Web Personalization

Presenter: Hyoung-rae Kim

Advisor: Philip K. Chan Ph.D

Committee: Debasis Mitra Ph.D
Marius-Calin Silaghi Ph.D
Alan C. Leonard Ph.D

Department of Computer Sciences
Florida Institute of Technology
Web personalization

- Web information retrieval
- User modeling
- Machine learning
Web information retrieval

• Non information retrieval systems
  – Contents validation (e.g., checking broken links, broken images, etc.)
  – Monitoring remote machines, etc.
Basics of Web information retrieval

- Stop list
- Stemming
- Phrase
Basics of WIR

• Stop list
  – Removing words that have no inherent meaning (e.g., “a”, “the”, “are”, or “to”).
  – Using a stop list in the range of 8 through 500 do not reduce the accuracy [Frakes and Baeza-Yates, 1992]

• Stemming
  – Common morphological variants of a word have similar meanings (e.g., “computer”, “computing”, “computerized”, and “compute”).

• Phrase
  – A composed term has more specific meaning (e.g., “apple tree”, “apple computer”).
Clustering Web contents

- Cluster related Web pages together.
- usually happens in Portal services.
- Does not use any personal information.
- Cluster large amounts of information in a short time.
An example of clustering Web contents

- **Grouper** [Zamir and Etzioni, 1999] dynamically groups the search results.
- The groups are ordered by their estimated coherence.
- Use STC (suffix tree clustering) algorithm.
- Each cluster consists of phrases extracted from the snippets.
- A drawback: the clusters can fail to capture the semantic distinctions the users were expecting.
Predicting navigation

- Predict future requests.
- Use other people’s information.
- Can support prefetching.
- When the server guesses correctly, the latency of the next request is greatly reduced.
- A drawback: hard to predict previously unvisited pages.
An example of predicting navigation

- Pitkow and Pirolli [1999] uses longest repeating sequences (LRS),
- With those sequences, they build All-$K^{th}$-Order Markov model.
- This reduces model size significantly while retaining the ability to make accurate predictions.
Personalized contents

- Reorganize contents & design of a Web page.
- Mainly controlled by the server side.
- Use other people’s information.
- Can support prefetching, but do not care much.
- Can support unvisited pages partially.
An example of personalized contents

• Anderson [2002] personalizes a web content for different audience.
  – elide-content, swap-siblings (rearranging content), add-shortcut (add navigation links)
  – PROTEUS: personalizes web browsing for visitors using wireless PDAs at many web sites, adapting each site in turn;

• combines visitors’ content interests with a model of their navigation behavior.

• applicable to both individual pages and newly created (never visited before) pages.
Assisting personal information

- Help a user to organize their own information improving the usability.
- Reside in a personal computer.
- Can use other people’s information.
- Do not much related to predicting navigation.
An example of assisting system

- **PowerBookmarks** [Li et al., 1999] is a Web information organization, sharing, and management tool.
- measures how often the user visits them and their link structure etc.
- monitors and utilizes users’ access patterns and provide useful personalized services:
  - automated URL bookmarking, document refreshing, bookmark expiration (e.g., dead link), subscription services for new or updated documents (for certain query criteria)
Implicit detection

- Detect the interests of a Web page from a person.
- Can happen in a Client side or Server side.
- (e.g., Web log information, user’s access pattern, user’s behavior etc)
An example for implicit detection

- Previous work [Liberman, 1995; Mobasher et al., 2000; Joachims et al., 1997; Pazzani et al., 1996; Stefani and Strapparava, 1999] has investigated user navigation behavior and access log files.
- Goecks and Shavlik [1999] extended this work
  - # of hyperlinks the user clicked on; # of user scrolling activity; and the amount of user mouse activity.
- They use backpropagation algorithm for learning.
User modeling

Mainly user models try to describe [Webb et al., 2001]:
- The cognitive processes of user’s action;
- The difference between the user’s skill and expert skills;
- The user’s behavioral pattern or preferences;
- The user’s characteristics
Adaptive hypermedia

• The differences from adaptive web sites lies in the application domain, such as help systems.

• Adaptive navigation support can improve
  – the speed of navigation [Kaplan et al., 1993]
  – learning [Brusilovsky and Pesin, 1998]
  – understanding of content [Boyle and Encarnacion, 1994].
An example of hypermedia

- Weber and Specht [1997] proposed a distance learning system.
- demonstrated that user modeling techniques are effective for adaptive guidance and for individualized help.
- used a combination of an overlay model (provide default path and short cut path) and an episodic user model (stores knowledge about the learner in terms of a collection of episodes, such episodes can be viewed as cases).
User modeling in Web

- Personalized Web contents/interface using user model
- Controlled by the client/server side.
- Can support more personalized service.
- Support Web information retrieval techniques
Human behavior based user model

- Learn user model by observing user’s actions
- such as Web log file, path, click, downloads, frequency, etc.
- Can support prefetching.
- Do not use the contents information.
- Hard to predict unvisited Web pages.
An example of behavior based model

• TELLIM [Hoerding, 1999] evaluates every presentation element whether the customer was interested in it or not.
  – monitors the behavior of a customer
  – recognize the user’s needs and preferences
  – by their personal experience some decision lists are generated (e.g., if the downloading of an integrated image was interrupted, then it has negative interest to the customer).
  – Based on those decision lists, rules are generated
  – The attributes are quite simple: kind of product, e.g. “car”, brand, e.g. “Ford”, and kind of information, e.g. “engine”.
  – (e.g., a rule: “if the size of the item is less than 20GB (hard drive), then the customer may not be interested”).
Contents based user model

- learn user model by the contents of Web pages
- use the contents information.
- good for predicting unvisited Web pages.
- usually need more number of training data.
An example of contents based model

- **Syskill & Webert** [Pazzani et al., 1996] is an intelligent agent that learns user profiles.
  - ask users for rating Web pages
  - informative words are used as Boolean features
  - uses Naïve Bayes Classifier in building user profiles.
  - learns a naïve Bayesian classifier to determine the interestingness of pages.
  - test with Backprop, Perceptron, ID3, NearN, Frequent, other than Bayes.
Hybrid way based user model

- learn user model by observing user’s actions and the contents of Web pages visited by a user.
An example of hybrid user model

- Mobasher et al. [2000] combine site usage based clustering and site content based approach.
- The user preference is automatically learned from Web usage data, and integrated with domain knowledge and the site content.
- These profiles could be used to perform real-time personalization.
- The integration of usage and content mining increases the usefulness and accuracy of the resulting recommendations.
Implicit/Explicit user modeling

- Explicit way
  - Time and effort in specifying her interests
  - User’s interest may change over time
- Implicit way
- Hybrid way
An example for hybrid user modeling

- Ardissono et al. [1999] use hybrid way.
- The initial model is built by asking users directly.
- Their user model is stereotype (e.g., age, gender, education level, field of job, traveling, going to cinema, etc.).
- The model is refined periodically after monitoring the user’s behavior (e.g., which news s/he selects and which one s/he does not).
- The obtained user models are used for dynamic generation of the web pages based on a knowledge base (which news, at which detail level and which advertisement).
- A drawback: difficulty of setting rules for revising user profile.
Machine learning

- **Supervised**
  - Characterization: Given only one label, find boundary
  - Classification: Given multiple labels, find boundaries

- **Unsupervised**
  - Clustering: Given no labels, find boundaries
  - Outlier detection: Given no labels, find minorities

Depth exam.
Symbolic method for Classification

- allows easier human comprehension.
- Learning decision trees
- Learning sets of rules
- Semantic network
Semantic networks

- It uses the metaphor
  - that objects are nodes in a graph,
  - that these nodes are organized in a taxonomic structure, and
  - that relations between nodes are represented by arrows.

- SiteIF [Stefani and Strapparava, 1999] represents a user model as a semantic network,
  - whose nodes are concepts and arcs between nodes are the co-occurrence relation of two concepts.
  - It is an effective way to represent data as they incorporate the inheritance mechanism that prevents duplication of data.

- A drawback: the network can be complex
Numerical method for Classification

- human cannot easily understand what is learned in Numerical method - the performance is quite good though.
- Hidden Markov model
- Naïve Bayes classifier
- Neural networks
- Instance based learning
Naïve Bayes classifier

- WebWatcher [Joachims et al., 1997] is a tour guide software agent.
  - Highlighting interesting hyperlinks.
  - Prefetching recommended Web pages.
  - Learning from experience to improve its advice-giving skills.
  - Leveraging data from different users (runs as a centralized server)
- User interest (user model) is represented by high-dimensional feature vectors, each dimension representing a word.
- A disadvantage:
  - Difficulty of interpretation since every vector has only probability
  - No description about short-term and long-term interests.
Clustering

- Five sub categories [Han and Kamber, 2000].
An example

- Bookmark Organizer [Maarek and Ben-Shaul, 1996]
- This uses the classical HAC, they build a non-binary cluster hierarchy by applying “slicing” technique.
Concluding remarks

• Discussed three research areas related to Web personalization.
  – Web information retrieval
  – User modeling
  – Machine learning

• Identified different groups of techniques in each area.

• Provided an example for each group.
Thanks
Depth exam.
• **Collaborative Filtering**: Guiding people's choices of what to read, what to look at, what to watch, what to listen to (the filtering part); and doing that guidance based on information gathered from some other people (the collaborative part) [Paul Resnick]

• **A recommender system** is any system which provides a recommendation, prediction, opinion, or user configured list of items that assists the user in evaluating items.
An example

• An n-gram is a sequence of n web request $<x_1, x_2, x_3, \ldots, x_n>$

• The conditional probability that a surfer transitions to an $n^{th}$ page given their previous $k=n-1$ page visits:

\[
P(x_n|x_{n-1}, \ldots, x_{n-k}) = \Pr(X_n=x_n|X_{n-1}, \ldots, X_{n-k}),
\]

where $X$ represents training data, $x$ testing data.
Another example

• A newsagent called News Dude [Billsus and Pazzani, 1999], learns which stories in the news a user is interested in.

• The newsagent uses a multi-strategy machine learning approach to create separate models of a user’s short-term and long-term interests.

• They use the Nearest Neighbor algorithm for modeling short-term interests and a Naïve Bayesian classifier for long-term interests.
Machine learning

• Distance functions
• Clustering techniques
• White box way of learning
  – Semantic network
  – Hidden Markov model
  – Learning sets of rules
• Black box way of learning
  – Naïve Bayes classifier
  – Neural networks
  – Instance based learning
Distance functions

- Desirable characteristics [Piatetsky-Shapiro, 1991]:
  - P1: if $A$ and $B$ are statistically independent, then $F$ is 0;
  - P2: $F$ monotonically increases with $P(A,B)$ when $P(A)$ and $P(B)$ remain the same;
  - P3: if $P(A)$ (or $P(B)$) increases when the rest of the parameters ($P(A,B)$ and $P(B)$ (or $P(A)$)) remain unchanged, then $F$ monotonically decreases.
Other properties

• Statistical independence can be measured by the determinant operator, where \( \text{Det} (A,B) = A \cap B \times \cap - A \cap \times \cap B \) [Tan et al., 2002 (not original author)].

• Measuring their \textit{cross product ratio (CPR)} can assess the significance [Rosenfeld, 1994].

• Symmetry under variable permutation

• Row/Column scaling invariance

• Antisymmetry under Row/Column permutation

• Inversion Invariance

• Null Invariance
Hierarchy of machine learning

Machine learning

Supervised
- Characterization
  - Symbolic method
- Classification
  - Numerical method

Unsupervised
- Clustering
- Outlier detection
- Partition
- Hierarchical
- Density-based
- Grid-based
Hierarchy of machine learning

- **Machine learning**
  - **Supervised**
    - Characterization
    - Classification
  - **Unsupervised**
    - Clustering
    - Outlier detection

Characterization: (know only good things, one clusters)
Classification: (each instances has their label)
Clustering: (no labels)
Outlier detection: (no labels)
Hierarchy of machine learning

- Machine learning
  - Supervised
    - Characterization
      - Symbolic method
    - Classification
      - Numerical method
  - Unsupervised
    - Clustering
      - Partition
      - Hierarchical
      - Density-based
      - Grid-based
    - Outlier detection
Instance based learning

• It simply stores the presented training data.
• Examples are nearest neighbor and locally weighted regression.
• The disadvantage of this approach is that the cost of classifying new instances can be high.
Machine learning

- Supervised
  - Characterization
  - Symbolic method
- Classification
- Unsupervised
  - Clustering
  - Outlier detection
- Numerical method
Learning sets of rules

• It directly learns rule sets to assign categories to a training set.
• TELLIM [Hoerding, 1999]
Hidden Markov model (HMM)

- It is to model sequence of data.
- Deshpande and Karypis [2001] used pruning techniques for HMM model to reduce the model size and improve the prediction.
- This approach is mainly used in predicting navigation in a Web server.