Intelligent Search Engine and Recommender Systems based on Knowledge Graph

阳德青

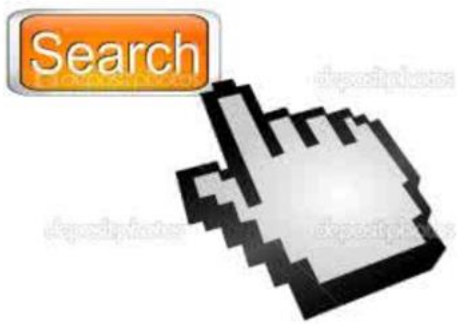
复旦大学知识工场实验室

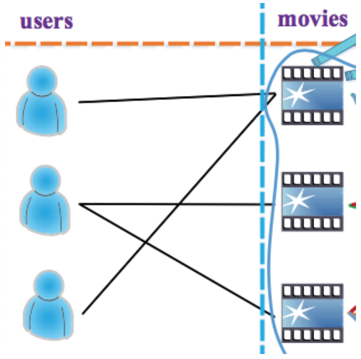
yangdeqing@fudan.edu.cn

2017-07-13

Background

- **Knowledge Graph** exhibits its excellent performance through the intelligent applications built on it
- As typical AI systems, **Search engine** and **recommender system** are very popular and promising in the era of large data
- Many previous literatures and systems have proved KG's merits on such AI's applications





KG-based Search Engine

The History of Search Engine

知識工場

- 1st: category-based
  - Yahoo, hao123
- 2nd: IR-based
  - Keyword-based, vector space, Boolean model
- 3rd: link-based
  - PageRank (Google)

result in

- The keyword of high click frequency are ranked higher
- The pages containing the keywords of more weights are ranked higher
- The pages having more important in-links are ranked higher

However, how to handle it
if users want to search something new or the ones of long tail?

The History of Search Engine

- When Knowledge Graph emerges

keyword/string → thing/entity, relevant entities

The History of Search Engine

知識工場

- Search 4.0: user-based search engine
  - Really understand the search intent of users, especially for those rare entities
  - Directly return pure answers (entities) instead of relevant web pages
  - Provide plentiful/relevant results with high confidence/interpretability

intelligent search engine

It is actually a recommendation.

Knowledge graph can help the engine achieve these goals

Entity Suggestion with Conceptual Explanation [1]

- More than search

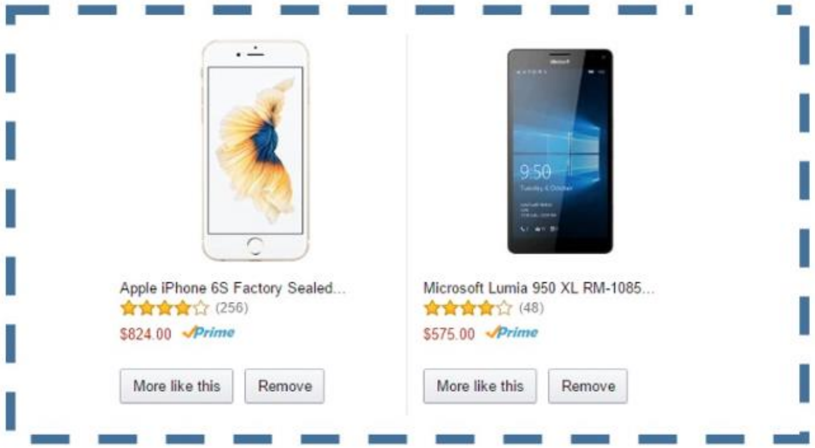
iPhone 6 Plus

or

Your Recently Viewed Items

Query Examples

You May Also Like ____, Because ____



Suggested Entities

Can any search engine return such results?

- Problem statement
  - Given a query *q* described by a set of entities, the engine should return some **relevant** entities along with some **fine-grained** concepts which can well explain the results

Entity Suggestion with Conceptual Explanation

知識工場

- What are the relevant entities and fine-grained concepts?

What entities should be suggested?

Baidu 百度

Chi

Entity Suggestion with Conceptual Explanation

- Previous Works
  - 1. Co-occurrence (text or query log), e.g. Google Set, Bayesian Set
  - 2. Web lists, e.g. SEAL, SEISA
  - 3. Overlap of properties
- Weakness
  - 1. Co-occurrence ≠ conceptual coherence
  - 2. Membership of lists → too general/specific concepts or a randomly combined list
  - 3. Overlap of properties ≠ conceptual coherence

Entity Suggestion with Conceptual Explanation

知識工場

Use **Probase** to find the relationships between concepts and entities

Find the best concepts which are **typical** and **fine-grained** to explain the suggested entities

- Solutions
  1. Probabilistic Relevance Model

$$\underset{e \in E-q}{\operatorname{argmax}}\, rel(q,e) = \sum_i \boxed{P(e|c_i)}\boxed{P(c_i|q)\delta(c_i)}$$

  2. Relative Entropy Model

$$\underset{e \in E-q}{\operatorname{argmin}}\, KL\big(P(C|q), P(C|q,e)\big) = \sum_{i=1}^{n} \delta(c_i) P(c_i|q) \log\left(\frac{P(c_i|q)}{P(c_i|q,e)}\right)$$

Entity Suggestion with Conceptual Explanation

知識工場

- $P(c_i|q)$ Computation

1. Naïve Bayes Model

$$P(c_i|q) = \frac{P(q|c_i)P(c_i)}{P(q)} \propto \prod_{e_j \in q} P(e_j|c_i)P(c_i)$$

$$\propto P(c_i) \prod_{e_j \in q, n(e_j, c_i) > 0} \lambda P(e_j|c_i) \prod_{e_j \in q, n(e_j, c_i) = 0} (1 - \lambda)P(e_j)$$

2. Noisy-or Model

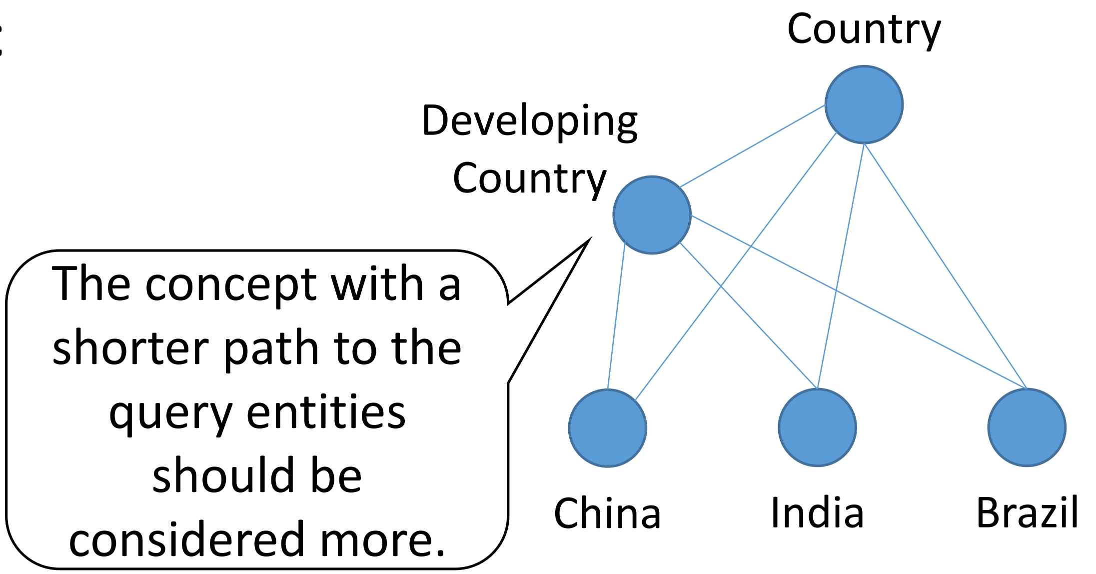
$$P(c_i|q) = 1 - \prod_{e_j \in q} (1 - P(c_i|e_j))$$

Entity Suggestion with Conceptual Explanation

- $\delta(c_i)$ computation
  - It quantifies the granularity of a concept, good concepts should not be too general or too specific

Concept	Number of Entities
Country	2648
Developing country	149
Growing market	18

Entity-based Approach



Hierarchy-based Approach

Entity Suggestion with Conceptual Explanation

知識工場

- $\delta(c_i)$ computation
  1. Entity-based Approach
     - Penalize popular concepts
     - $\delta(c_i) = \frac{1}{P(c_i)}$

  2. Hierarchy-based Approach（ Average first passage time ）
     - $\underset{C_q^k}{\operatorname{argmax}} \sum_{c \in C_q^k} \sum_{q_i \in q} h(q_i|c)$
     - $\begin{cases} h(q_i|c) = 0, & if\ q_i = c \\ h(q_i|c) = 1 + \sum_{c' \in c(c')} P(c'|q_i)h(c'|c) & if\ q_i \neq c \end{cases}$

Long Concept Query using Conceptual Taxonomies [2]

- Problem statement
  - Given a long concept, the engine should return the correct entities belonging the concept
  - An example

American blind male musician

Long Concept Query using Conceptual Taxonomies

知識工場

- A naïve solution
  - Divide the long concept
    - American blind male musician ⟹ {American musician, blind musician, male musician}
  - Challenge: Data Sparsity/Error by Hearst Patterns
    - For example:

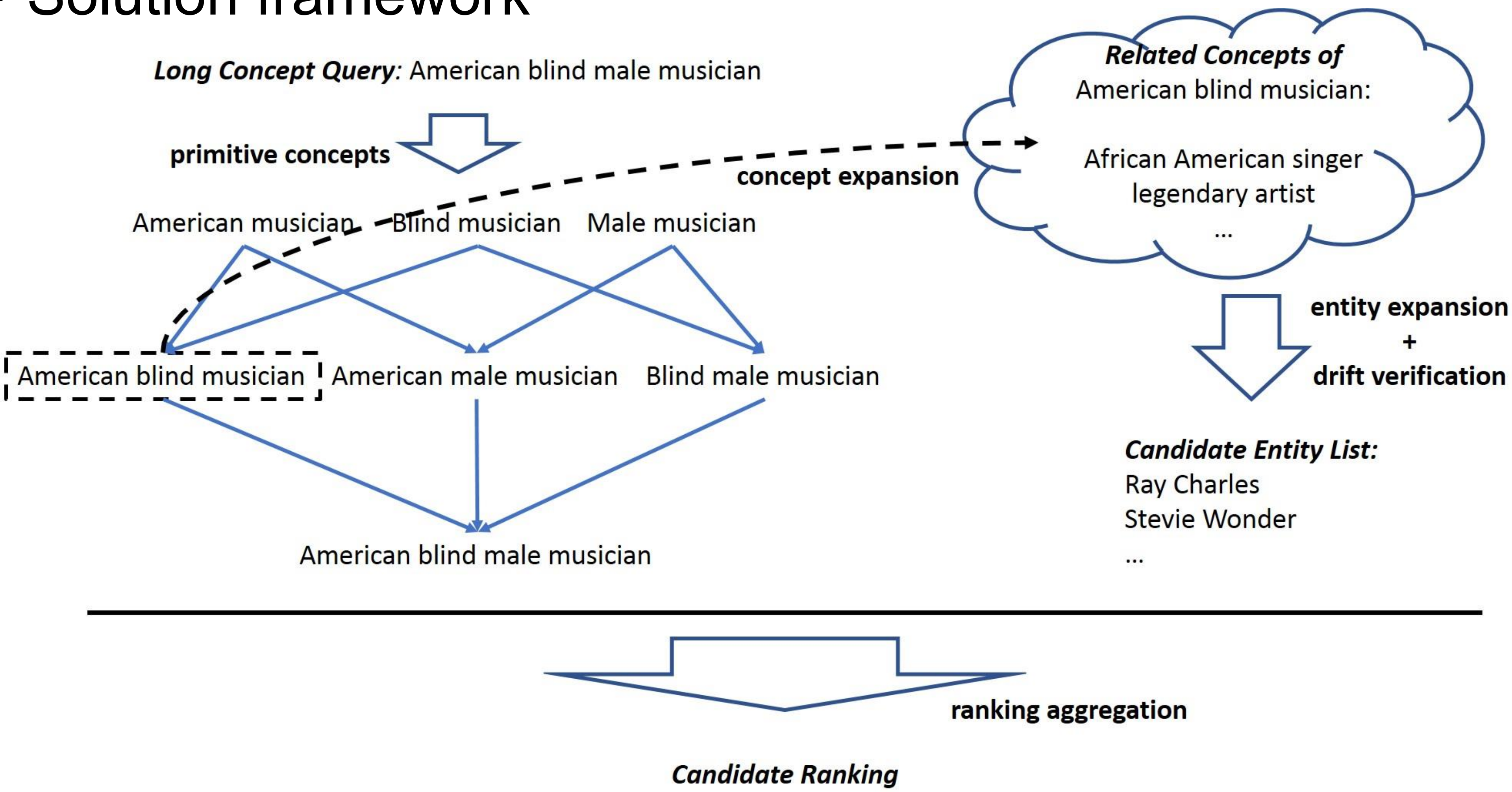
Concept	Entity
Blind musician	Ray Charles, Stevie Wonder, Kobzari, Jose Feliciano
Male musician	Vocalist, Jeffrey star, famous rock stars-guitarists, drummer

Finding right intersection is difficult

Long Concept Query using Conceptual Taxonomies

知識工場

- Solution framework



KG-based Recommender Systems

Traditional Recommendation Algorithms

知識工場

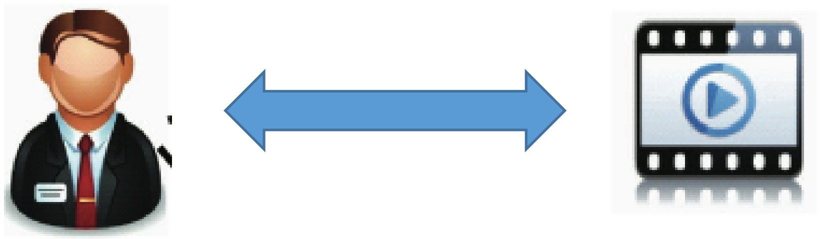
- Collaborative Filtering (CF) based
  - Use historical user-item interactions to describe the similarity between users/items, such as rating/purchase/review records…
- Content-based
  - Specify user/item’s content by tags, text description/reviews, purchase records, social ties…
- Hybrid

However, how to handle the new users/items with sparse explicit data?

Cold Start is an inevitable problem!

KG-based Recommendation

Finding precise **relationships** between users and items is the ultimate goal in recommender systems



- When Knowledge Graph emerges
  - Understand and characterize users and items more comprehensively
  - Uncover the semantic/latent relationships between users and items other than the user-item interactions
  - Uncover the semantic/latent relationships between the features of users/items

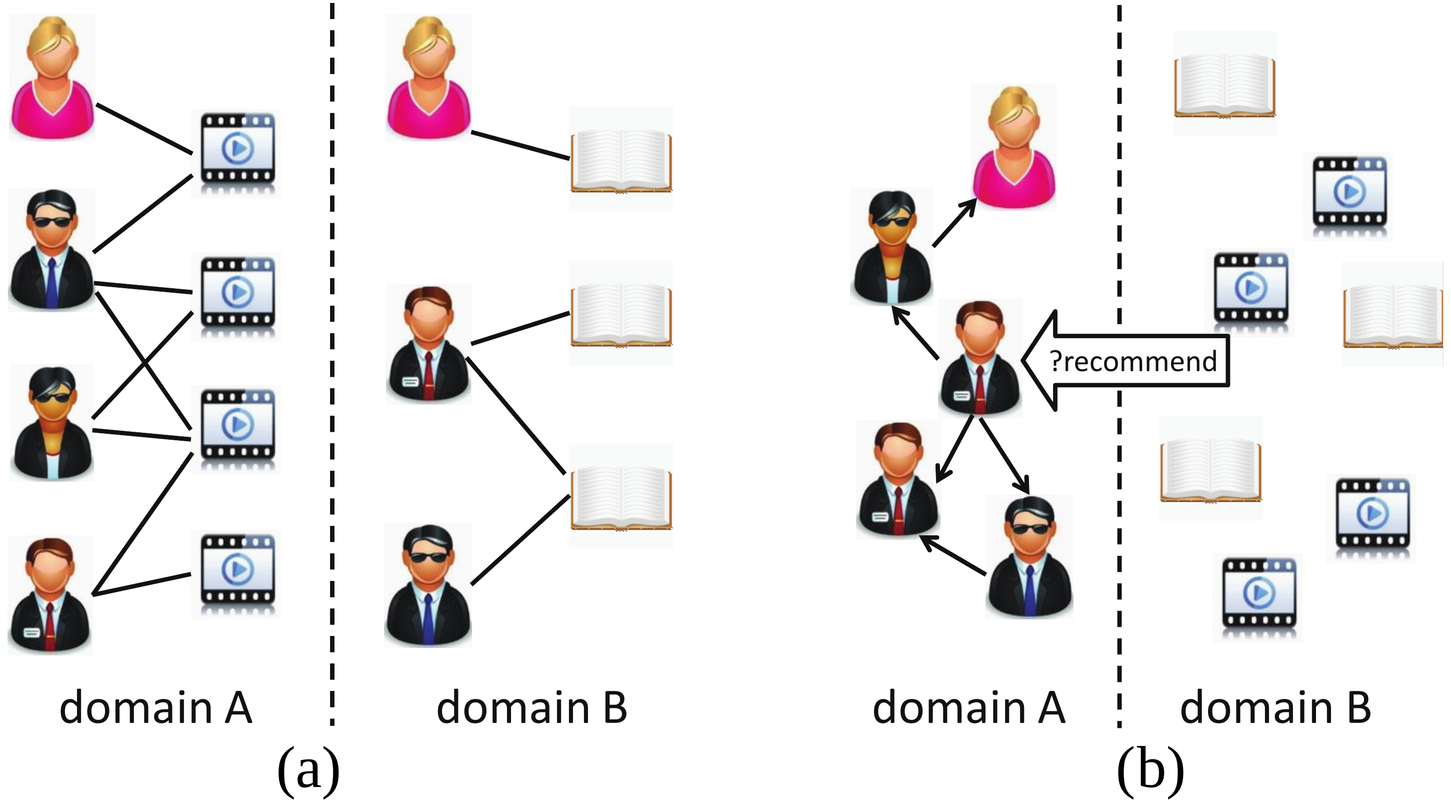
What Knowledge Can be Used?

- Example1
  - In KG, two movie/entity ‘警察故事’ and ‘十二生肖’ both have an attribute ‘主演：成龙’
  - Such knowledge can be used to find the relationship/similarity between the two movies rather than those discovered based historical user-movie ratings
- Example2
  - In a taxonomy KG, ‘华为Mate’ and ‘iPhone’ both ‘isA high-end business cellphone’, ‘荣耀’ and ‘小米’ both ‘isA low-end cellphone’
  - Such knowledge can used to infer the more precise characteristics of items
- Example3
  - For tag-based recommenders, the semantic relationships between tags is very important, e.g., ‘摄影’ and ‘旅游’, which can be learned by KG

Semantic-based Recommendation across Heterogeneous Domains [3,4,5]

知識工場

- Scenario of cross-domain recommendation



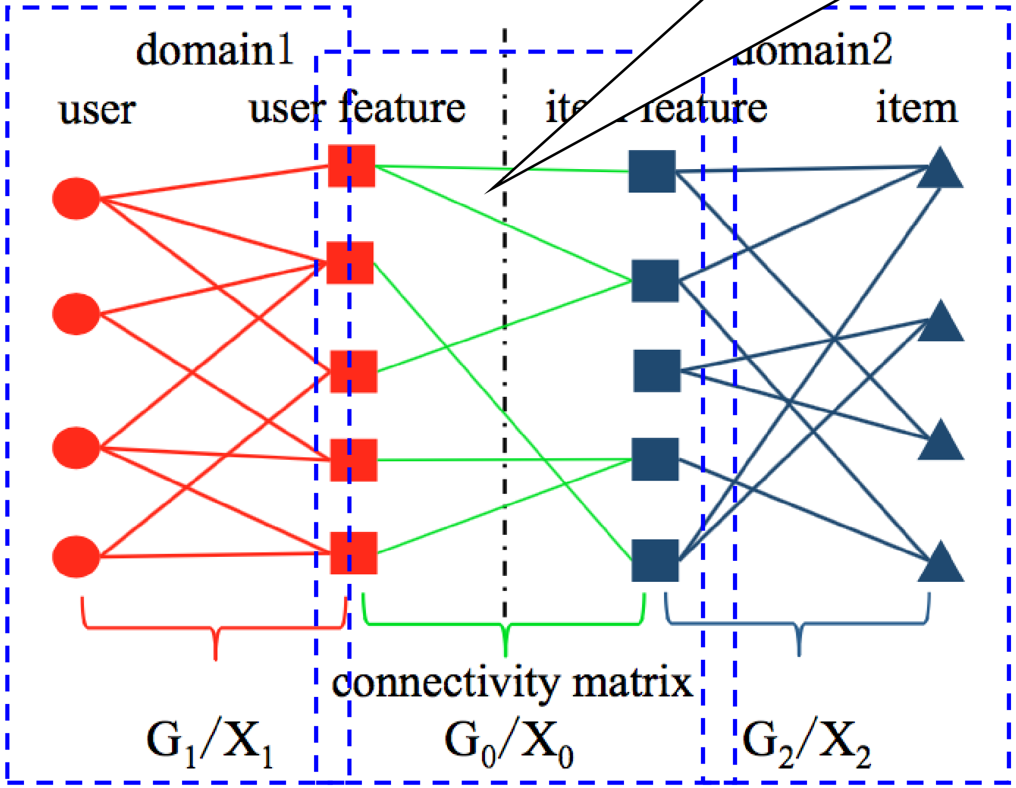
How to find the relationships between the users and items which are distributed in different domains?

Semantic-based Recommendation across Heterogeneous Domains

知識工場

- Solution
  - A multi-partite graph based algorithm

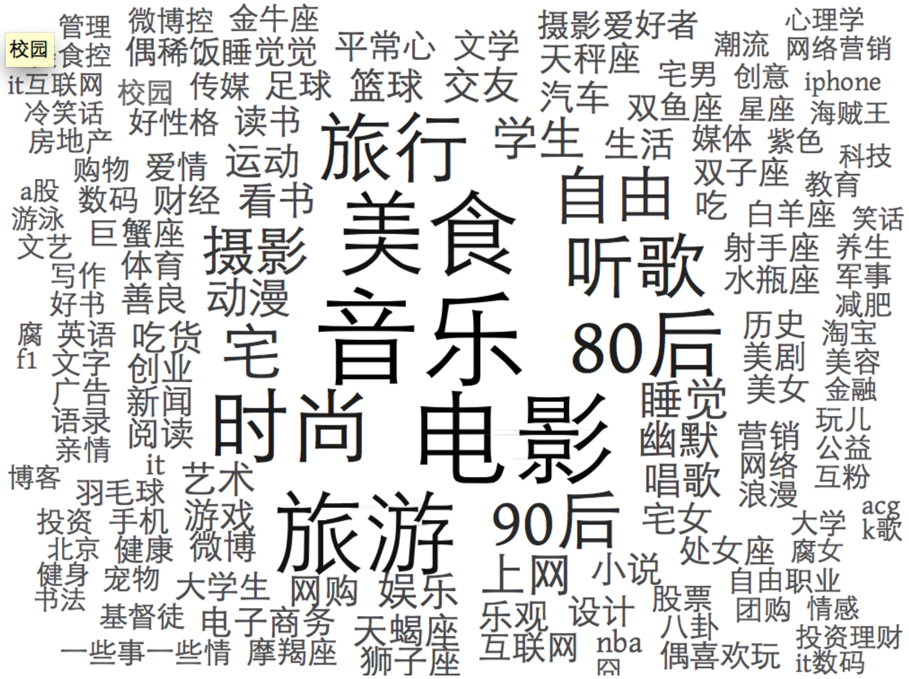
Capturing the correlations between the features distributed in different domains is critical.



Semantic-based Recommendation across Heterogeneous Domains

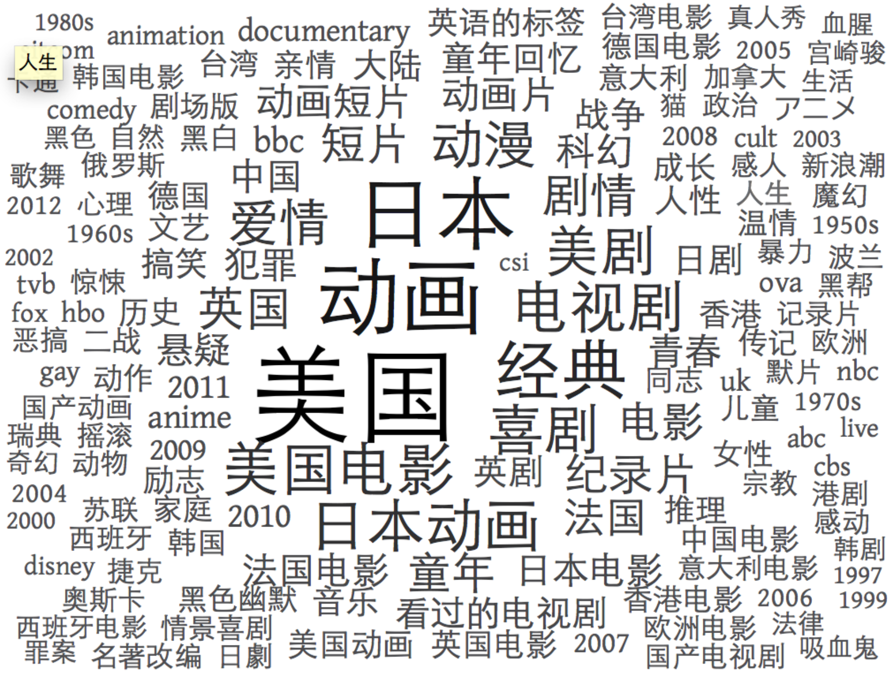
知識工場

- Solution
  - Discover the semantic relationships rather than lexical similarity between tags (features) by using the concepts/entities in KG



Weibo users' tags

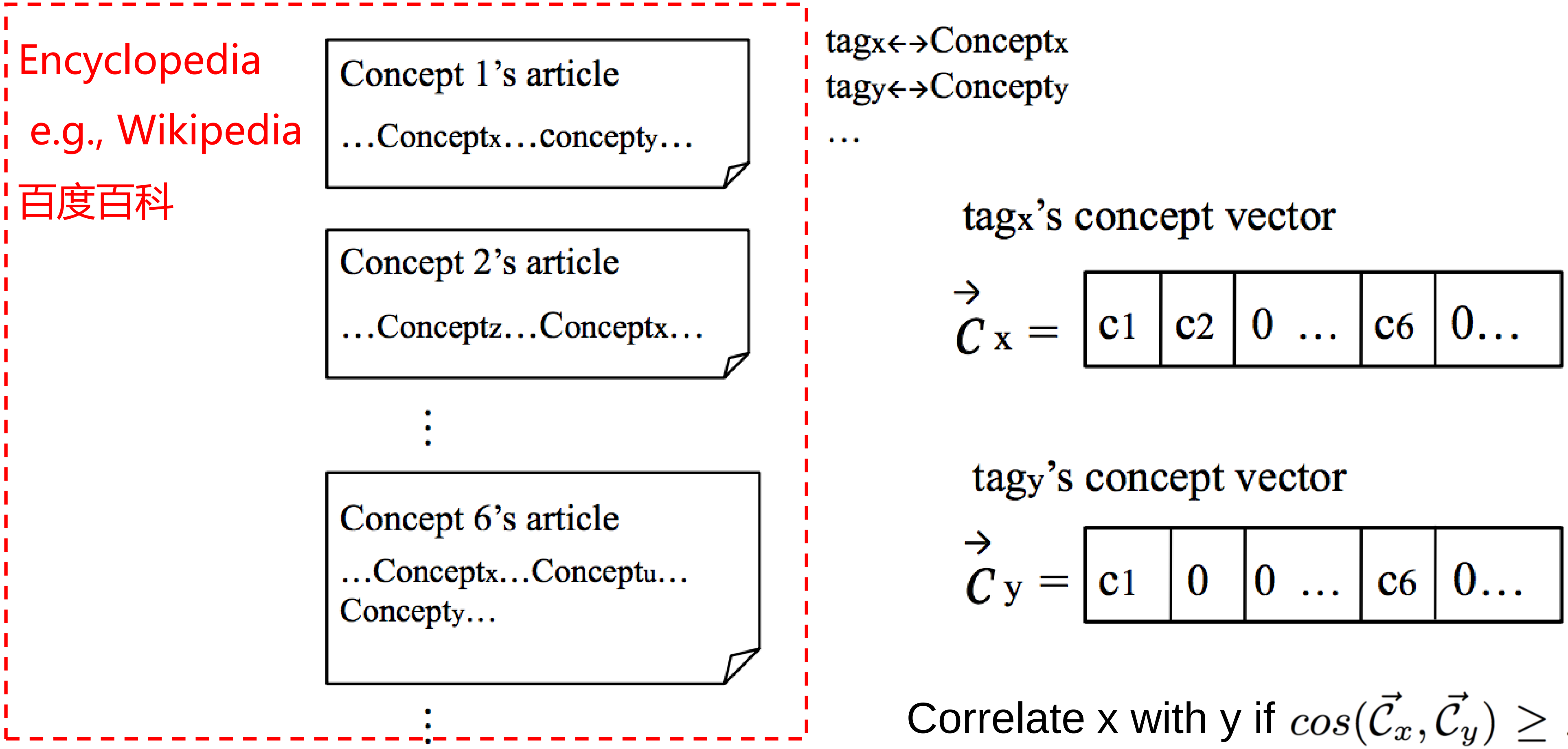
VS



Douban movies' tags

Semantic-based Recommendation across Heterogeneous Domains

- ESA based semantic matching

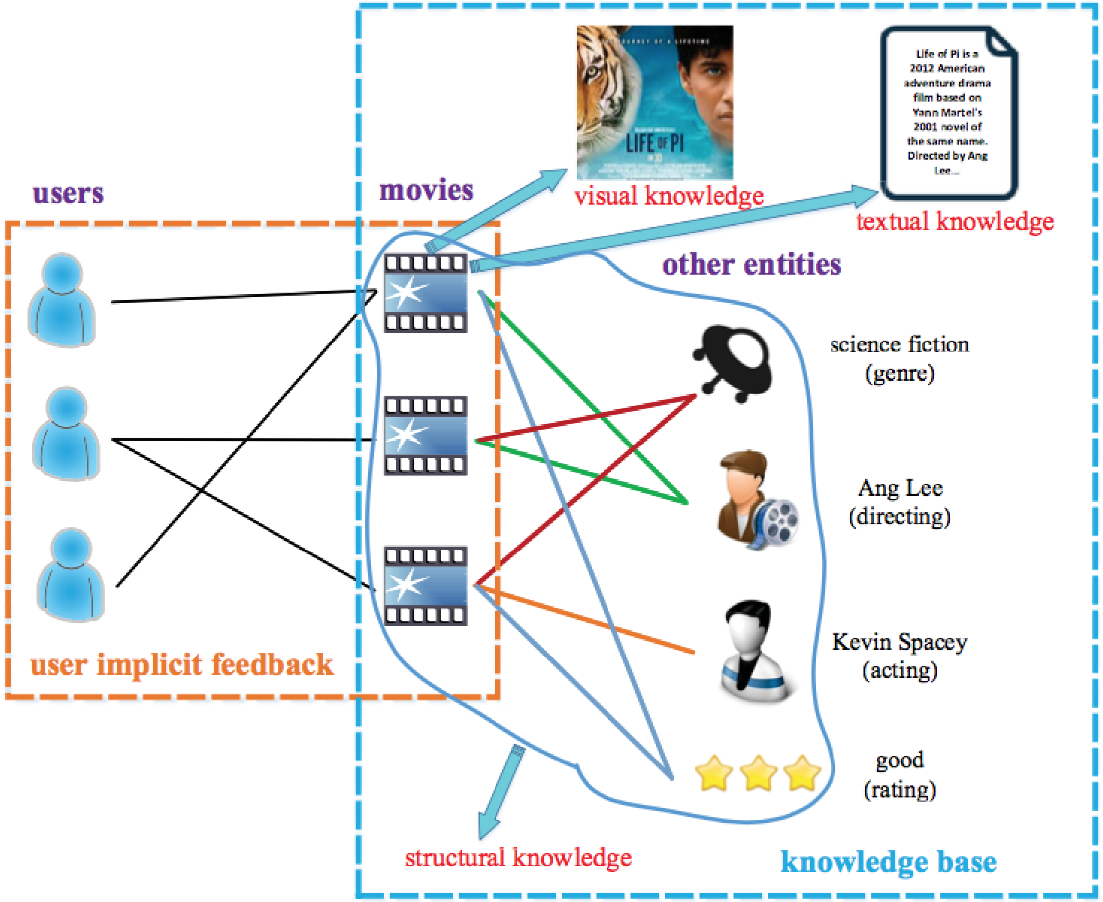


Correlate x with y if $cos(\vec{C}_x, \vec{C}_y) \geq \lambda$

Collaborative Knowledge Base Embedding for Recommender Systems [6]

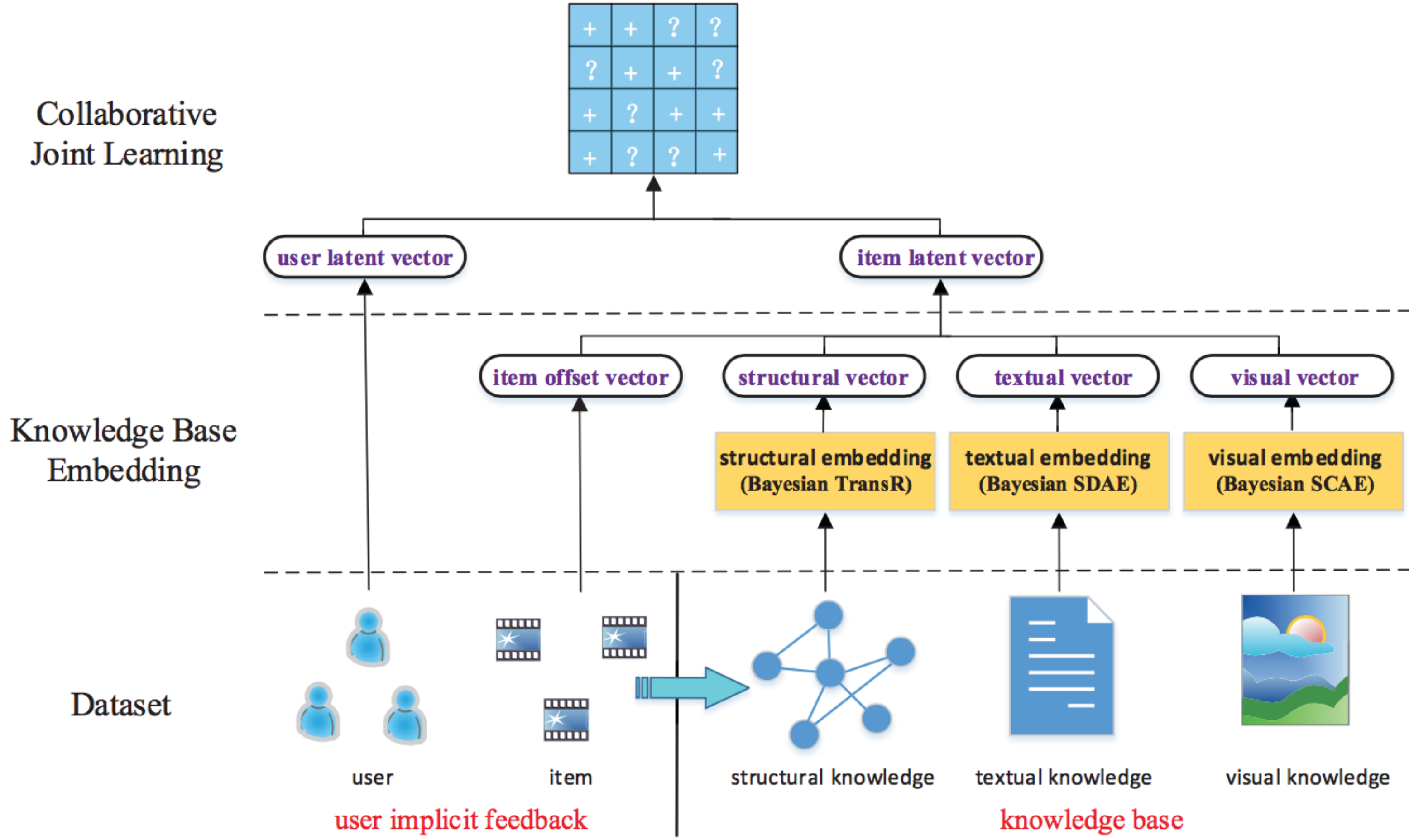
知識工場

- A perfect work to accomplish recommendation by fusing KG and DL algorithms



Collaborative Knowledge Base Embedding for Recommender Systems

- Three kinds of knowledge are utilized



References

[1] Yi Zhang, Yanghua Xiao, Seung-won Hwang, Haixun Wang, X. Sean Wang, Wei Wang: Entity Suggestion with Conceptual Explanation, in IJCAI 2017.

[2] Yi Zhang, Yanghua Xiao et al.: Long Concept Query using Conceptual Taxonomies, under reviewing.

[3] **Deqing Yang**, Jingrui He, Huazheng Qin, Yanghua Xiao, Wei Wang: A Graph-based Recommendation across Heterogeneous Domains, in CIKM 2015.

[4] **Deqing Yang**, Yanghua Xiao, Yangqiu Song, Wei Wang: Semantic-based Recommendation Across Heterogeneous Domains, in ICDM 2015.

[5] **Deqing Yang**, Yanghua Xiao, Yangqiu Song, Junjun Zhang, Kezun Zhang, Wei Wang: Tag propagation based recommendation across diverse social media, in WWW 2014.

[6] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, Wei-Ying Ma: Collaborative Knowledge Base Embedding for Recommender Systems, in KDD 2016

知識
工場

Thank YOU！
Q & A

Our LAB: Knowledge Works at Fudan University

http://kw.fudan.edu.cn