Web Personalization

- “Web personalization is the process of customizing a Web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user’s navigational behaviour (usage data) in correlation with other information collected in the Web context, namely structure, content and user profile data.”
Why Web Personalization

- Continuous growth in the size and the use of WWW
- Huge financial interests in attracting and retaining customers
- Need for predicting user needs in order to improve the usability and user retention of a web site

Why ISP-Centric Web Personalization

**Traditional approach**
- Owners of web pages insert recommendations based on offline processing results. Traditional algorithms:
  - collaborative filtering, contextual parameters (location etc.)
  - demographics, previous purchases
  - local navigational patterns
- Issues
  - Global navigational behavior is not taken into account
  - Navigational behavior changes through time
  - User contextual data are rarely known

**ISP centric personalization**
- Full Knowledge of the visited Web graph by all the subscribers-navigational patterns.
- ISP has all users’ demographic and context related data.
- Thus: Recommendations can be added in every web page the user visits in real time
- Potential for an ISP Advertising framework (targeted), as effective as Google ad-words
System Architecture

Recommendations are inserted in every web page as HTTP traffic goes through the router.*

* Requirement for low latency (< ~0.5 sec per web page)

Online Personalization Process [1]
Online Personalization Process [2]

1. Request Web Page
2. Web Page Transmitted
3. Temporary store Web Page and parsing
4. Send Web Page and User feature vector (bag of words with frequencies) <<20KB/page>>
5. Select top-k matching recommendations
6. Send Recommendation list <<1KB>>
7. Embed Recommendation list

Web Page Feature Vector Extraction

- 1\textsuperscript{st} Approach: Known Vector Space
  - Server updates Router periodically with the vector space, constructed by the ads repository stored at server
  - Fast string matching on web page source

- 2\textsuperscript{nd} Approach: Unknown Vector Space
  - Use of Information Theory methods to extract to N terms on web page
Weight Calculation of Features

- On feature vector, every feature corresponds to a weight representing the importance of each term.

- Attributes that usually are used:
  - Linguistic Information (e.g. Noun or not)
  - Capitalization
  - Anchor Text
  - Meta Section features
  - Title
  - URL
  - Relative Location (The beginning of a document often contains an introduction or summary)

Recommendation Mechanism
Matching Model 1st Alternative

- Use Terms as the coordinates in a vector space
  - Compute the vector sum \((vs)\) of user and Web page vectors
  - Compute cosine similarity between recommendation vectors and \(vs\)
  - Select the K-nearest recommendations

Matching Model 2nd Alternative

- User’s vector as vector space
  - Projection of web page vector to the new space
  - Projection of recommendations vectors to the new space
  - Compute cosine similarity between web page and recommendation vectors
  - Select the K-nearest recommendations

- 3rd Alternative: Web Page vector as vector space
Recommendations Ranker

- A ranking algorithm will take into account
  - The similarity rank from matcher
  - Click Through Rate (CTR)
  - Corresponding Bid
- something like
  - \( F(x) = a \times \text{similarity} + b \times \text{CTR} + c \times \text{bid} \)
  - \( \text{Where } a > b > c \)
- Problem with new ads => 0 impressions

Recommendations Filter

- Filter based on
  - \( K \) : number of recommendation that an be shown on the current web page
  - A relevance threshold : irrelevant recommendations to the users
    - Increases users annoyance
    - Reduces system reliability
Challenges [1]

- **Storage and Processing Data:**
  - Provider Server
    - Fast Match, Ranking, Filtering (Use of Inverted Indexes)
  - In User Router
    - Fast string match or feature extraction
    - Limited Capacity
    - ISP Provides a Specific Router
      - HTTP proxy on router that supports ICAP (Internet Content Adaptation Protocol)

Challenges [2]

- **Recommendations insertion**
  - **Browser plugin**
    - Easy to develop (plug-in, toolbar, etc)
    - Not embedded in the HTML content
    - ISP should motivate users to install such software
  - In Router
    - Embedded in the HTML content (user friendly)
    - Alter the html content of every website raises *legal issues*
References

- George Forman, Evan Kirshenbaum, Extremely Fast Text Feature Extraction for Classification and Indexing, HP Laboratories, HPL-2008-91R1
- Wentau Yih, Joshua Goodman, Vitor R. Carvalho, Finding Advertising Keywords on Web Pages, WWW 2006 ACM
- Deepayan Chakrabarti, Deepak Agarwal, Vanja Josifovski, Contextual Advertising by Combining Relevance with Click Feedback, WWW 2008 / Refereed Track: Search - Ranking & Retrieval Enhancement
- Anderei Broder, Marcus Fontoura, Vanja Josifovski, Lance Riedel, A semantic approach to contextual advertising, ACM
- Sarah K. Tyler, Yi Zhang, Dou Shen, Exploring Collaborative Filtering for Sponsored Search, SIGIR’09, 2009
- Berthier Ribeiro Neto, Marco Cristo, Paulo B. Golgher, Edlene Silva de Moura, Impedance Coupling in Content Targeted Advertising, SIGIR 2005 ACM

Thank You