Browser-Oriented Universal Cross-Site Recommendation and Explanation based on User Browsing Logs

Yongfeng Zhang, Tsinghua University
zhangyf07@gmail.com

Outline

- Research Background
- Research Topic
- Current Achievements
- Research Plans

Personalized Recommender Systems

- Personalized Recommender Systems
  - Attempts to recommend the items of potential interests
- Widely integrated into many commercial systems
  - Especially the many online shopping websites
- Help to increase the online traffic and profits
  - Amazon online book shop: 30% profits comes from RS
  - Forrester: 1/3 of the customers takes the recommendation when they noticed them

Related Work

- Content-based Recommendation
  - Content-based Recommender Systems [Pazzani2005]
  - Auto Profile Construction for Recommendation [Sugiyama2004]
- Collaborative Filtering based Recommendation
  - User-based Collaborative Filtering [Resnick1994]
  - Item-based Collaborative Filtering [Sarwar2001]
  - Matrix Factorization based Collaborative Filtering [Koren2009]
- Hybrid Recommendation models
  - Hybrid Recommender Systems [Burke2002]
  - Content-based, Collaborative Recommendation [Marko1999]

Problems

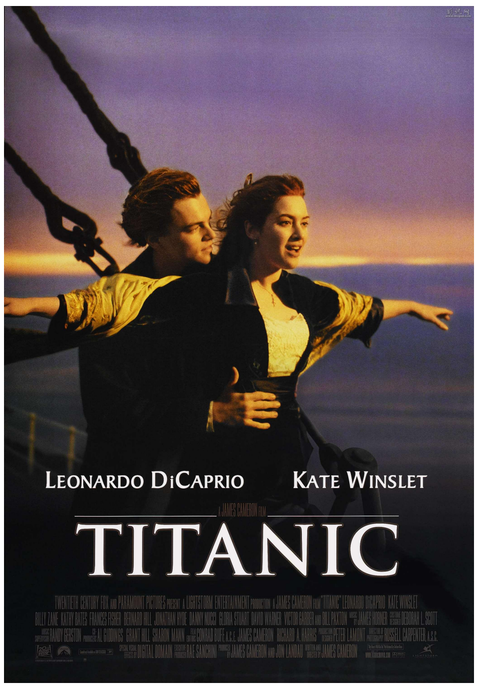
- They still focus on the vertical recommender systems
  - Although RS is getting more and more noticed
  - Still mostly restricted in inner-site/domain recommendation
    - Product recs in online shopping
    - Related article recs in online medias
    - Video, movie or music recs
  - The recommendation engine of the Web mainly consists of many independent vertical recommenders.

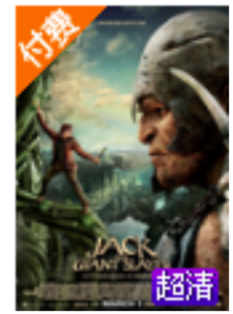
Research Topic

Browser-Oriented Universal Cross-Site Recommendation and Explanation based on User Browsing Logs

Understanding Universal Recommendation

- Universal recommendation: A CASE STUDY.





Homogeneous items from the same sites

Heterogeneous items from other websites.

Understanding Universal Recommendation

- Major characteristics of vertical recommendation
  - Single domain： Recs are usually from the same product domain
  - Inner-site： Recs are usually the products/items from the inner-site
  - Additional： Usually comes in the form of an additional application
- Major characteristics of universal recommendation
  - Cross-domain: The ability to recommend items from other domains
  - Inter-site: Recommended items are not necessarily from the same site
  - Fundamental: Comes in the form of a fundamental application when the user is surfing online

Why to Construct Universal Recommenders

- Why inter-domain/site
  - Vertical recs are homogeneous
    - Single domain / Inner-site
  - This applies to most current RS
    - E.g. Related video / article recommendation

- Underlying problem
  - Difficult to discover user needs from other potential aspects
  - Further restricts the application and business model of RS

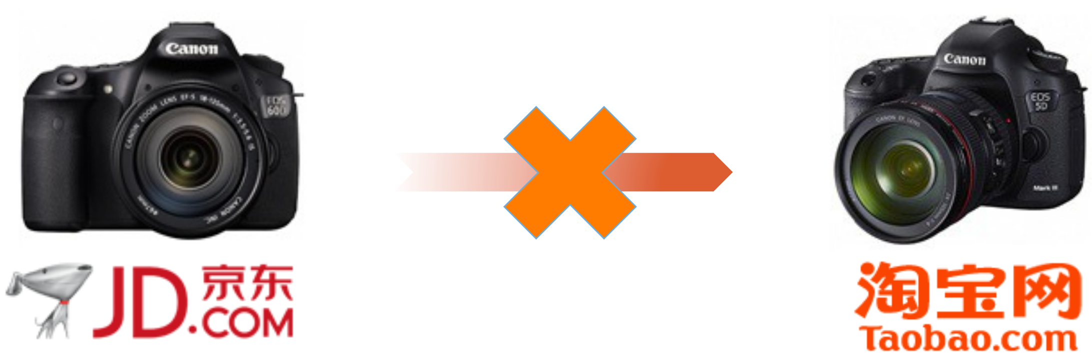
Music?

Purchasing?

Games?

Difficulties of Universal Recommender

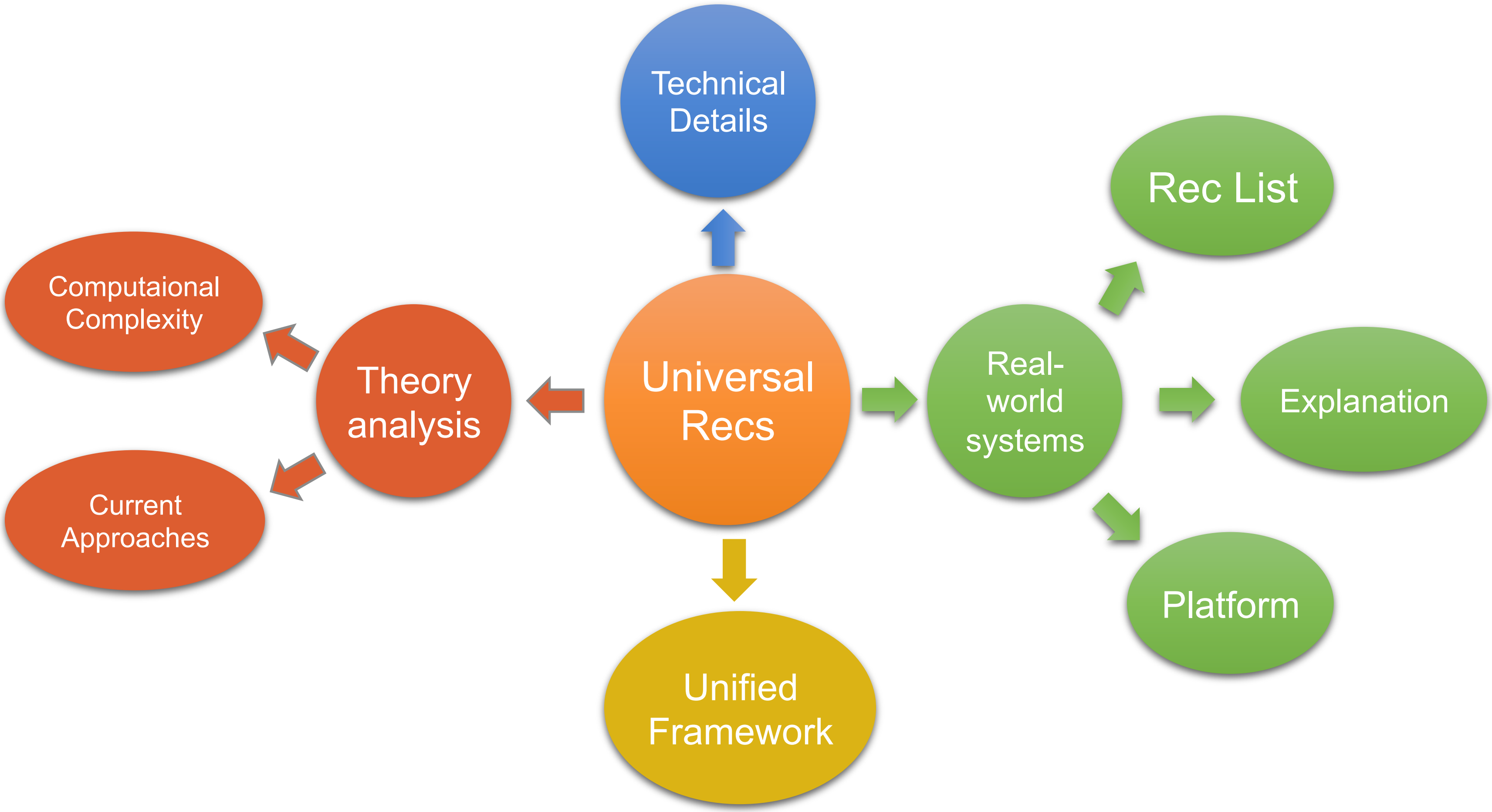
- Difficulties of constructing a universal recommender system
  - Lack of inter-site user behavior data
  - Restrictions from the business aspects



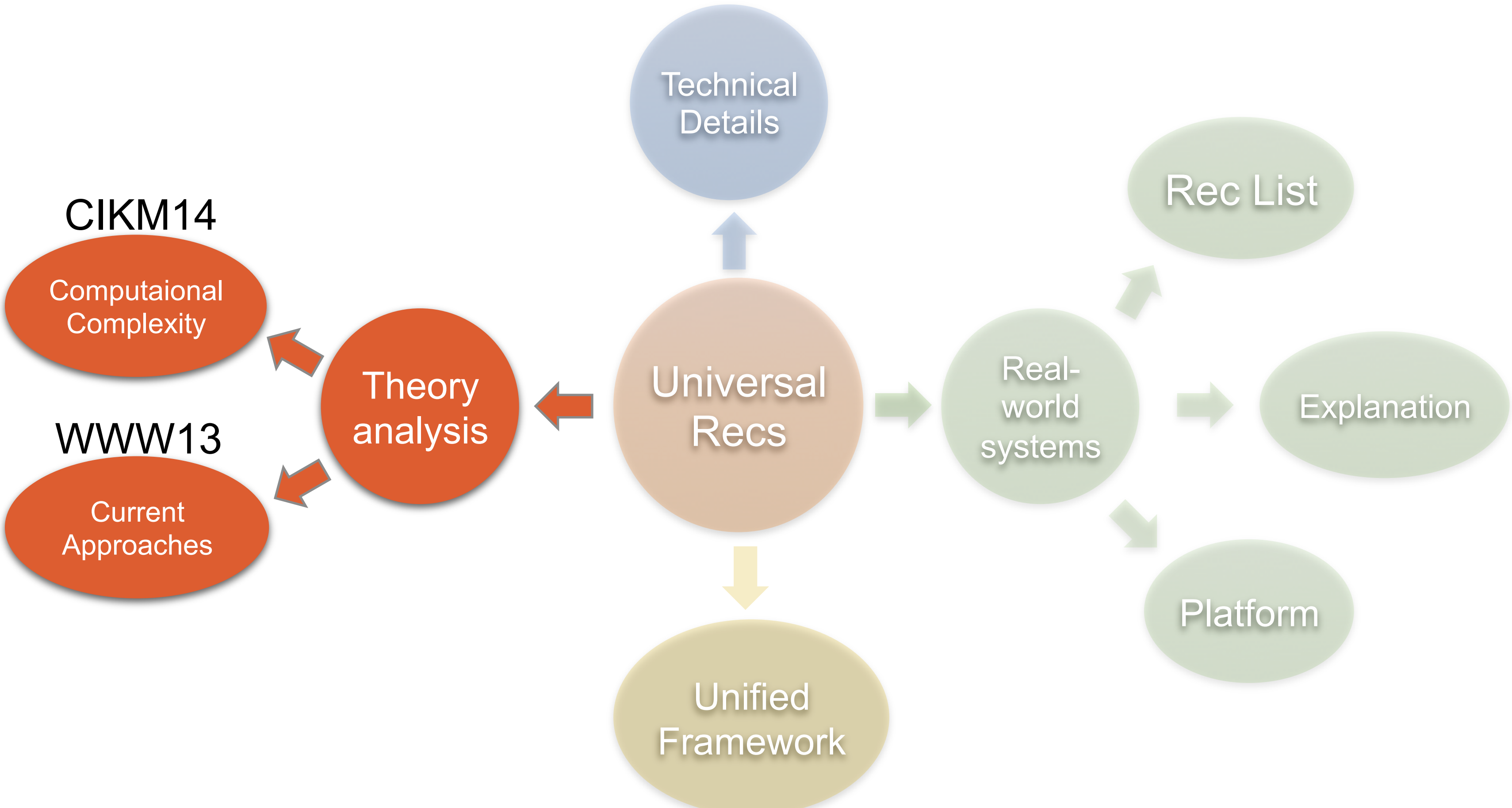
The Solutions

- Corresponding solutions to the problems
  - Incorporating search engine / browser log data
  - Browser-oriented recommendation: which offers a recommendation platform independent of a specific web

The Approach



Theoretical Analysis (CIKM14 & WWW13)



Theoretical Analysis (CIKM14 & WWW13)

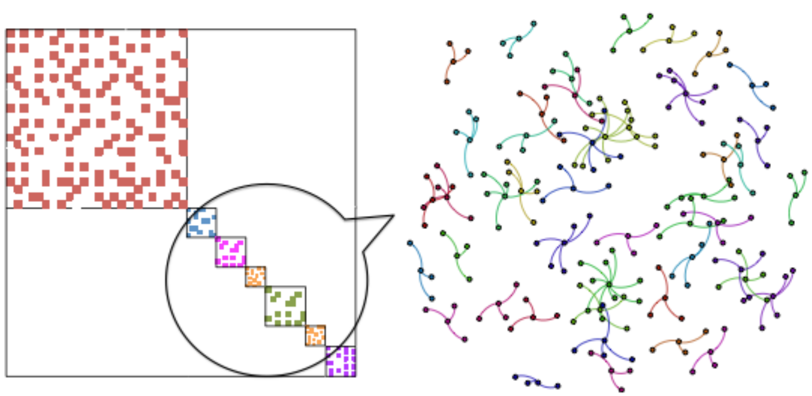
- Universal recommendation from a technical point of view
  - The classical rating prediction problem for universal rec
    - Rating prediction on a Block Diagonal Form (BDF) matrix

$$
\begin{array}{c|cccc|}
 & \mathcal{J}_1 & \mathcal{J}_2 & \cdots & \mathcal{J}_k \\
\hline
\mathcal{I}_1 & D_1 & & & \\
\mathcal{I}_2 & & D_2 & & \\
\vdots & & & \ddots & \\
\mathcal{I}_k & & & & D_k \\
\hline
\end{array}
$$

  - Lack of inter-site user behavior data
    - Which means that these is no data in off-diagonal areas

Theoretical Analysis – Computational Complexity (CIKM14)

- Lack of inter-site data brings severe problems [CIKM14]
  - Nearest-neighbors methods are invalid: no way to compute similarity
  - CF based on Matrix Factorization are also invalid
    - We prove these are at least O(r!) equal valued minima， r=50~100
    - Predictions for off-diagonal areas given by MF are meaningless



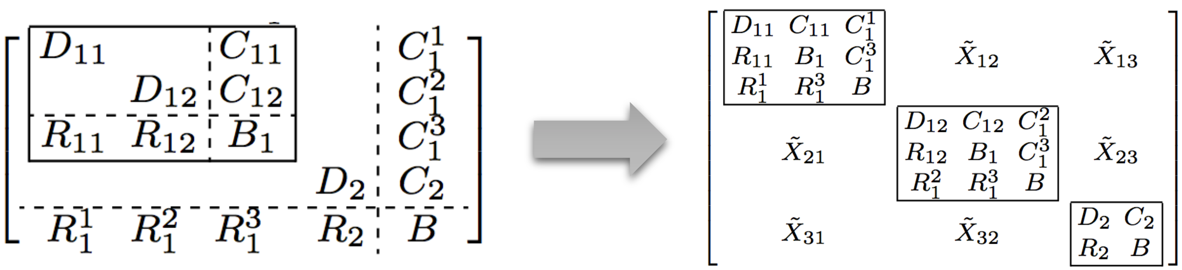
(a) Yelp dataset Matrix (b) Yelp dataset Graph

$$\begin{array}{c} \\ \mathcal{I}_1 \\ \mathcal{I}_2 \\ \vdots \\ \mathcal{I}_k \end{array} \begin{array}{c} \begin{array}{cccc} \mathcal{J}_1 & \mathcal{J}_2 & \cdots & \mathcal{J}_k \end{array} \\ \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_k \end{bmatrix} \end{array}$$

[1]CIKM'14,Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservebles.

Theoretical Analysis – Parallelization (WWW13)

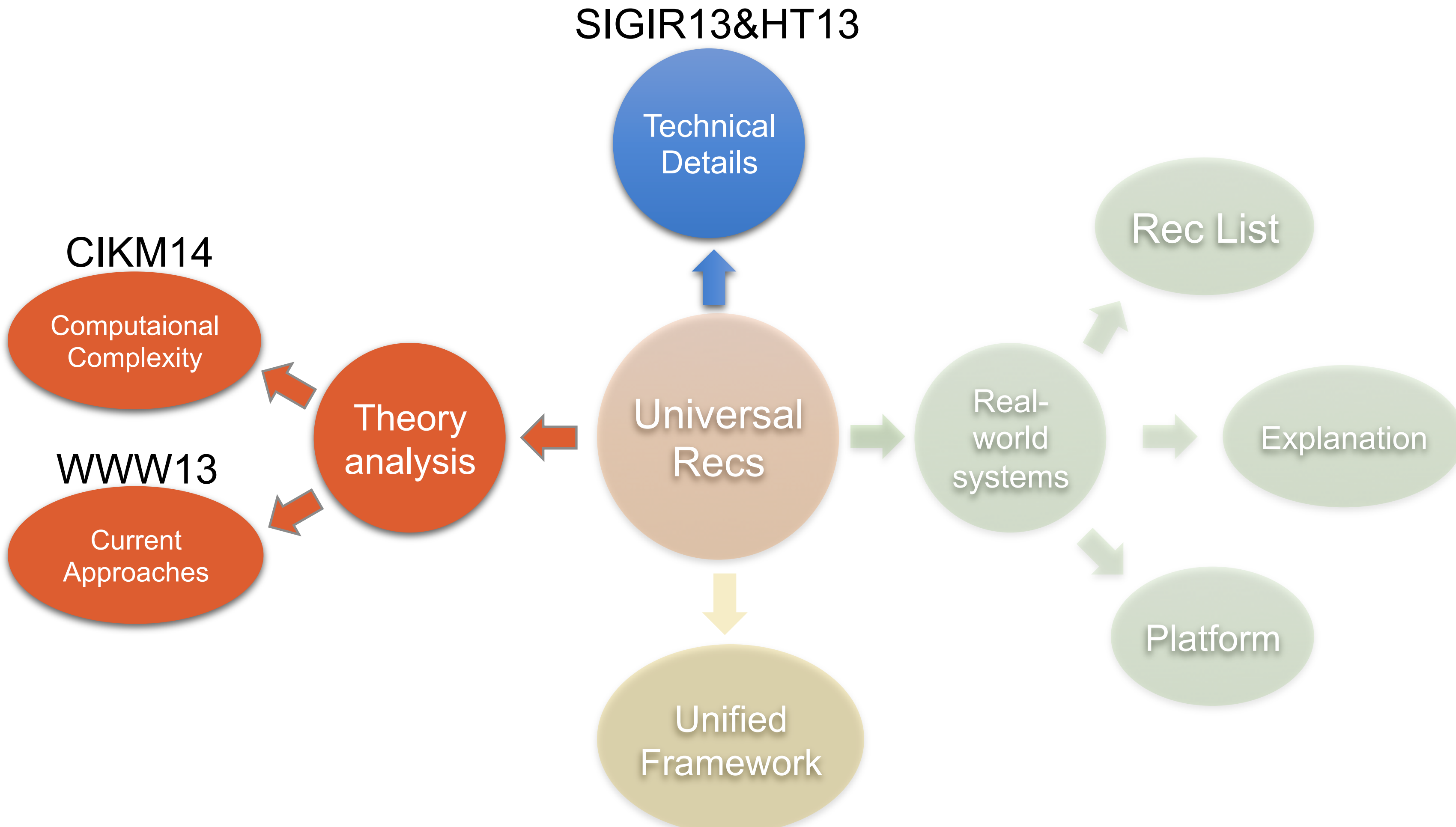
- Localized Matrix Factorization (LMF) based on BDF matrices [WWW13]:
  - Many commonly used MF algorithms (SVD/NMF) can be parallelized in LMF
  - Equal to its single routine algorithm, rather than its approximation
    - Which offers a unified framework for large-scale parallel MF and CF



$$X = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_k \end{bmatrix} \approx f(UV^T) = f\left(\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{bmatrix} \begin{bmatrix} V_1^T & V_2^T & \cdots & V_k^T \end{bmatrix} \right) = f\left(\begin{bmatrix} U_1V_1^T & U_1V_2^T & \cdots & U_1V_k^T \\ U_2V_1^T & U_2V_2^T & \cdots & U_2V_k^T \\ \vdots & \vdots & \ddots & \vdots \\ U_kV_1^T & U_kV_2^T & \cdots & U_kV_k^T \end{bmatrix} \right)$$

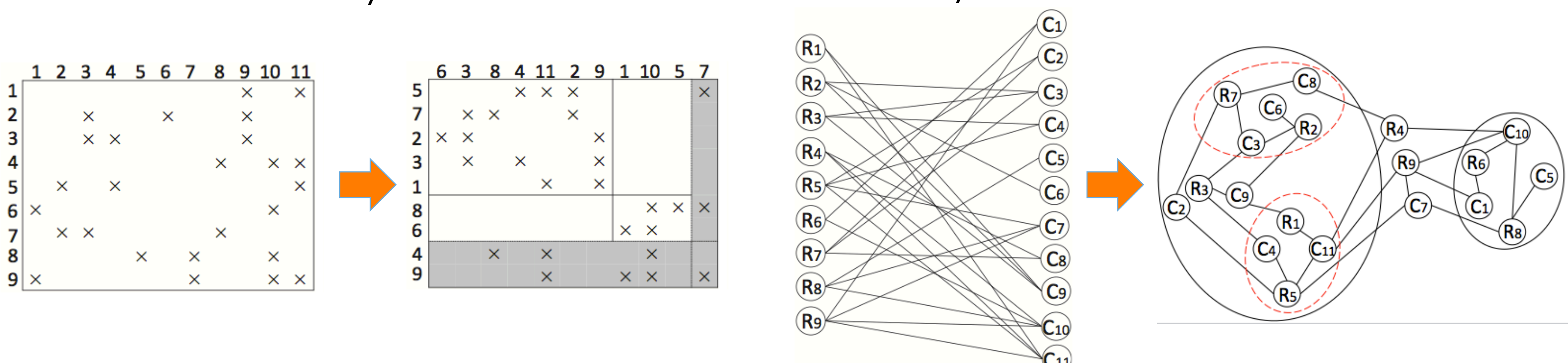
[1]CIKM'14,Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservebles.
[2]WWW'13, Localized Matrix Factorization for Recommendation based on Matrix Block Diagonal Forms.

Technical Details (SIGIR13 & HT13)



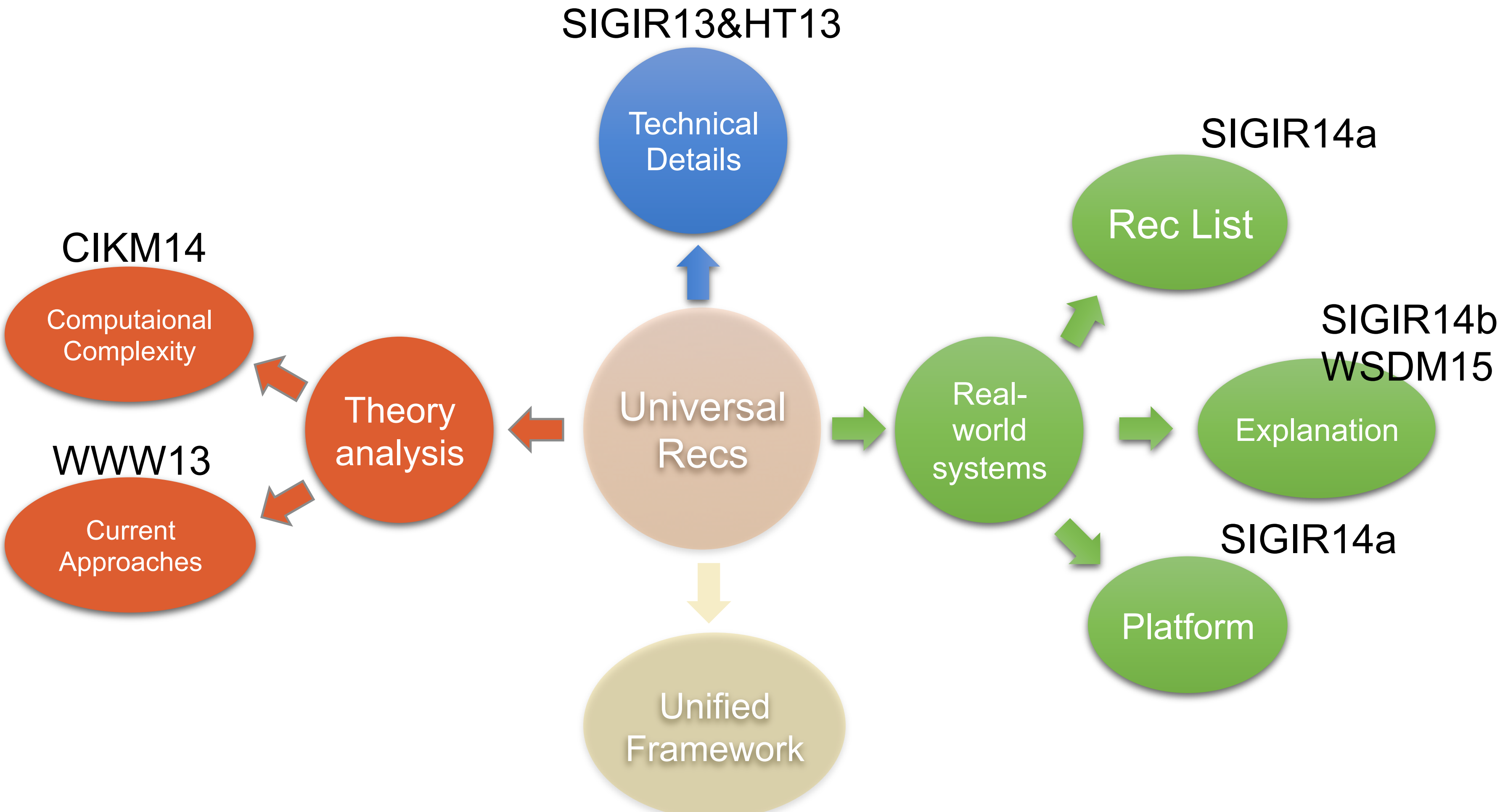
Technical Details (SIGIR13 & HT13)

- Use LMF in real-world systems [SIGIR13,HT13]
  - We prove that the bordered BDF structure on matrices is equal to conducting community detection on its bipartite graph
  - Which offers an intuition of the application of LMF in real-world systems
  - And provides a unified mathematical framework for the application of community detection in recommender systems



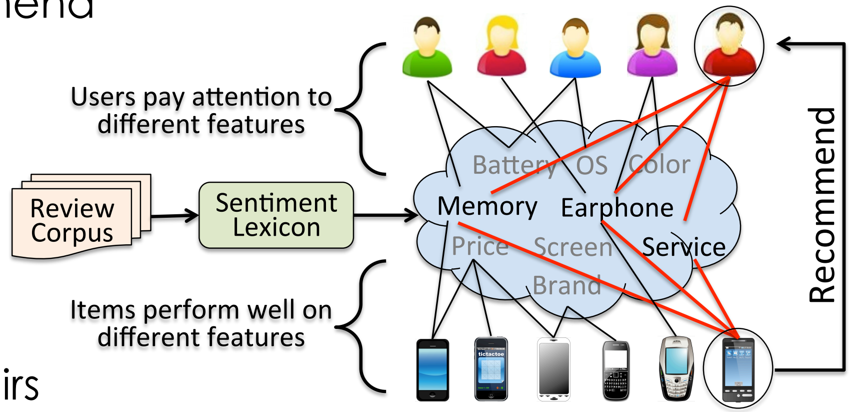
[1]SIGIR'13, Improve Collaborative Filtering through Bordered Block Diagonal Form Matrices.
[2]Hypertext'13, A General Collaborative Filtering Framework based on Matrix Bordered Block Diagonal Forms.

Real-world System (SIGIR14a & SIGIR14b)



Realization in Real-world Systems – the Recommendation Explanation (SIGIR14a/b)

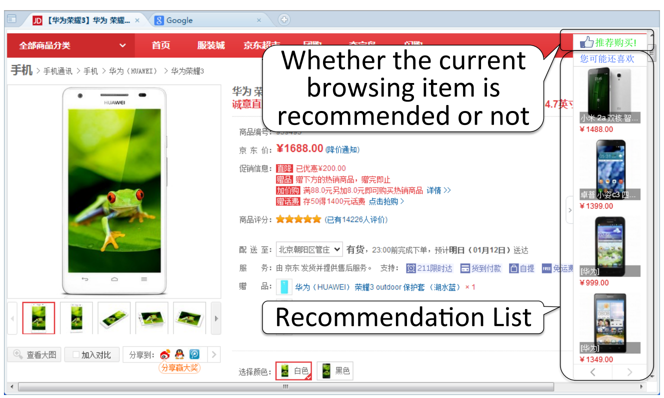
- Recommendation Explanations [SIGIR14a]
  - Seems more important in universal recommend
  - To persuade a user to examine a rec in an unfamiliar website
- Phrase-level sentiment analysis [SIGIR14b]
  - Mine the product features automatically
  - Extract Product Feature – User Opinion pairs
    - e.g. Camera Lens – Long
  - Construct the feature-level recommendation explanations automatically
- Examples
  - This camera performs well on Lens, which feature you may concern
  - We know the concerned features of a user from his/her historical reviews



[1]SIGIR 2014a, Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis.
[2]SIGIR 2014b, Boost Phrase-level Polarity Labeling with Review-level Sentiment Classification.

Realization in Real-world Systems - Browser-Oriented Recommendation (SIGIR14a)

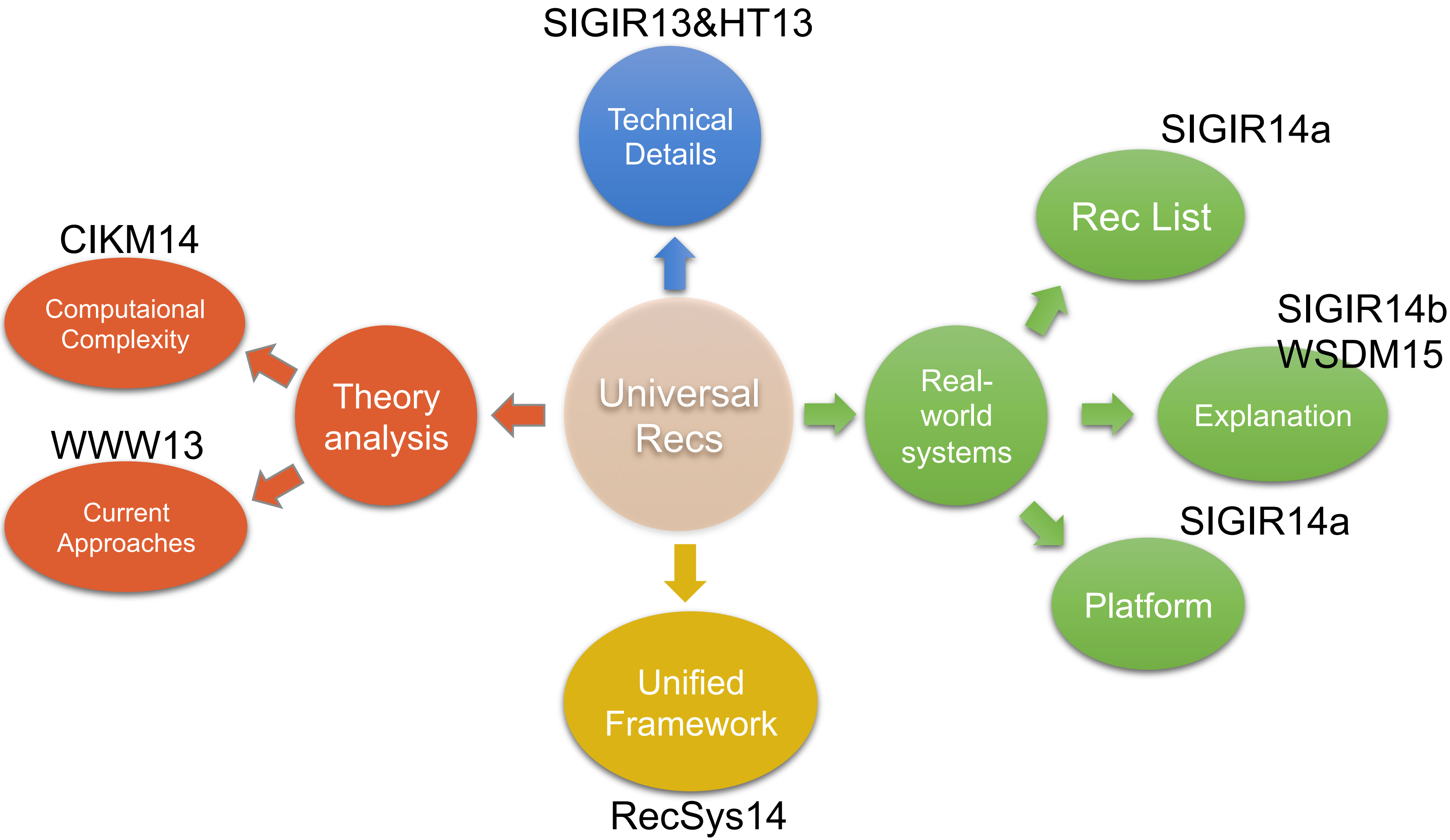
- How to provide cross-site recommendation?
  - Provide recommendations by web browsers directly! [SIGIR14a]
  - An independent recommendation platform from specific websites.
    - To solve the problem that the websites have no intention to provide recommendations from other sites.
    - Also offer us brand new business models
  - Browsers further enrich cross-site user behavior data with browsing logs
    - Help to better understand the user needs, even real-time needs



Realization in Real-world Systems - Content-based Recommendation

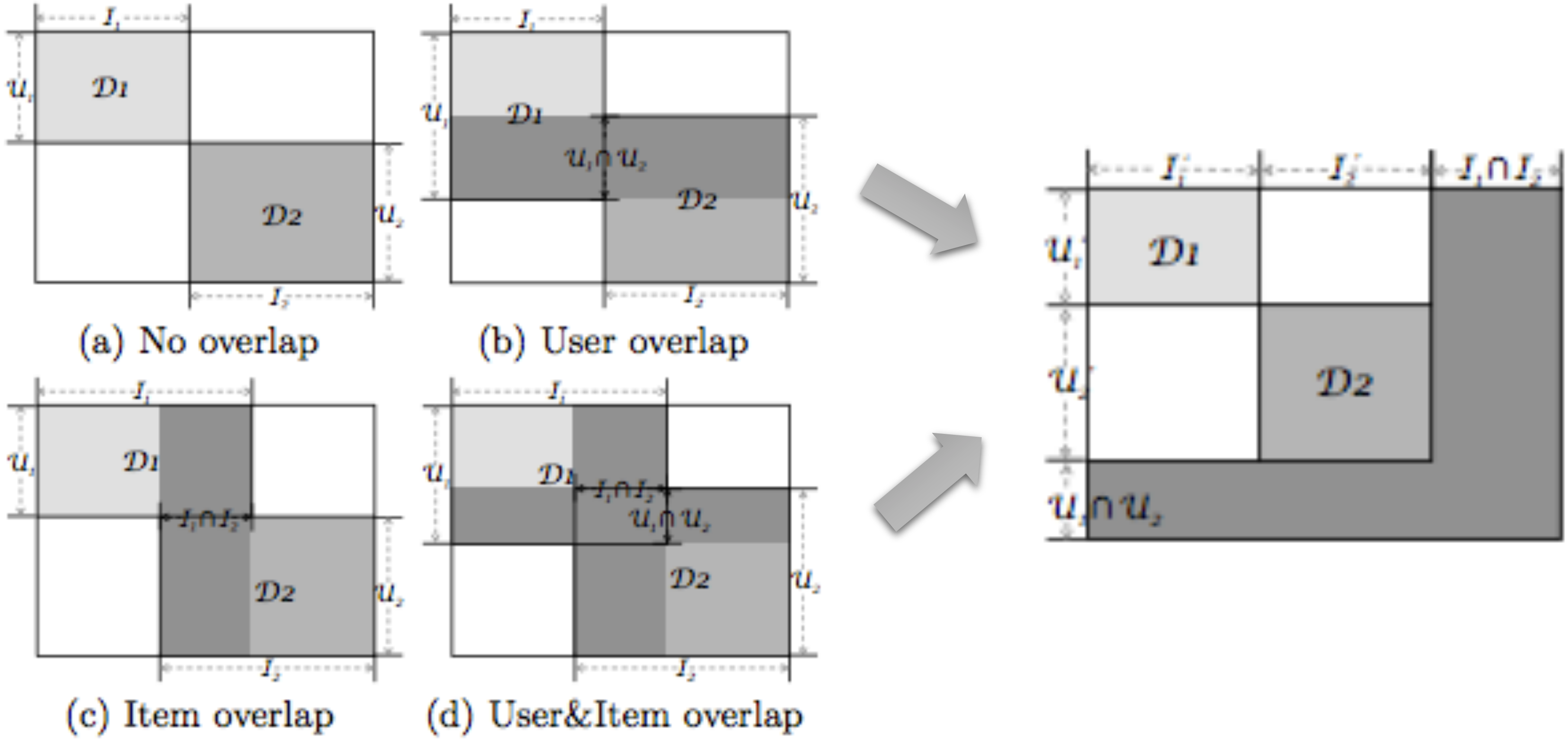
- Search engines and browsers provides rich content information
  - User queries
  - Textual content
  - User profiles
  - Item/product profiles
- Content-based rec further improves performance
  - By enriching the results of CF algorithms
  - To provide more informed recommendations

Unified Framework (RecSys14)



Unified Framework for Universal Recommendation (RecSys14)

- We propose a unified framework to incorporate inter-site information for universal recommendation [RecSys14]:



- This can be generalized to multiple site relations

[1] RecSys'14,Browser-Oriented Universal Cross-Site Recommendation and Explanation based on User Browsing Log

Incorporating Inter-site User Behavior Data by Search Engines / Browsers

- Search engine / browser logs as inter-site data
  - Provides row and column borders for BDF structures

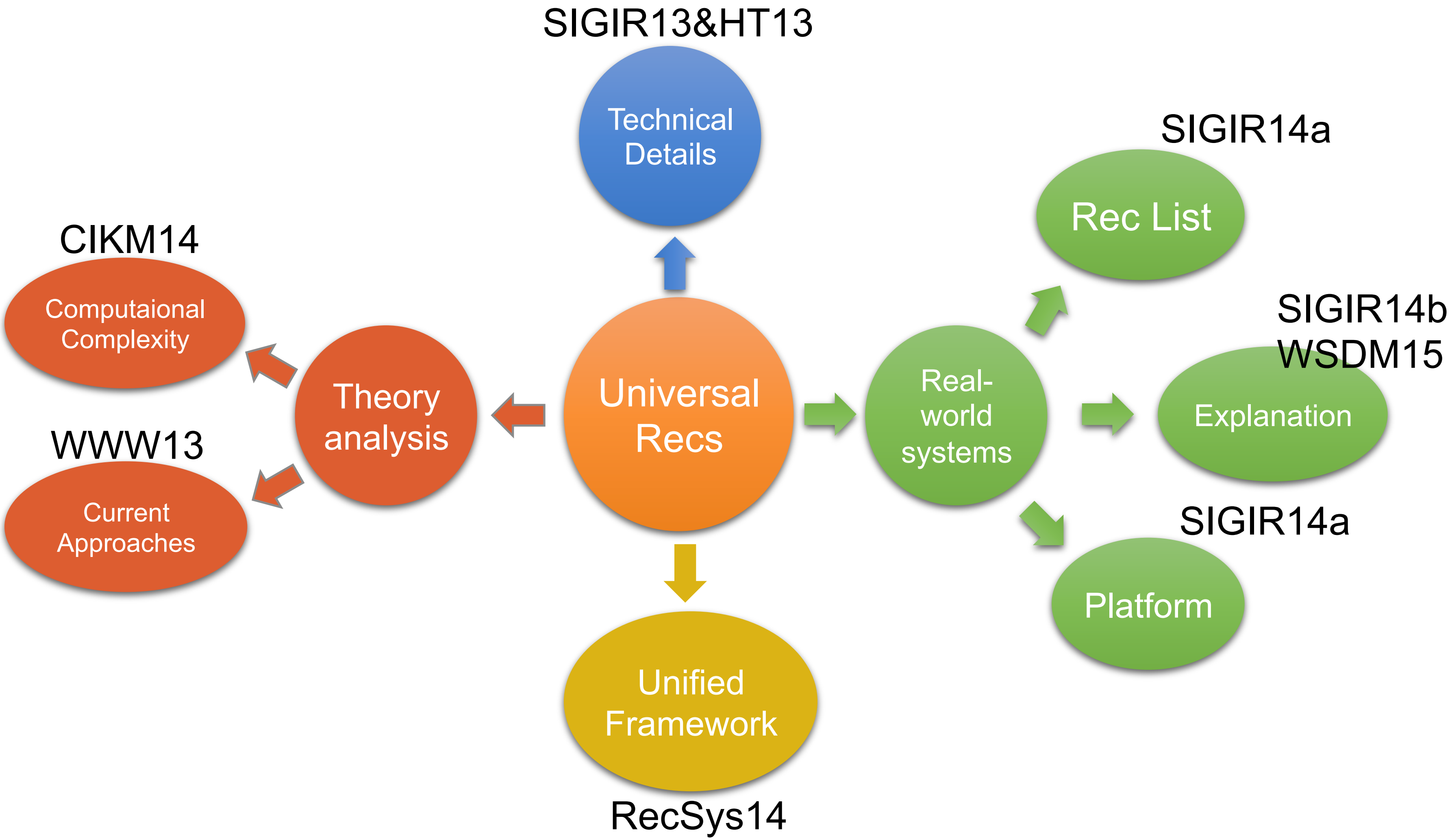
$$
\begin{array}{c|cccc|}
 & \mathcal{J}_1 & \mathcal{J}_2 & \cdots & \mathcal{J}_k \\
\hline
\mathcal{I}_1 & D_1 & & & \\
\mathcal{I}_2 & & D_2 & & \\
\vdots & & & \ddots & \\
\mathcal{I}_k & & & & D_k \\
\hline
\end{array}
$$

$$
\begin{array}{c|cccc:c|}
 & \mathcal{J}_1 & \mathcal{J}_2 & \cdots & \mathcal{J}_k & \mathcal{J}_B \\
\hline
\mathcal{I}_1 & D_1 & & & & C_1 \\
\mathcal{I}_2 & & D_2 & & & C_2 \\
\vdots & & & \ddots & & \vdots \\
\mathcal{I}_k & & & & D_k & C_k \\
\hdashline
\mathcal{I}_B & R_1 & R_2 & \cdots & R_k & B \\
\hline
\end{array}
$$

  - e.g. The product co-occurrence information in sessions
    - Users search for the Ocean Heart necklace after searching for Titanic
    - Provides row borders to the BDF structured matrix
  - e.g. Item dis-ambiguity results of the user queries
    - Provides column borders to the BDF structured matrix
  - They make it possible to conduct collaborative filtering[1,2]

[1]WWW 2013, Localized Matrix Factorization for Recommendation based on Matrix Block Diagonal Forms.
[2]SIGIR 2013, Improve Collaborative Filtering through Bordered Block Diagonal Form Matrices.

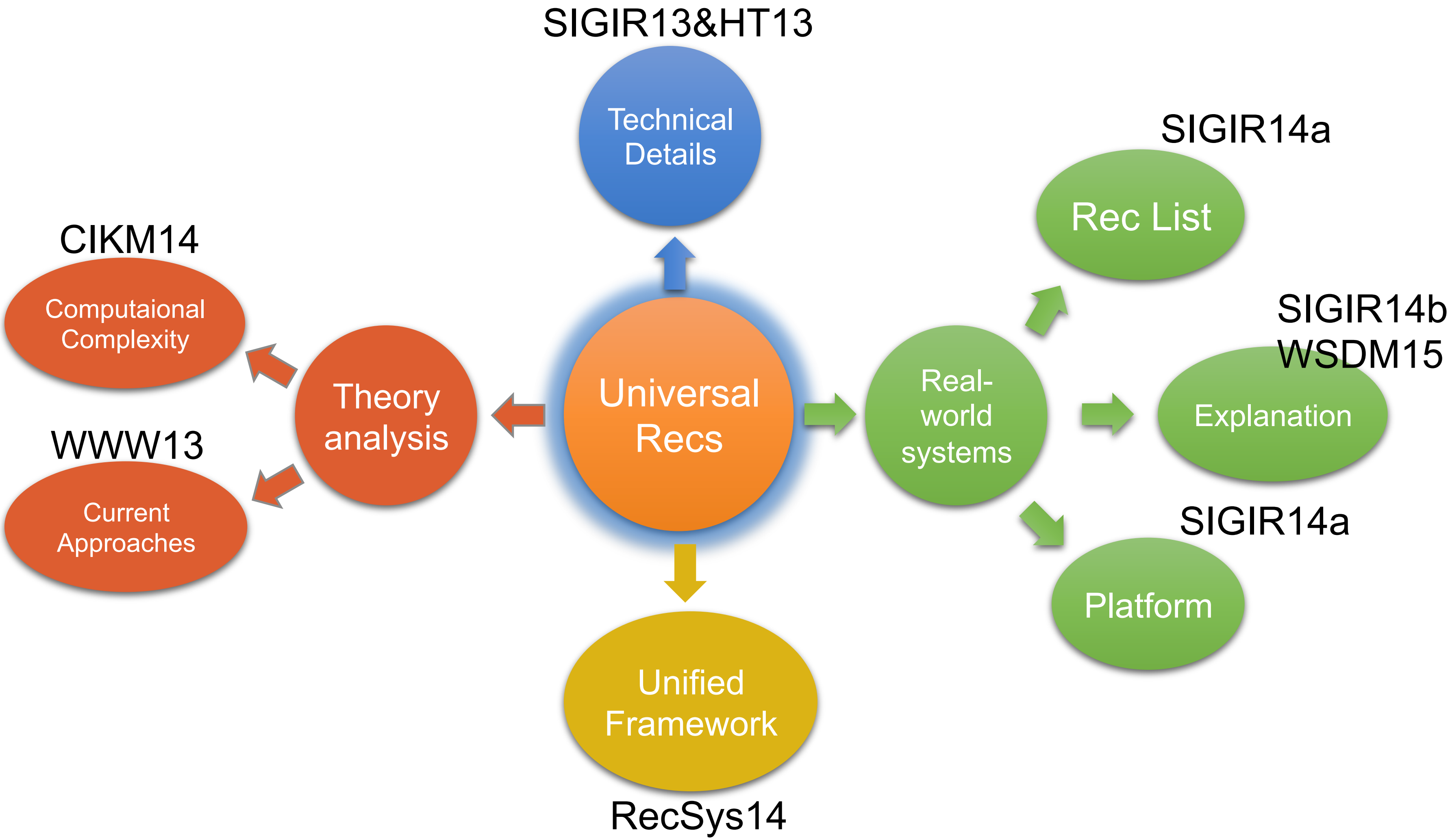
Universal Recommendation



Other Research Bases

- Data acquirements and Research Platform
  - Our lab cooperates with a major commercial search engine company in China (SoGou.com)
  - We have large-scale search engine and browser logs
  - The company has a famous web-browser product and large-scale real-world users
    - helps to conduct real scenario experiments
  - Pre-research results provide firm research foundations for this research topic

Universal Recommendation




THANKS