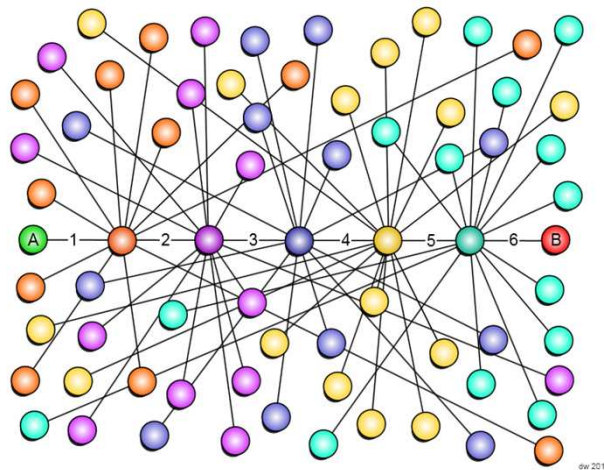- Monte-Carlo Technique/Modeling
- Experiment and results

Introduction

- ## Recommender Performance:

  - *Quantitative methods* collect numerical data and analyze it using statistical methods - relying on precise measurement outcome to yield conclusions

  - Evaluation metrics: coverage and accuracy metrics

    - precision, recall and F1-measure

    - mean absolute error (MAE) and root mean square error (RMSE)

- ## Sources of Uncertainty:

  - dataset chosen for testing, and data sparseness due to new users or few ratings (cold start)

  - Input data representation

  - similarity computation – several algorithms

  - Custom transformations to traditional approaches

… Sources of Uncertainty　　　　　　　(cont'd)

on Social Recommenders
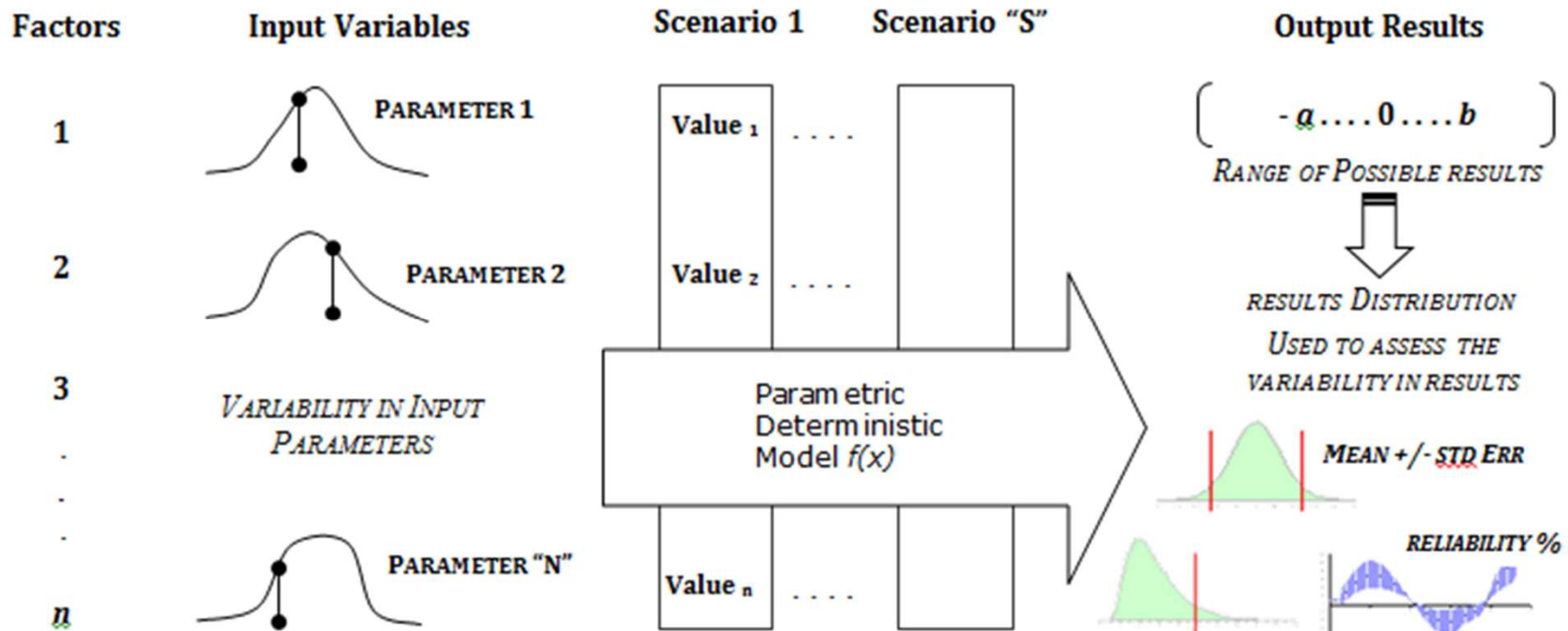

friendship formation mechanism

Relationship building patterns
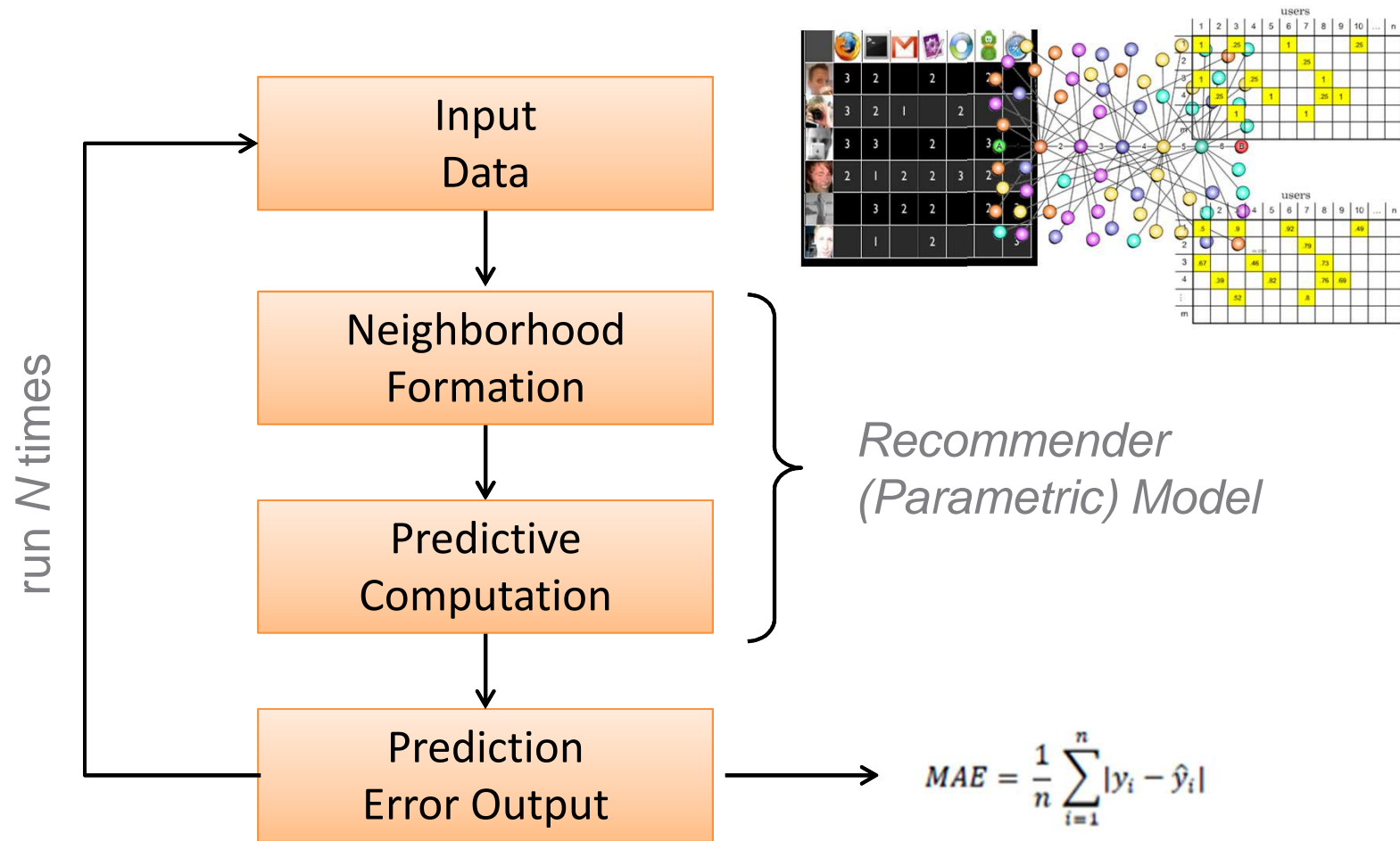
Dynamic process

Quality of recommendations

Snapshot

Monte-Carlo Technique

Summary Statistics: measures of *Location*, *Dispersion*, *Shape*, and *Order*

Monte-Carlo Modeling

run *N* times

Input Data

Neighborhood Formation

Predictive Computation

Prediction Error Output

Recommender (Parametric) Model

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Experimental Results

- Dataset

- Evaluation metrics and results – model robustness
  - Summary Statistics: mean, deviation, kurtosis, skewedness, Percentiles, quartiles, intervals, overall performance probability
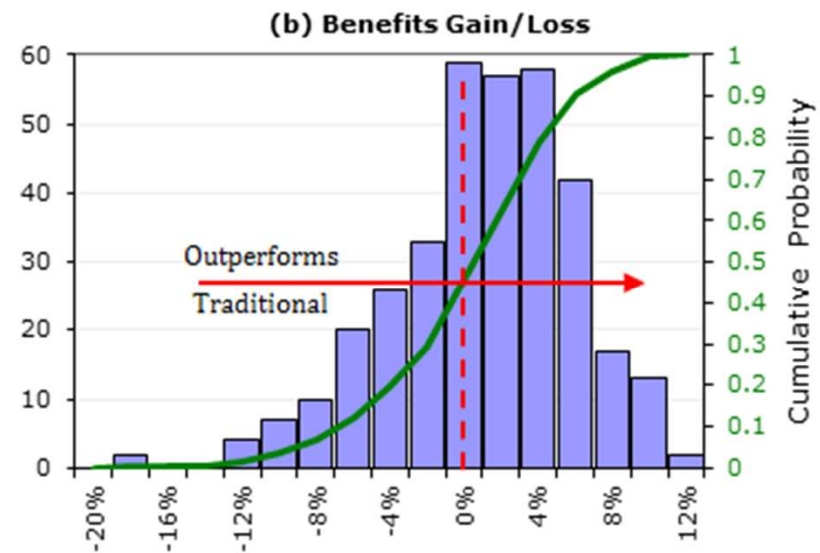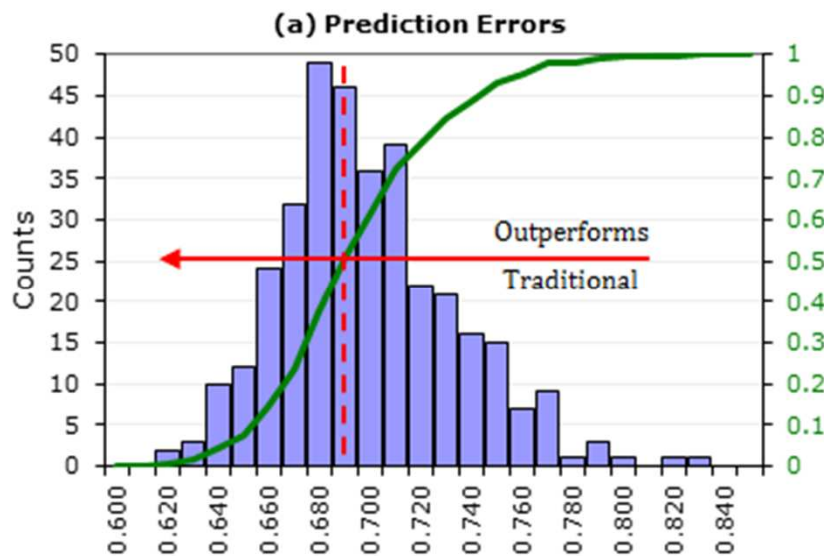
Conclusions

- Traditional Evaluation – Deterministic
- Monte Carlo Evaluation – Probabilistic
  - improved understanding, higher confidence, longer lasting value, and better depiction of recommender predictions – model robustness vs. performance
  - Can be employed on any recommender implementation

- Future Work / Challenges
  - extend the evaluation modeling strategy to account for effects of *data input representation*, different *evaluation metrics*, *similarity calculation* algorithms, etc.

Merci

- Any questions or comments ?