Search Intents: Understanding and Representation

Hongyuan Zha

Software Engineering Institute
East China Normal University
and
College of Computing
Georgia Institute of Technology

Outline

- Complex search intents
- Entity-centric search
- Interactive and adaptive search intent solicitation
  — user preference solicitation in CF

Search Intents and Search Space

- Search intents:
  - information need $\Rightarrow$ search intents $\Rightarrow$ queries
    — task $\Rightarrow$ sub-tasks $\Rightarrow$ queries
  - Scope: simple, complex, task-oriented, entity-centric
  - Modality: keywords, natural language, images, patient records, code snipppets, protein sequences, ...
- Search Space:
  - Content: textual, image, sensor readings, spatial-temporal, abstract and physical objects, ...
  - Structure: hyperlinks, social networks, ...
    — deep understanding of content and structure $\Rightarrow$ entity and relationship
  - Interaction data

Search Intents: Existing Technology

- Intent classification: navigational, informational, commercial, $\cdots$,
  — intent detection and inference
  — leveraged in ranking algorithm design in the learning to rank paradigm
- Query suggestion/expansion, Query auto-completion
- Relevance feedback

Complex Intents and Search Task Identification

- Complex intents:
  - Sustained interaction and engagement with information
  - The use of multiple search sessions across several days or longer
  - Build a deeper, internalized understanding of a problem or topic
- Task identification from query logs

  Intent representation: a query hierarchy

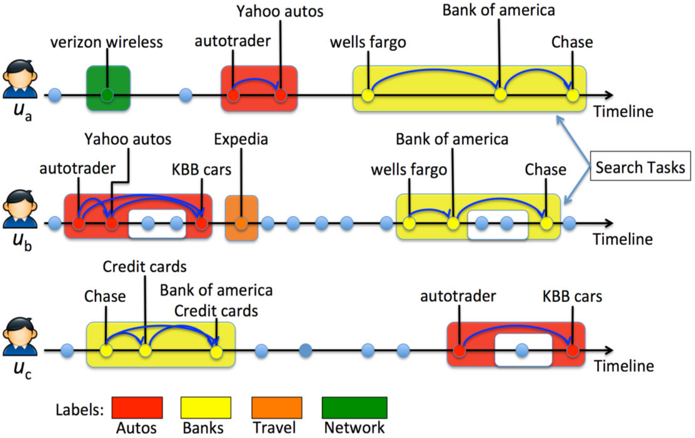
  — a tree with non-leaf nodes representing search tasks/subtasks and leaf nodes representing queries

  — a search task contains all queries on its descendant nodes
- Usage of query hierarchy:
  - Given a partial query sequence, find the corresponding task/subtask based on the hierarchy
  - Task-oriented query/sub-task/task completion pathways suggestions

Query Log Data and User Interaction Data

- Queries/subsequence of queries: measure semantic closeness
  — potentially augmented by topic models
- Time stamps of query submission: measure temporal closeness
- Content of documents: surrogates for query/sub-task closeness
  — can leverage click models
- Hyperlinks among returned documents: whether transitions exist between two sub-tasks
- User behavior/interaction:
  — click data $\Rightarrow$ multi-session click models

Query Streams to Search Tasks



L. Li, H. Deng, A. Dong, Y. Chang and H. Zha. Identifying and Labeling Search Tasks via Query-based Hawkes Processes. SIGKDD, 2014.

Temporally-Weighted Query Co-Occurrence

Two consecutive or temporally-close queries, issued many times by the same user or many others users

- Query co-occurrence modulated by temporal proximity
  — Query LDA model augmented by temporal point process for query submission time stamps
  — Potential personalization
  — No sub-task hierarchy yet

Scientific Challenges

- Fine-grained multi-resolution task-oriented search intent classification, detection and inference
- Modeling and tracking of tasks, processes and states (of completion)
- Extended models for task representation: sub-task dwell time, sub-task transition probabilities
- Task-oriented user search behaviors (interplay between tasks and user click behaviors), relevance models and evaluation metrics

Entity-centric Search and Knowledge Graph

- Entity-centric search (entities & relations)
  — deep understanding of content and structure
  — Web of things $\Rightarrow$ physical objects and links
- Comparison of search intent under complex tasks and entity-centric search
  - Representation: a query hierarchy/tree $\leftrightarrow$ complex graph
  - Resource: Observed query log $\leftrightarrow$ Knowledge graph built from additional resources
  - Node: query $\leftrightarrow$ entity
- Difference between tranditional search and entity based search
  - Query: keyword/natural language query $\leftrightarrow$ entity-centric query
  - Result: document $\leftrightarrow$ entity and relation

Query Annotation Leveraging Knowledge Graphs

Mapping between natural language queries and entity-centric queries

- Challenge:
  - Need to understand complex relation embedded in query to understand search intent
  - Need to identify entity type and relationship
- Leverage contextual information to avoid disambiguation in query annotation
- Use document-entity link to help entity disambiguation

Search Knowledge Graphs

natural language queries $\Rightarrow$ entities and relations

- Embedding of knowledge graphs
  - entities $\Rightarrow$ ℓ-dimensional vectors in Euclidean spaces
  - relations $\Rightarrow$ mappings (can be nonlinear) between vectors
  - deep learning and joint matrix factorization
    D. Zhou, S. Zhu, K. Yu, X. Song, B. Tseng, H. Zha, C. L. Giles. Learning Multiple Graphs for Document Recommendations. WWW, 2008.
- Geometry search

  — distance search with non-euclidean metrics

  — variations of R-trees

Scientific Challenges

- Effective mapping between keyword/natural language queries and entity-centric queries
  — knowledge graph based query annotation
  — leveraging data sources for entity/relation disambiguation
  — more complex and ambiguous queries: Web of things
- Deep-learning inspired knowledge graph embedding/link prediction algorithms: scalability, fast algorithms, fast updating methods
  — search intent inference as joint matrix/tensor completion problem
  — probabilistic methods
- Knowledge graph exploration: navigation patterns, integrating CF to aid knowledge graph exploration (search destination/path recommendation)
- Fast algorithm for distance search in knowledge graph embedding space

Interactive Search Intent Solicitation

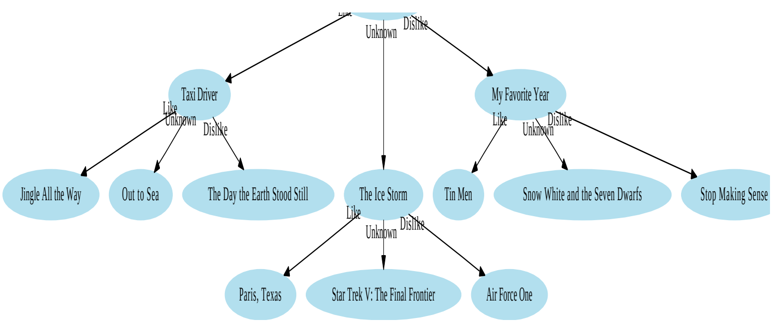
Search intent inference and understanding through an interactive interview process

- To understand a user's search intent, we can conduct an interview process
  — asking several interview questions to infer her search intent
  — analyzing the responses to the questions and build an intent representation for the user
- Key: Construct a mapping from the set of responses to interview questions to user profiles
  — In what form should be the mapping?

Interview Process as a Decision Tree

- **Key Observation** Questions asked should adapt to users' responses to previous questions
  — *Open-loop* optimization vs. *close-loop* optimization
  — Dynamic programming
- The interview process can be organized by a decision tree
  — At each node, an interview question is presented to the user.
  — Direct the user to one of the child nodes according to her answer.

An Example Decision Tree Based on Movielens



K. Zhou, S.H. Yang and H. Zha. Functional matrix factorizations for cold-start recommendation. SIGIR, 2011.
M. Sun, F. Li, J. Lee, K. Zhou, G. Lebanon and H. Zha. Learning Multiple-Question Decision Trees for Cold-Start Recommendation. WSDM, 2013.

Scientific Challenges

- Partially observed Markov decision process (POMDP) for search intent solicitation
  — close connection with RL
- Predictive models for tracking user intent shift
  — shifts triggered by feedback from search results
- Integrating an interview process to help with search intent solicitation based on knowledge graph exploration: we can estimate the user's search preference and recommend good starting entities or subgraphs to investigate