Ontology-Based Collaborative Recommendation

Ahu Sieg, *Bamshad Mobasher*, Robin Burke

Center for Web Intelligence
School of Computing
DePaul University, Chicago, Illinois, USA

ITWP 2010
Big Island, Hawaii
June 20, 2010

Introduction

Problem we are addressing:

- Contextual information access
  - Context-sensitivity is the key to lifting the burden of information access from users
- User context is an essential ingredient in building more intelligent personalized applications
  - Web search
  - Navigation Support
  - Recommender systems
- Allan et al. 2004 - “Contextual Retrieval: Combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user’s information needs”

Introduction

Solution:

- Develop a user modeling framework in which the user's "information access context" is learned
- Integrate the critical elements that make up the user's information context:
  - Short-term user interests or information need
  - Semantic knowledge about the domain
  - Long-term user profiles that reveal trends in user preferences
- Translate context representation into a form that can be used to support different information access activities
- **Ontological user profiles**
  - Annotated instances of a reference domain ontology

Information Access Context

❖ Web Search Scenario

?

Query: “Madonna and Child”

?

PeopleNews

E-MAIL IM

Madonna Ready for Another Baby?

Wednesday Nov 24, 2004 1:55pm EST

By Todd Gold

Three months after finishing her Re-Invention Tour, Madonna is currently enjoying quiet time with her family in London, she's just published her fourth book for young readers, *The Adventures of Abdi* – and, at 46, she tells PEOPLE she wouldn't mind getting pregnant again.

She's not making any definite plans, but the pop icon says: "I'm going to have fun with my husband and see what happens."

CREDIT: JO HALE / GETTY

Need to “learn” user’s profile and to disambiguate the domain of discourse:

*** User is an art historian?**

*** User is a pop music fan?**

Information Access Context

Recommendation Scenario

- Steve’s purchases on Amazon:
  - mystery-detective fiction “Da Vinci Code” (for himself)
  - “Python Programming” (for work)
  - “Green Eggs and Ham” (for his daughter)
- How should we represent Steve’s *interest in books*?
- System needs to know the difference between children books and computer books
- What should be recommended if Steve is reading reviews for a book on Perl Scripting?

Outline

- Ontological User Profile as the Context Model
- Updating User Context by Spreading Activation
- Ontological Approach to Collaborative Recommendation
- Experimental Evaluation

Domain Ontology

- **Represents concepts and relationships in a particular domain of interest**
  - Hierarchical concept structure and instances within the knowledge base
  - Rather than being associated with single atomic entities like individual books, users' choices and preferences are associated with relevant concepts in the ontology

- **Reference Ontology**
  - An existing domain ontology on which all ontological profiles are based
  - Underlying ontology can be modularized to allow for adoption to a variety of application domains
  - E.g., Amazon’s Book Taxonomy for a collaborative book recommender system

Ontological User Profiles

- **Ontological user profile is an instance of the reference ontology**
  - Each concept is annotated with an interest score
- **Whenever the system acquires new evidence about user interests, such as page views or explicit ratings, the user profile is updated with new interest scores**
- **Framework is designed to maintain and update the ontological user profiles based on the user behavior and ongoing interaction**
- **Profile Normalization**
  - Relative importance of concepts in the profile reflect the changing interests and varied information contexts of the user

Augmenting Collaborative Recommendation

- **Standard User-Based Collaborative Filtering**
  - Operates by selecting the *k* most similar users to the target user based on ratings on individual items
  - Formulates a prediction by combining the preferences of these neighbors
- **Collaborative Filtering with Ontological Profiles**
  - User similarities are computed based on their interest scores across ontology concepts, instead of their ratings on individual items
    - This helps broaden the recommendations and alleviate typical problems with CF: "cold start," "diversity," "serendipity"
  - Additional filtering is performed by selecting only neighbors that have significant interest in the concept of the "target item"
    - This helps in identifying the relevant "information access context" and improves accuracy

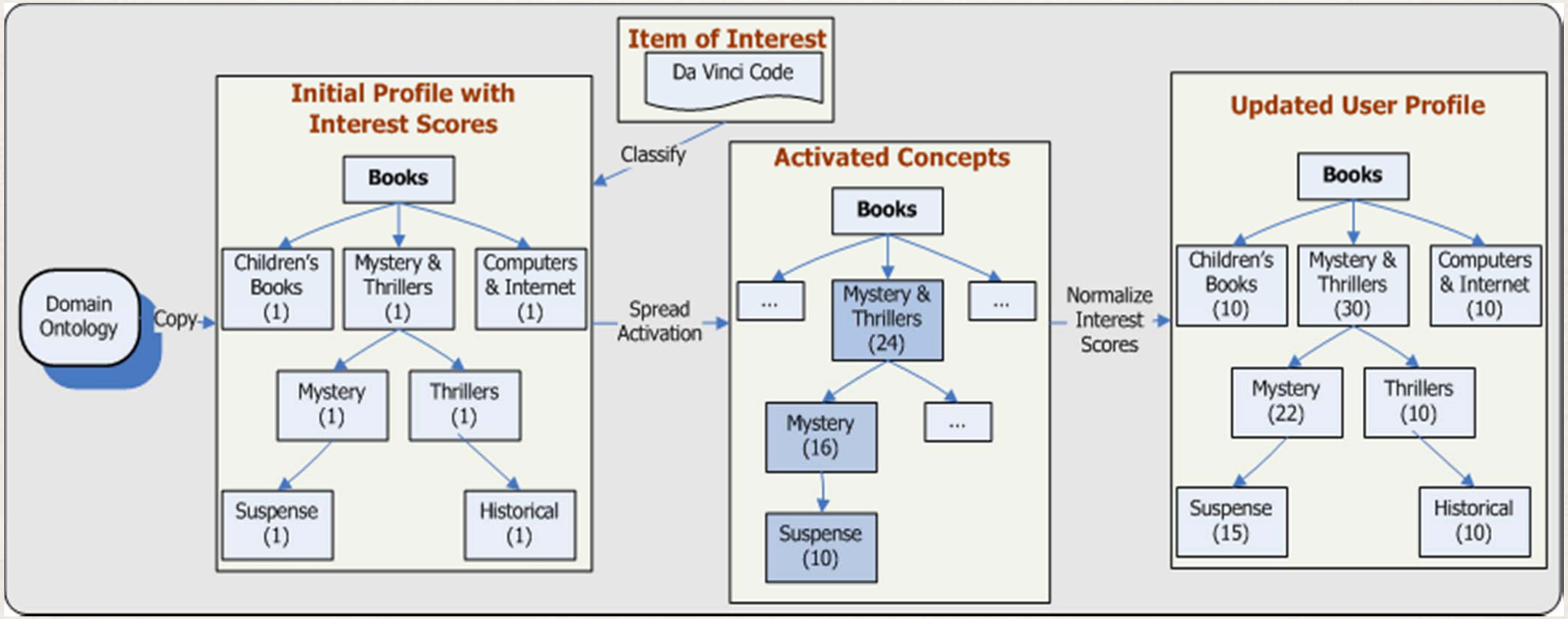
Updating User Context by Spreading Activation

- **Interest score**
  - Indicates the importance of a concept for the user
  - Gets incremented or decremented based on the user's behavior over many interactions
- **Spreading Activation**
  - Ontological User Profile is treated as the semantic network
  - Interest scores updated based on activation values
  - Initial set of concepts is assigned an initial activation value based on similarity to user's short-term interests
  - Activate other concepts based on a set of weighted relations
    - Relationship between adjacent concepts is determined based on the degree of overlap
  - Obtain a set of concepts and their respective activations

Spreading Activation Algorithm

- **Input**
  - Ontological user profile with interest scores
  - An item the user is interested in (e.g. book)
- **Spreading Activation**
  - Initial set of concepts are the concepts the item belongs to
  - Each concept is assigned an initial activation value
    - $IS(C_j)$
    - Priority queue in non-increasing order of activation values
  - Concept with highest activation
    - Remove from queue
    - Propagate activation to neighbors
      - activation value * activation weight
    - Add activated concepts to the priority queue
    - Reorder the queue
  - Add resulting activation values to existing interest scores for the concepts
  - Normalize interest scores
    - Increment or decrement

Profile Updating Illustrated



Ontology-Based Collaborative Recommendation

- **Semantic Neighborhood Generation**
  - Compare the ontological user profiles for each user to form semantic neighborhoods
  - Euclidean Distance

$$\text{distance}_{u,v} = \sqrt{\sum_{j \in C} \left(IS(C_{j,u}) - IS(C_{j,v})\right)^2}$$

- C - set of all concepts in the reference ontology
- $IS(C_{j,u})$ – interest score for concept C_j for target user u
- $IS(C_{j,v})$ – interest score for concept C_j for target user v

  - Normalize the distance
  - Calculate similarity based on the inverse of the normalized distance

Ontology-Based Collaborative Recommendation

Prediction Computation

- Compute the prediction for an item *i* for target user *u*
  - Select most similar *k* neighbors
  - Concept-based filtering on the neighbors

- Variation of Resnick's standard prediction formula

$$\boxed{p_{u,i} = \bar{r}_u + \frac{\sum_{v \in V} sim_{u,v} * (r_{v,i} - \bar{r}_v)}{\sum_{v \in V} sim_{u,v}}}$$

  - We use concept-based mean ratings for the target user and specific neighbors
  - *V* – set of *k* similar users

Experimental Evaluation

Questions

- Can the semantic evidence provided by the ontological user profiles be utilized to meet the user's recommendations needs?
  - Prediction Accuracy
  - Coverage
  - Cold-start performance
  - Personalization and diversity
- User profile convergence and accuracy

Experimental Setting

- **Reference Ontology**
  - Amazon's Book Taxonomy
    - ISBN – unique identifier for each book
    - Category, title, URL, and editorial reviews
    - 4,093 concepts and 75,646 distinct books
- **Ontological user profile: an annotated instance of Amazon's Hierarchy**
  - Concepts are annotated with interest scores (initially set to 1)
  - User's short-term interest (e.g. books of interest) is matched against concepts in the hierarchy
  - Spreading Activation is used to incrementally update the interest scores
- **Evaluation using the book ratings collected by Ziegler**
  - 4-week crawl from the BookCrossing community

Experimental Evaluation

❖ Experimental Data Set

- 72,582 book ratings belonging to users with 20 or more ratings
- Training data utilized for spreading activation
- Test data used for predicting ratings
- K-fold cross validation, $k = 5$

Experimental Evaluation - Metrics

- **Prediction Accuracy**
  - Mean Absolute Error (MAE)

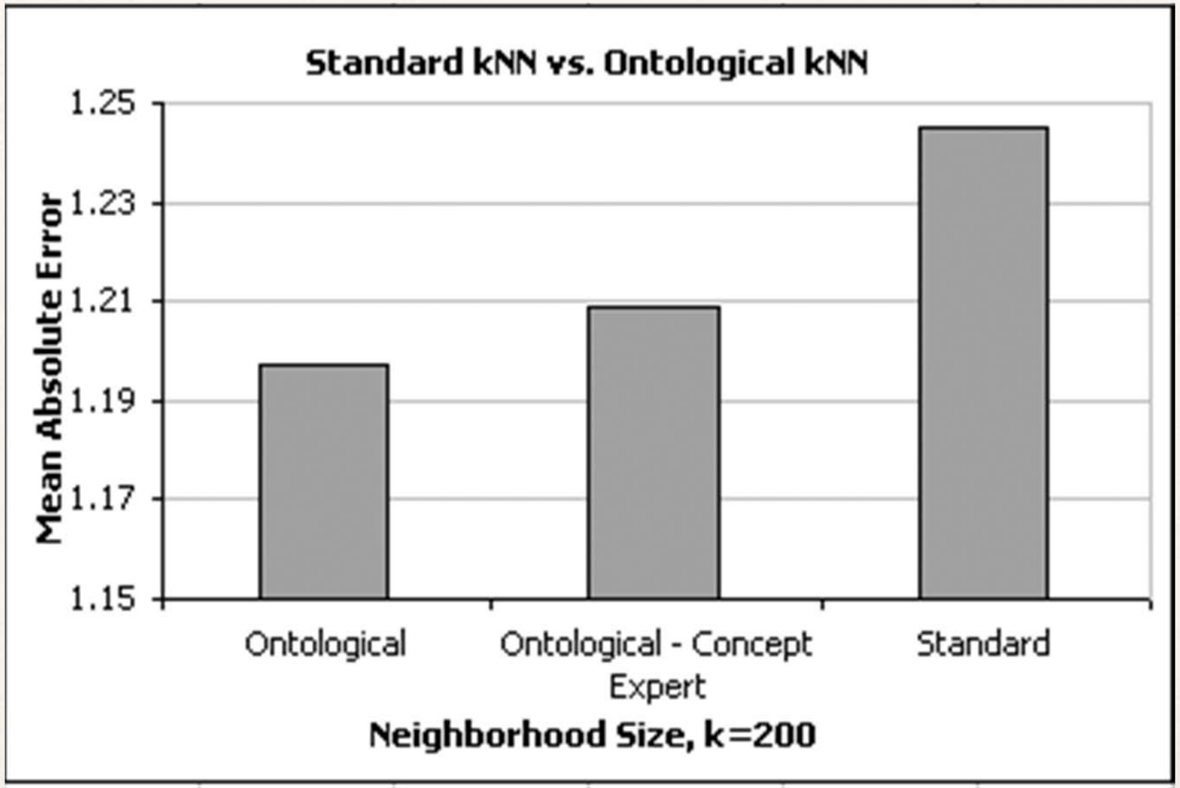
- **Top-N Recommendation Effectiveness**
  - Hit Ratio

- **Recommendation Diversity**
  - Personalization
    - measures the uniqueness of different users' recommendation lists based on inter-list distance
  - Surprisal
    - measures the unexpectedness of a recommended item relative to its overall popularity

Experimental Results

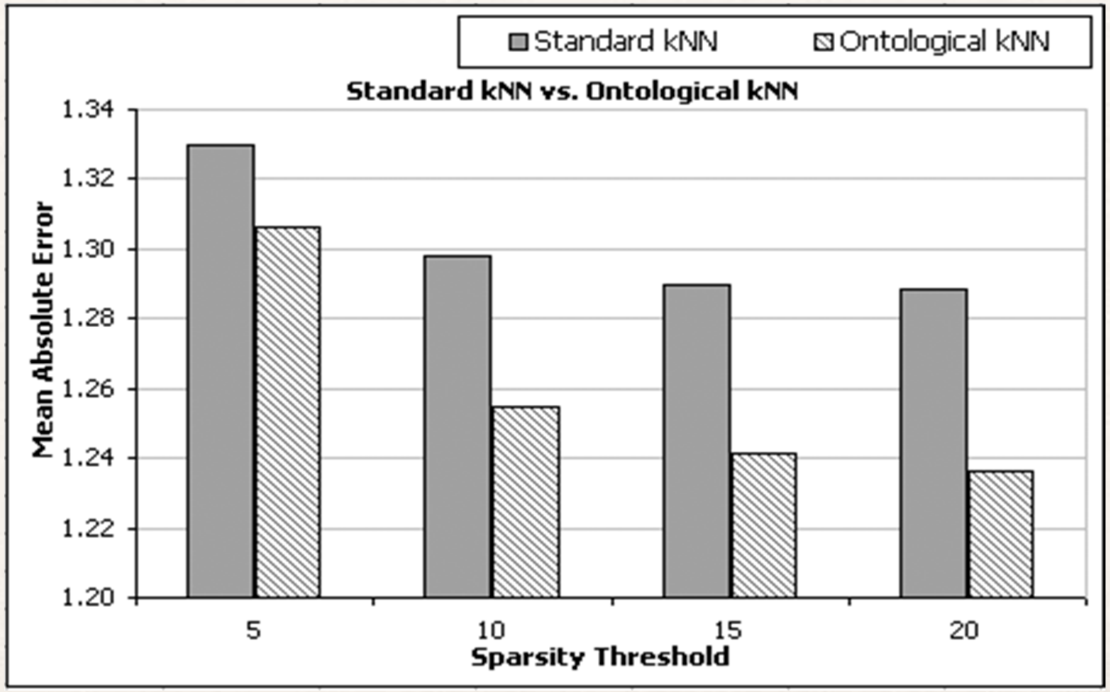
❖ **Mean Absolute Error, k=200**

- ANOVA significance test with 99% confidence interval, *p-Value < 0.01 (6.9E-11)*



Experimental Results

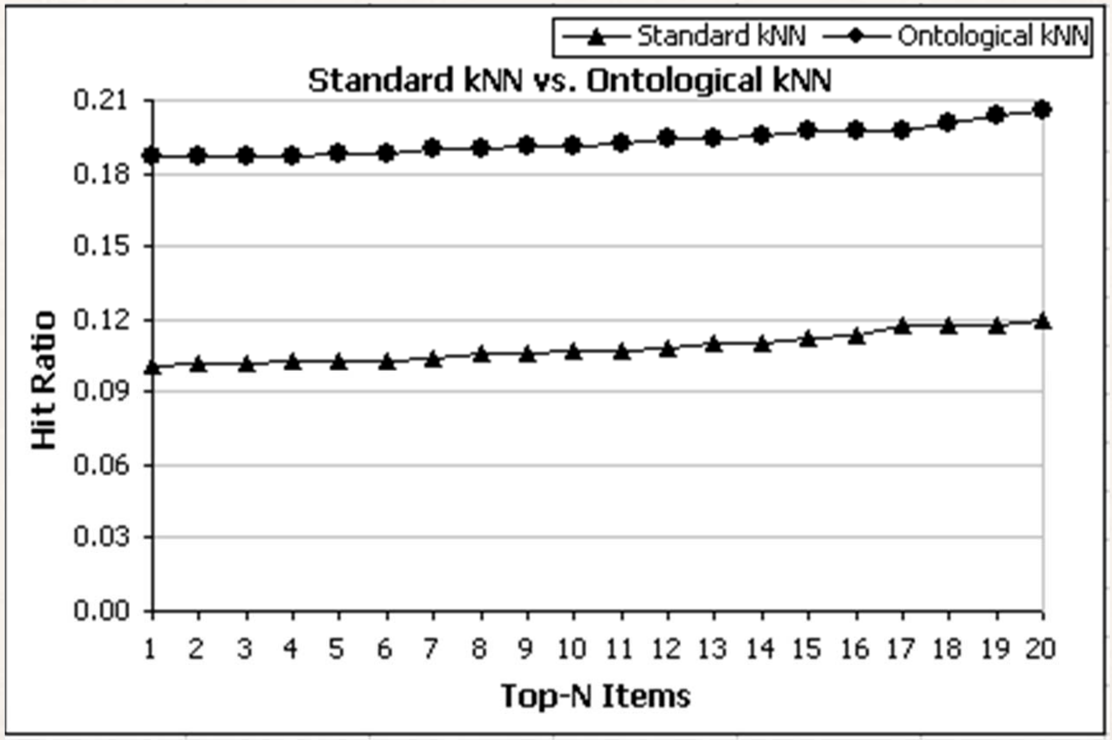
- **Cold-start Problem**
  - Item cannot be recommended until it has been rated by a substantial number of users
  - Computed MAE for items with 20 or fewer ratings to demonstrate improved cold-start performance



Dissertation pages: 110-111

Experimental Results

- **Top-N Recommendation**
  - Improved Hit Ratio
    - Computed by determining whether a hit exists within the top-N items in the recommendation list



Experimental Results

Recommendation Diversity

- Personalization
  - uniqueness of different users' recommendation lists based on inter-list distance
  - q_{ij} - number of common items in the top *N* recommendations for two given users *i* and *j*

$$d_{ij}(N) = 1 - \frac{q_{ij}(N)}{N}$$

- Surprisal
  - unexpectedness of a recommended item relative to its overall popularity
  - *i* – given item in a user's recommendation list
  - $frequency_i$ – number of overall positive ratings for *i* divided by the total number of users

$$I_i = \log_2\left(\frac{1}{frequency_i}\right)$$

Experimental Results

- **Recommendation Diversity**
  - Improved Personalization
  - Improved Surprisal

Algorithm	Personalization, d(20)	Surprisal/Novelty, I(20)
Standard kNN	0.922	6.544
Ontological kNN	0.975	7.286
ANOVA p-value	1.9417E-276	4.9221E-181

Conclusions and Outlook

- **Framework for contextualized recommendation using ontologies**
  - Semantic knowledge embedded in an ontology combined with long term user profiles can be effectively used to improve collaborative recommendation
  - Integration of ontological profiles help identify user's relevant information access context, thus improving accuracy
  - The semantic context also broadens the set of recommended items, thus improving the diversity and serendipity of recommendations
- **Future work**
  - Long-term stability and convergence patterns of profiles
  - Additional metrics to measure the "quality" of recommendations
  - Experiments with additional data sets

Questions

?