Facet Selection Algorithms for Web Product Search

Damir Vandic, Flavius Frasincar, Uzay Kaymak

vandic@ese.eur.nl
Erasmus University Rotterdam

ERASMUS UNIVERSITEIT ROTTERDAM

Introduction

- Facet selection in multifaceted search
- We propose and evaluate several automatic facet selection methods
  - aim: effectively partition the search space and allow the user to drill down in the least amount of time

Problem statement

- Based on problem statement in [7, 9]
- Four underlying assumptions (similar to [7])
  - Users have a particular target product in mind
  - Search session ends if target appears in top-10
  - User knows if target product has displayed facet
  - Users select facet(s) that the target product has

Problem statement

- Utility of displaying a set of facets $F_p \subseteq F$, using approach M, with respect to query q and already selected facets S ([7]):

$$
\begin{aligned}
U^M_{q,S}(F_p) &= E[X|q,S] - E_M[X|q,S,F_p] \\
&= \sum_{\substack{d \in D_{q,S} \\ r^S_q(d) > m}} p(d = d_q) r^S_q(d) - E_M[X|q,S,F_p]
\end{aligned}
$$

Problem statement

- Using the previous definition, we can formulate the *k* facet selection problem as:

$$F^{*}_{p,S,M} = \arg \max_{\substack{F_p \subseteq \mathcal{F} \\ |F_p| < k}} U^{M}_{q,S}(F_p)$$

- Problem is NP-hard, we use an iterative greedy algorithm [9]

Facet selection methods

- Best Facet I estimation ([9])

$$U_{q,S}^{M_{BI}}(F_p) = E[X|q,S] - E_{M_{BI}}[X|q,S,F_p]$$
$$= \sum_{\substack{d \in D_{q,S} \\ r_q^S(d) > m}} p(d = d_q) \max_{f \in F_{p,S,d}} |D_{q,S,d}^f|$$

with

$$F_{p,S,d} = F_{p,S} \cap C(d), \text{ and}$$

$$D_{q,S,d}^f = \{d' \in D_{q,S} : r_q^S(d') < r_q^S(d) \wedge f \notin C(d')\}$$

Facet selection methods

- Conjunctive estimation ([9])

$$U_{q,S}^{M_C}(F_p) = E[X|q,S] - E_{M_C}[X|q,S,F_p]$$
$$= \sum_{d \in D_{q,S} \wedge r_q^S(d) > m} p(d = d_q) \cdot |D_{q,S,d}^{F_{p,S,d}}|$$

with

$$D_{q,S,d}^{F_{p,S,d}} = \{d' \in D_{q,S} : r_q^S(d') < r_q^S(d) \wedge F_{p,S,d} \not\subseteq C(d')\}$$

Facet selection methods

- In order to investigate the influence of the ranking, we experiment with Best Facet II

$$
\begin{aligned}
U_{q,S}^{M_{BII}}(F_p) &= E[X|q,S] - E_{M_{BII}}[X|q,S,F_p] \\
&= \sum_{\substack{d \in D_{q,S} \\ r_q^S(d) > m}} \max_{f \in F_{p,S,d}} p(d = d_q) \cdot |D_{q,S,d}^f|
\end{aligned}
$$

Facet selection methods

- Two hybrid methods:
  - Probabilistic Entropy
    - conjunctive / entropy approach
  - Probabilistic Conjunctive
    - conjunctive / random approach

Facet selection methods

- Baseline methods
  - Entropy approach greedily selects facets based on highest entropy (i.e., facets that evenly split the results)
  - Weighted Residuals Coverage, 'Most Probable' heuristic in [7]
  - Greedy count, 'Most Frequent' heuristic in [7]

Evaluation

- Data set of 980 products and 487 facets (key/value pairs)
- Experimental setup
  - 1000 generated queries
  - many target products
  - 3, 5, or 7 facets to display
  - 3 user simulation strategies [9]

Evaluation

- Evaluated measures
  - average number of clicks
  - average total utility
  - top-10 promotion percentage
- Statistical tests for difference in mean
  - Bonferroni corrected p-value
    => $0.05/972 = 5.144033 \times 10^{-05}$

Results

- Average number of clicks
  - Greedy count scores the worst
  - The Prob. Conjunctive (P.C.) and Prob. Entropy (P.E.) score best (< 3 clicks on average)
  - For ‘select all’ strategy, we see a decrease with the number of facets

Results

- Average utility
  - No significant patterns across number of selected facets
  - Methods that strongly depend on ranking, perform poorly for l = 3 (e.g., Best Facet II)
  - Prob. Conjunctive and Prob. Entropy again score the best

Results

- Top-10 Promotion Percentage
  - P.C. and P.E score best, with P.C. > P.E.
  - However, for the probabilistic strategy there is no clear winner (results are not statistically significant)
  - Furthermore: performance of P.C. and P.E seems to decrease with the number of total selected facets

Conclusions

- Main conclusion:
  - Hybrid (dithering) methods perform best
  - The Conjunctive algorithm, shown to be superior in the work of [9], can be improved by escaping from local minima.
- Future work:
  - ranking facets vs. selecting, evaluation with user study, and learning from click behavior

Thank you.

Damir Vandic
vandic@ese.eur.nl

ERASMUS UNIVERSITEIT ROTTERDAM