A Literature Review: Qualitative Evaluation of Recommendation Systems for Social Networks Based on Data Quality Issues

CS846 - Software Eng. in Big Data
Course project
Cheriton School of Computer Science

Nazli Medghalchi and Ahmed Aljmiai

Department of ECE and CS
University of Waterloo
Outline

• Introduction
  • Social Network-based Recommender
  • Location-based Social Network Recommender
• Problem
• Motivation
• Goal
• Approach
• Expected Results
• Future Work and Contribution
• References
Introduction

- Decision making within big data requires a recommender system
- Recommendation systems e.g. search engines, friend suggestions, local restaurant suggestions, e-government, e-business, and, etc.

Image source: [top] https://smallbiztrends.com/2016/05/popular-social-media-sites.html
Introduction

• Traditional methods for recommendation systems:
  
  • Content-based (CB)
  
  • Collaborative filtering (CF)
  
  • Knowledge-based (KB)
  
  • Hybrid — Many recommendation systems often rely on a combination of techniques
    
    • Linear combination
    
    • Sequential combination
Introduction

- Advanced methods for recommendation systems:
  - Social network-based recommender system
  - Fuzzy recommender
  - Context awareness-based recommender
  - Group recommender systems
  - Location-based social network recommender
Social Network-based Recommender

- Type of relationship (e.g. popular people, friends, classmates, co-workers, neighbours, relatives etc.)
- Similarity
- geo-location
- Weight-value
Location-based Social Network Recommender

- User (reviewer) popularity
- Local-people behaviours
- Tourists’ behaviours
- User’s friends/relatives
- User’s history
- Time-based
Location-based Social Network Recommender

- Tourists / Local-people behaviours
  - Place
  - Weekday or weekend
  - Time (e.g. 8:00 am, 12:00 pm)
Location-based Social Network Recommender

• Friends/relatives recommendation
  • Used in many social media
  • According to [Ye & et al.] 87.7% of friends share nothing.
  • Among 96% of friends, only 10% places is shared/visited.
    • proximity
    • interests
Problem

• Popular Reviewers Trustworthiness
  • Are they always accurate and serious?
  • Can we always trust them?
  • Do we trust them as our closest friends & relatives?
  • Should we completely ignore non-popular reviewers?
  • Who do we trust more tourists or local people?
Problem

• Recommendation systems and big data problems:
  
  • **Cold-Start**: Lack of sufficient new user/item history in recommendation systems
  
  • **Sparse Matrix**: Dealing with sparse Collaborative Filtering Matrix (NaN values)
  
  • **Gray Sheep Users**: The level of recommendation predictiveness
  
  • **Redundancy**: Over-specialization known as redundancy in suggestion (flexibility)
  
  • **Scalability** of algorithm based on data
Problem

- Data quality regarding data coming from many sources as
  - Data from the Internet of Things
  - Data collected by various industries
  - Scientific experimental and observational data
Problem

• How to choose a most efficient approach as recommender systems based on data quality, algorithm strengths and trustworthiness?
Motivation

- Many papers on certain method on certain data set but less focused on data quality

- Previous literatures focus on a quantitative evaluation of recommendation systems in the context of concrete social network databases
Goal

• Qualitative evaluation on location-based and Social network-based methods used as in Yelp social network
Approach

• Review literatures on comparing recommender algorithms specifically advanced methods

• Observing differences on recommender systems results regarding their application on social media data sets e.g. Yelp data set
Expected Results

• Recommendation systems algorithms accuracy and effectiveness relies on data set quality regarding application

• Hybrid methods can solve the problem of less accuracy for some data sets (e.g. gray-sheep user problem)

• Observing the effectiveness of certain methods on data sets with certain quality level

• SVM regression method is more accurate on sparse data set than CF based method
Future Work and Contribution

• Suggesting comprehensible framework of qualitative evaluation of recommendation systems methods

• Providing solid information on relation between data quality factors and useful recommendation systems
References


2. F. EL Jamiy, A. Daif, M. Azouazi and A. Marzak, “The potential and challenges of Big data - Recommendation systems next level application”, Hassan II University, Casablanca, Morocco.


4. Z. Xu, X. Wei, X. Luo, Y. Liu, L. Mei, C. Hu, L. Chen, “Knowle: A semantic link network based system for organizing large scale online news events”, The Third Research Institute of Ministry of Public Security, Shanghai, China, Shanghai University, Shanghai, China, Tsinghua University, Beijing, China, Shanghai Institute of Technology, Shanghai, China, Future generation computer systems 43-44(2015) 40-50.


References


14. Zh. Xiang, Z. Schwartz, J. H, Gerdes Jr., M. Uysal, “What can big data and text analytics tell us about hotel guest experience and satisfaction?”, Department of Hospitality and Tourism Management, Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, USA, Department of Hotel, Restaurant & Institutional Management, University of Delaware, Newark, DE 19716, USA, Department of Integrated Information Technology, College of Hospitality, Retail, & Sport Management, University of South Carolina, Columbia, SC, USA.

15. Á. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, E. Herrera-Viedma, “A quality based recommender system to disseminate information in a university digital library”, University of Granada, Department of Computer Science and Artificial Intelligence, Granada, Spain, University of Jaén, Department of Computer Science, Jaén, Spain, Department of Librarianship and Information Science, Granada, Spain, Information Sciences 261 (2014) 52–69.

References

17. J. He and W. W. Chu, “A Social Network-Based Recommender System (SNRS)”, Computer Science Department University of California, Los Angeles, CA.


References


Thank you!

Questions?
Terminology

• **Data Quality**: Data quality refers to the level of quality of data. There are many definitions of data quality but data are generally considered high quality if "they are fit for their intended uses in operations, decision making and planning."

• **Qualitative Approach**: is primarily exploratory research. It is used to gain an understanding of underlying reasons, opinions, and motivations.

• **Recommendation systems**: Basically filtering information, that can filter unseen information and can predict whether a user would like a given resource.

• **Collaborative filtering**: 1) Memory-based approaches 2) Model-based approaches
Terminology

- **Content-based (CB):** Content-based (CB) recommendation techniques recommend articles or commodities that are similar to items previously preferred by a specific user.

- **Collaborative filtering (CF):** to make choices based on the opinions of other people who share similar interests [19].

- **Knowledge-based (KB):** Knowledge-based (KB) recommendation offers items to users based on knowledge about the users, items and/or their relationships.

- Importance degree, similarity degree, satisfaction degree, preference degree.
Terminology

- Precision, Recall

- F1 measure: \[ F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \]