

Μία αξιωματική προσέγγιση για τη διαφοροποίηση των αποτελεσμάτων



ΜΑΘΗΜΑ

Ανάκτηση Πληροφορίας

Παππάς Χρήστος
Ιωάννινα, Ιανουάριος 2010

Διάρθρωση

■ Εισαγωγή

- Πρόβλημα
- Σημαντικότητα
- Ενδιαφέροντα θέματα

■ Τεχνικό περιεχόμενο

- Ορισμοί εννοιών
- Μοντελοποίηση προβλήματος
- Συναρτήσεις-Αλγόριθμοι
- Ειδικές περιπτώσεις συναρτήσεων απόστασης

Διάρθρωση(συνέχεια...)

■ Πειραματική αποτίμηση

- Μέτρα εκτίμησης
- Semantic disambiguation
- Product disambiguation

■ Κριτική-Επεκτάσεις εργασίας

- Πλεονεκτήματα
- Μειονεκτήματα
- Επεκτάσεις

ΕΙΣΑΓΩΓΗ

Πρόβλημα

- Ένα σύστημα αναζήτησης εγγράφων, βάσει επερωτήσεων, είναι αποτελεσματικό όταν ικανοποιεί τις ανάγκες του χρήστη.
- Όταν υπάρχει ασάφεια των επιδιώξεών του, τότε το σύστημα θα πρέπει να διαφοροποιήσει τα αποτελέσματα.
 - Τα αποτελέσματα θέλουμε να αντιστοιχούν ιδανικά σε κάθε πιθανό σκοπό του χρήστη.
- Το σύστημα δεν αρκεί να επιστρέφει τα πιο «σχετικά» έγγραφα, αλλά να συνυπολογίζει και την ποικιλομορφία τους.

Σημαντικότητα

- Ο χρήστης είναι πιο πιθανό να ικανοποιηθεί.
 - Τα αποτελέσματα καλύπτουν διαφορετικές πιθανές προθέσεις του.
- Το σύστημα δεν κάνει «τυχαία» πρόβλεψη
 - Τυχαία με την έννοια της επιλογής πιο πιθανού σκοπού χρήστη, βάση πιθανοτικού μοντέλου.
- Το σύστημα λαμβάνει υπόψη διαφορετικές πιθανές επιδιώξεις του χρήστη.

Ενδιαφέροντα θέματα

- Πρόκειται για πρόβλημα βελτιστοποίησης 2 κριτηρίων
 - Σχετικότητα
 - Ποικιλομορφία
- Συνδυασμός ranking-clustering.
- Ανάπτυξη πολλών συναρτήσεων διαφοροποίησης και αντίστοιχων αλγορίθμων βελτιστοποίησής τους.

ΤΕΧΝΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ

Ορισμοί εννοιών

- Σύνολο εγγράφων $U = \{u_1, u_2, \dots, u_n\}$ $n \geq 2$
- Σύνολο ερωτήσεων Q .
- Δοθέντων ερώτησης $q \in Q$ και ακεραίου k , θέλουμε ένα υποσύνολο $S_k \subseteq U$ που να είναι και «σχετικό» και «πρωτότυπο».
- «Σχετικότητα» ορίζεται: $w: U \times Q \rightarrow \mathbb{R}^+$
- «Πρωτοτυπία» ορίζεται: $d: U \times U \rightarrow \mathbb{R}^+$
 - Μεγάλη απόσταση σημαίνει μεγάλη πρωτοτυπία.

Ορισμοί εννοιών(συνέχεια...)

- Συνάρτηση επιλογής υποσυνόλου

$$f : 2^U \times Q \times w \times d \rightarrow \mathbb{R}$$

- Στόχος

Δοθέντων q, w, d και ακεραίου k , να βρούμε το υποσύνολο $S_k \subseteq U$ τέτοιο ώστε:

$$S_k^* = \operatorname{argmax}_{S_k \subseteq U} f(S_k, q, w(\cdot), d(\cdot, \cdot))$$

Μοντελοποίηση προβλήματος

- Ορίζεται ένα σύνολο αξιωμάτων
 - Αναμένουμε να ικανοποιούνται από ένα σύστημα διαφοροποίησης.
- Έπειτα ορίζουμε συναρτήσεις διαφοροποίησης
 - Ικανοποιούν διαφορετικά υποσύνολα των ορισθέντων αξιωμάτων
 - Δεν υπάρχει συνάρτηση που να ικανοποιεί όλα τα αξιώματα
 - Αξιώματα αποτελούν βάση για σύγκριση συναρτήσεων-συστημάτων

Μοντελοποίηση προβλήματος (συνέχεια...)

■ Αξιώματα

- **Scale invariance:** Η f δεν επηρεάζεται από κλιμάκωση σε σχετικότητα και απόσταση

$$S_k^* = \operatorname{argmax}_{S_k \subseteq U} f(S_k, q, \alpha \cdot w(\cdot), \alpha \cdot d(\cdot, \cdot))$$

- **Consistency:** Δοθέντων $\alpha : U \rightarrow \mathbb{R}^+$ and $\beta : U \times U \rightarrow \mathbb{R}^+$, αλλάζουμε τις w, d ως εξής:

$$w(u) = \begin{cases} w(u) + \alpha(u) & , u \in S_k^* \\ w(u) - \alpha(u) & , \text{otherwise} \end{cases} \quad d(u, v) = \begin{cases} d(u, v) + \beta(u, v) & , u, v \in S_k^* \\ d(u, v) - \beta(u, v) & , \text{otherwise} \end{cases}$$

Η f μεγιστοποιείται πάλι από το ίδιο σύνολο

Μοντελοποίηση προβλήματος (συνέχεια...)

■ Αξιώματα(συνέχεια...)

- **Richness**: Πρέπει να υπάρχουν συναρτήσεις w, d ώστε, για οποιοδήποτε $k \geq 2$, να υπάρχει μοναδικό S_k^* που μεγιστοποιεί την f .
- **Stability**: Η f πρέπει να οριστεί έτσι ώστε $S_k^* \subseteq S_{k+1}^*$
- **Independence of irrelevant attributes**: Δοθέντος συνόλου S , θέλουμε το $f(S)$ να είναι ανεξάρτητο από:
 $w(u)$ for all $u \notin S$ $d(u, v)$ for all $u, v \notin S$
- **Monotonicity**: Δοθέντων $w(\cdot), d(\cdot, \cdot), f$ and $S \subseteq U$,
θέλουμε $f(S \cup \{x\}) \geq f(S)$ για κάθε $x \notin S$

Μοντελοποίηση προβλήματος (συνέχεια...)

■ Αξιώματα(συνέχεια...)

- **Strength of relevance:** Η f δεν αγνοεί τη συνάρτηση σχετικότητας. Δοθέντων $w(\cdot)$, $d(\cdot, \cdot)$, f and S , πρέπει για κάθε $x \in S$:

- Υπάρχουν $\delta_0 > 0$ and $a_0 > 0$ ώστε: $f(S, w'(\cdot), d(\cdot, \cdot), k) = f(S, w(\cdot), d(\cdot, \cdot), k) + \delta_0$

Η w' προέκυψε από w με την αλλαγή $w'(x) = a_0 + w(x)$

- Αν $f(S \setminus \{x\}) < f(S)$, υπάρχουν $\delta_1 > 0$ and $a_1 > 0$ ώστε:

$$f(S, w'(\cdot), d(\cdot, \cdot), k) = f(S, w(\cdot), d(\cdot, \cdot), k) - \delta_1$$

Η w' προέκυψε από w με την αλλαγή $w'(x) = a_1 + w(x)$

Μοντελοποίηση προβλήματος (συνέχεια...)

■ Αξιώματα(συνέχεια...)

- **Strength of similarity:** Η f δεν αγνοεί τη συνάρτηση ομοιότητας(απόσταση). Δοθέντων $w(\cdot)$, $d(\cdot, \cdot)$, f and S , πρέπει για κάθε $x \in S$:

- Υπάρχουν $\delta_0 > 0$ and $b_0 > 0$ ώστε: $f(S, w(\cdot), d'(\cdot, \cdot), k) = f(S, w(\cdot), d(\cdot, \cdot), k) + \delta_0$

Η d' προέκυψε από d με την αλλαγή ότι αυξάνουμε τα

$d(x, u)$ ώστε $\min_{u \in S} d(x, u) = b_0$

- Αν $f(S \setminus \{x\}) < f(S)$, υπάρχουν $\delta_1 > 0$ and $b_1 > 0$ ώστε:

$$f(S, w(\cdot), d'(\cdot, \cdot), k) = f(S, w(\cdot), d(\cdot, \cdot), k) - \delta_1$$

Η d' προέκυψε από d με την αλλαγή ότι μειώνουμε τα

$d(x, u)$ ώστε $\max_{u \in S} d(x, u) = b_1$

Συναρτήσεις-Αλγόριθμοι

■ **Max-sum diversification**

Μεγιστοποίηση αθροίσματος σχετικότητας και απόστασης.

$$■ f(S) = (k-1) \sum_{u \in S} w(u) + 2\lambda \sum_{u,v \in S} d(u,v)$$

- $\lambda > 0$ ρυθμίζει το trade-off

Χαρακτηρισμός: Ικανοποιεί όλα τα αξιώματα, εκτός από το **stability**.

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Max-sum diversification**(συνέχεια...)

Βελτιστοποίηση με αναγωγή στο

MaxSumDispersion πρόβλημα του **facility dispersion**.

- $f(S) = \sum_{u,v \in S} d'(u,v)$

- Ανάγεται στο $f(S) = (k-1) \sum_{u \in S} w(u) + 2\lambda \sum_{u,v \in S} d(u,v)$, αν θεωρήσουμε $d'(u,v) = w(u) + w(v) + 2\lambda d(u,v)$

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Max-sum diversification**(συνέχεια...)

Το MaxSumDispersion είναι NP-hard.

Υπάρχει 2-προσεγγιστικός αλγόριθμος

- Δεδομένου ότι d' είναι **metric**

```
Input  : Universe  $U$ ,  $k$ 
Output: Set  $S$  ( $|S| = k$ ) that maximizes  $f(S)$ 

Initialize the set  $S = \emptyset$ 
for  $i \leftarrow 1$  to  $\lfloor \frac{k}{2} \rfloor$  do
  Find  $(u, v) = \operatorname{argmax}_{x, y \in U} d(x, y)$ 
  Set  $S = S \cup \{u, v\}$ 
  Delete all edges from  $E$  that are incident to  $u$  or  $v$ 
end
If  $k$  is odd, add an arbitrary document to  $S$ 
```

Algorithm 1: Algorithm for MAXSUMDISPERSION

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Max-min diversification**

Μεγιστοποίηση ελάχιστης σχετικότητας και απόστασης.

- $f(S) = \min_{u \in S} w(u) + \lambda \min_{u, v \in S} d(u, v)$

- $\lambda > 0$ ρυθμίζει το trade-off

Χαρακτηρισμός: Ικανοποιεί όλα τα αξιώματα, εκτός από τα **consistency** και **stability**.

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Max-min diversification**(συνέχεια...)

Βελτιστοποίηση με αναγωγή στο

MaxMinDispersion πρόβλημα του **facility dispersion**.

- $g(P) = \min_{v_i, v_j \in P} d(v_i, v_j)$

- Ανάγεται στο $f(S) = \min_{u \in S} w(u) + \lambda \min_{u, v \in S} d(u, v)$, αν

θεωρήσουμε $d'(u, v) = \frac{1}{2}(w(u) + w(v)) + \lambda d(u, v)$

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Max-min diversification**(συνέχεια...)

To MaxMinDispersion είναι NP-hard.

Υπάρχει 2-προσεγγιστικός αλγόριθμος

- Δεδομένου ότι d' είναι **metric**

```
Input : Universe  $U$ ,  $k$   
Output: Set  $S$  ( $|S| = k$ ) that maximizes  $f(S)$   
  
Initialize the set  $S = \emptyset$ ; Find  
 $(u, v) = \operatorname{argmax}_{x, y \in U} d(x, y)$  and set  $S = \{u, v\}$ ; For  
any  $x \in U \setminus S$ , define  $d(x, S) = \min_{u \in S} d(x, u)$ ;  
while  $|S| < k$  do  
    Find  $x \in U \setminus S$  such that  $x = \operatorname{argmax}_{x \in U \setminus S} d(x, S)$ ;  
    Set  $S = S \cup \{x\}$ ;  
end
```

Algorithm 2: Algorithm for MAXMINDISPERSION

Συναρτήσεις-Αλγόριθμοι(συνέχεια...)

■ **Mono-objective formulation**

Ορίζει μία τιμή για κάθε έγγραφο που συνδυάζει σχετικότητα και απόσταση

- $w'(u) = w(u) + \frac{\lambda}{|U|-1} \sum_{v \in U} d(u, v)$

- $\lambda > 0$ ρυθμίζει το trade-off

Μεγιστοποιεί την $f(S) = \sum_{u \in S} w'(u)$

Χαρακτηρισμός: Ικανοποιεί όλα τα αξιώματα, εκτός από το **consistency**.

Ειδικές περιπτώσεις συναρτήσεων απόστασης

■ Semantic distance

Χρήση στο πρώτο πείραμα

- Αφορά web pages

$d(u, v) = 1 - \text{sim}(u, v)$, όπου u, v ιστοσελίδες

Θεωρούμε Jaccard ομοιότητα: $\text{sim}(u, v) = \frac{|S(u) \cap S(v)|}{|S(u) \cup S(v)|}$

Για το S

- Hash function $h: U \rightarrow [0, 1]$
- Min hash συνόλου $A=U$ ορίζεται: $MH_h(A) = \underset{x \in A}{\text{argmin}} \{h(x)\}$
 - Επέκταση για πολυσύνολο: $MH(A) = \underset{i}{\text{argmin}} \{h(x, i) | x \in A, 1 \leq i \leq c_x\}$
 - c_x συχνότητα στοιχείου x
- Τότε, sketch σελίδας $d: S(d) = \{MH_{h_1}(d), MH_{h_2}(d), \dots, MH_{h_k}(d)\}$
 - h_1, h_2, \dots, h_k hash functions

Ειδικές περιπτώσεις συναρτήσεων απόστασης(συνέχεια...)

■ **Categorical distance**

Χρήση στο δεύτερο πείραμα

- Αφορά προϊόντα

Αρχικά, ταξινομούνται τα προϊόντα σε κατηγορίες.

- Δημιουργείται δέντρο κατηγοριών

$$d(u, v) = \sum_{i=1}^{l(u)} \frac{1}{2^{e(i-1)}} + \sum_{i=1}^{l(v)} \frac{1}{2^{e(i-1)}} \quad \begin{array}{l} \blacksquare u, v \text{ κατηγορίες} \\ \blacksquare e \geq 0 \end{array}$$

- $l(\cdot)$ βάθος κατηγορίας στο δέντρο

$$d_c(x, y) = \sum_{u \in C_x, v \in C_y} \min(C_x(u), C_y(v)) \operatorname{argmin}_v d(u, v)$$

- Επέκταση για προϊόντα πολλαπλών κατηγοριών
- C_x, C_y διανύσματα με κατηγορικές πιθανότητες

ΠΕΙΡΑΜΑΤΙΚΗ ΑΠΟΤΙΜΗΣΗ

Μέτρα εκτίμησης

- Σκοπός να χαρακτηρίσουμε την επιλογή συνάρτησης διαφοροποίησης
 - Διενέργεια 2 πειραμάτων
 - Αποτίμηση κατάλληλων μέτρων

- Μέτρα
 - πρωτοτυπία(novelty)
 - σχετικότητα(relevance)

Semantic disambiguation

■ Πρώτο πείραμα

- Q αμφίσημες wikipedia σελίδες.
- S_q θέματα που σχετίζονται με σελίδα q .
- $p_q(x, s)$ πιθανότητα ένα αποτέλεσμα x να αναπαριστά ένα θέμα $s \in S_q$
 - $p_q(x, s)$ υπολογίζεται μέσω **semantic distance**

■ Διαδικασία διαφοροποίησης(υπενθύμιση)

- Ανακτούνται τα top n έγγραφα $R(q)$, ερώτησης $q \in Q$, από τη μηχανή αναζήτησης.
- Εφαρμογή αλγορίθμου διαφοροποίησης για να πάρουμε τα top k διαφοροποιημένα έγγραφα $D(q)$.

Semantic disambiguation(συνέχεια...)

- Συγκρίνουμε τα $D(q)$ και $R_k(q)$ με χρήση των **novelty** και **relevance**

- $R_k(q)$ τα πρώτα k έγγραφα της $R(q)$.

■ Novelty

- Εκτίμηση αριθμού θεμάτων, ερώτησης q , που καλύπτονται σε μία λίστα αποτελεσμάτων L .

- $$\text{Novelty}_q(L) = \frac{1}{|S_q|} \sum_{s \in S_q} \mathbf{I} \left(\sum_{x \in L} p_q(x, s) > \theta \right)$$

- θ κατώφλι

- \mathbf{I} συνάρτηση ίση με 1 ή 0(συνθήκη A ή Ψ ,αντίστοιχα)

Semantic disambiguation(συνέχεια...)

■ Relevance

- Εκτίμηση σχετικότητας, ερώτησης q , των αποτελεσμάτων λίστας S

- $$R(S, q) = \sum_{s \in S'} \left| \frac{1}{r_s} - \frac{1}{r'_s} \right|$$

- S' η ιδανική wikipedia κατάταξη
- r_s, r'_s η διάταξη του s στις S και S' , αντίστοιχα
- Το r_s της $R_k(q)$ βρίσκεται κάνοντας αναζήτηση περιορισμένη στα wikipedia sites
- Το r_s της $D(q)$ βρίσκεται ταξινομώντας βάσει

$$Rel(s, q) = \sum_{d \in D(q)} \frac{1}{pos(d)} p_q(d, s), \text{ όπου } pos(d) \text{ η θέση του } d \text{ στην } D(q)$$

Semantic disambiguation(συνέχεια...)

■ Συγκρίνουμε τα $D(q)$ και $R_k(q)$

■ Novelty
$$FN_q = \frac{\text{Novelty}_q(D(q)) - \text{Novelty}_q(R_k(q))}{\max(\text{Novelty}_q(D(q)), \text{Novelty}_q(R_k(q)))}$$

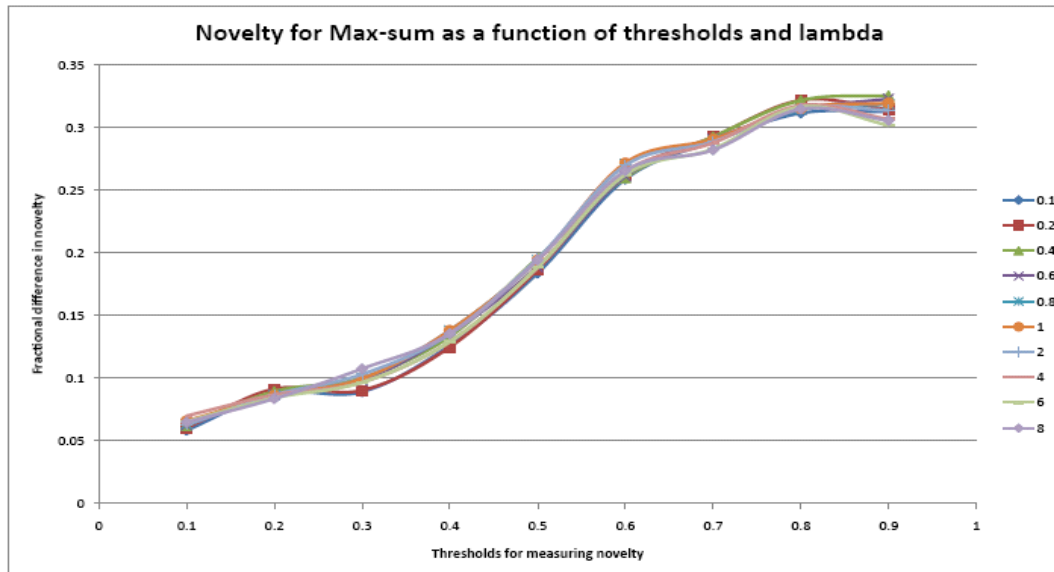
■ Relevance
$$FR_q = \frac{\text{Relevance}_q(D(q)) - \text{Relevance}_q(R_k(q))}{\max(\text{Relevance}_q(D(q)), \text{Relevance}_q(R_k(q)))}$$

■ $\text{Relevance}_q(S) = R(S, q)$

Semantic disambiguation(συνέχεια...)

■ Αποτελέσματα

- $n=30$, $k=10$ και $|Q|=100$



- $FN_q > 0$

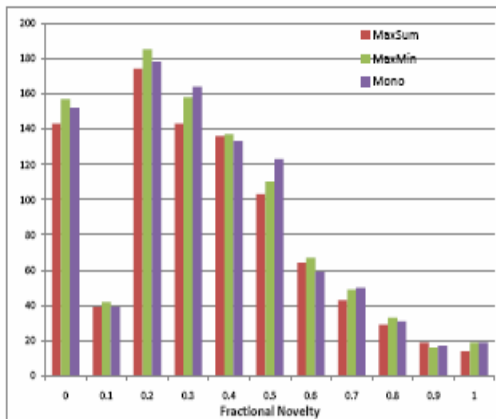
- FN_q αυξάνει με λ

- FN_q αυξάνει με θ

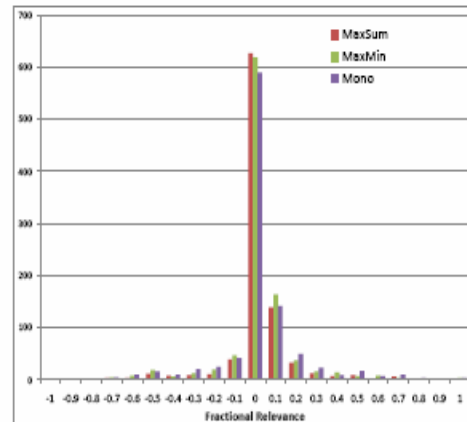
Figure 1: [Best viewed in color] The effect of varying the value of the trade-off parameter λ , and the threshold for measuring novelty on the output of the search results from MAXSUMDISPERSION.

Semantic disambiguation(συνέχεια...)

■ Αποτελέσματα(συνέχεια...)



(a) Novelty



(b) Relevance

a)-MaxMinDispersion

-75% queries produce more diversified results

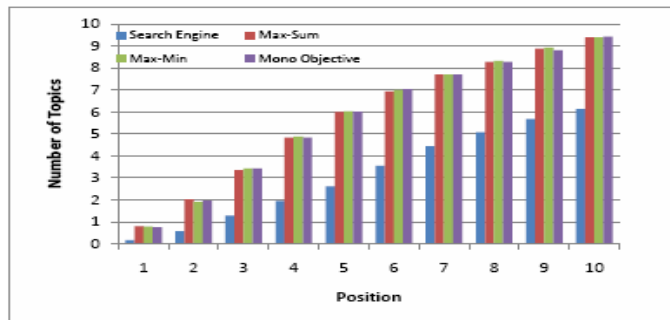
b)-MonoObjective

-Γενικά καλό ranking

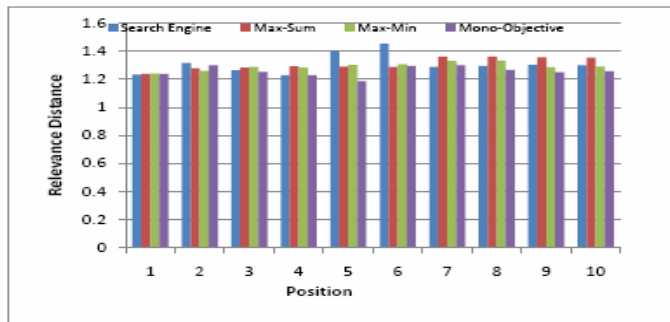
Figure 2: [Best viewed in color] The histogram of fractional difference in novelty (a) and relevance (b) plotted over a 1000 ambiguous queries.

Semantic disambiguation(συνέχεια...)

■ Αποτελέσματα(συνέχεια...)



(a) Novelty



(b) Relevance

$\lambda=1.0$ και $\theta=0.5$

a)-Topics αυξάνει με position

-Αλγόριθμοι υπερτερούν μηχανής αναζήτησης

-Ισάξιοι μεταξύ τους

b)-Διαφορές αν position ≥ 5

-MonoObjective μικρότερη distance

-MaxSumDispersion μεγαλύτερη distance

Figure 3: [Best viewed in color] The positional variation in the novelty (a) and relevance (b) of the result set.

Product disambiguation

■ Δεύτερο πείραμα

- 100 ερωτήσεις προϊόντων.
- 50 πρώτα αποτελέσματα, βάσει δημοτικότητας.
- χρήση **categorical distance**

■ Διαδικασία διαφοροποίησης(υπενθύμιση)

- Ανακτούνται τα top n έγγραφα $R(q)$, ερώτησης $q \in Q$, από τη μηχανή αναζήτησης.
- Εφαρμογή αλγορίθμου διαφοροποίησης για να πάρουμε τα top k διαφοροποιημένα έγγραφα $D(q)$.

Product disambiguation(συνέχεια...)

- Συγκρίνουμε τα $D(q)$ και $R_k(q)$ με χρήση των **novelty** και **relevance**
 - $R_k(q)$ τα πρώτα k έγγραφα της $R(q)$.
- **Novelty**
 - Εκτίμηση αριθμού κατηγοριών προϊόντων, ερώτησης q , που καλύπτονται σε μία λίστα αποτελεσμάτων L .
 - $$\text{Novelty}_q(L) = \frac{2}{|L|(|L|-1)} \sum_{u,v \in L} I(\text{lca}(u,v) \notin \{u,v\})$$
 - lca ο ελάχιστος κοινός προκάτοχος
 - I συνάρτηση ίση με 1 ή 0(συνθήκη A ή Ψ ,αντίστοιχα)

Product disambiguation(συνέχεια...)

■ Relevance

- Εκτίμηση σχετικότητας ερώτησης q και αποτελεσμάτων λίστας L

- Δεν υπάρχει ground truth(π.χ. wikipedia)
- Χρήση απόστασης των αντίστοιχων κατηγοριών στην ταξινόμηση

- $$\text{Relevance}_q(L) = \frac{1}{|L|} \sum_{u \in L} \frac{1 + d(\text{lca}(q, u), q)}{\text{pos}(u)}$$

- $\text{pos}(u)$ η θέση του u στην L
- d η categorical distance

Product disambiguation(συνέχεια...)

■ Αποτελέσματα

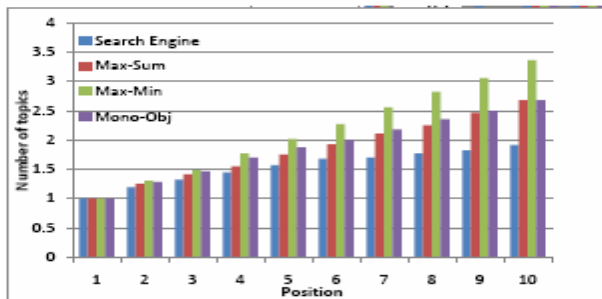
Product Search Engine	Diversified Results
Sony SCD CE595-SACD changer	Sony SCD CE595-SACD changer
Sony CDP CE375-CD changer	Sony CDP CE375-CD changer
Sony CDP CX355-CD changer	Teac SR-L50-CD player/radio
Teac SR-L50- CD player/radio	Bose Wave Music System Multi-CD Changer
Bose Wave Music System Multi-CD Changer	Sony S2 Sports ATRAC3/MP3 CD Walkman D-NS505
Sony RCD-W500C-CD changer/CD recorder	COBY CX CD109-CD player
Sony CD Walkman D-EJ011-CD player	JVC XL PG3-CD player
Sony S2 Sports ATRAC3/MP3 CD Walkman D-NS505	Pioneer PD M426-CD changer
Sony Atrac3/MP3 CD Walkman D-NF430	Sony SCD XA9000ES-SACD player
COBY CX CD109-CD player	Yamaha CDR HD1500-CD recorder/HDD recorder

- Diversified results contain more product brands
- Relative positions of brands still remains

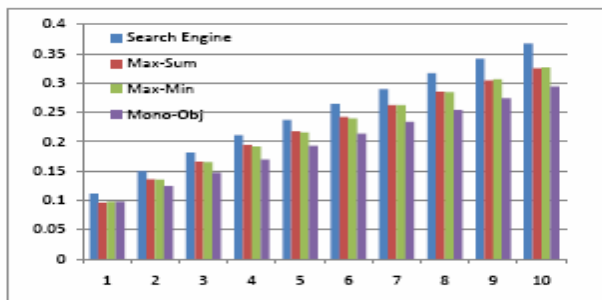
Table 1: The difference in the top 10 results for the query cd player from a commercial product search engine and the diversified results produced by running MAXSUMDISPERSION with $n = 30$ and $k = 10$

Product disambiguation(συνέχεια...)

■ Αποτελέσματα(συνέχεια...)



(a) Diversity



(b) Relevance

a)-MaxMinDispersion
-Αλγόριθμοι υπερτερούν
μηχανής αναζήτησης

b)-Search engine
-MonoObjective παραδόξως
δεν αυξάνει relevance
συγκριτικά με τους άλλους

Figure 4: [Best viewed in color] The positional variation in the diversity (a) and relevance (b) of the product result set averaged over 100 queries with $n = 30$, $k = 10$, $\lambda = 1.0$ and $\theta = 0.5$.

ΚΡΙΤΙΚΗ-
ΕΠΕΚΤΑΣΕΙΣ
ΕΡΓΑΣΙΑΣ

Πλεονεκτήματα

- Ο τυχαίος χρήστης είναι πιο πιθανό να ικανοποιηθεί σε σχέση με τα αποτελέσματα της μηχανής αναζήτησης.
- Η χρήση των αξιωμάτων επιτρέπει ένα θεωρητικό χαρακτηρισμό των συναρτήσεων διαφοροποίησης, ανεξάρτητα από τις συναρτήσεις «σχετικότητας» και «απόστασης».
- Τα μέτρα πρωτοτυπίας και σχετικότητας, κατά την πειραματική ανάλυση, ποσοτικοποιούν κατάλληλα την απόδοση των συναρτήσεων.

Μειονεκτήματα

- Η πειραματική αξιολόγηση εξαρτάται έμμεσα από τα αξιώματα, καθώς οι συναρτήσεις διαφοροποίησης ικανοποιούν διαφορετικά σύνολα αξιωμάτων.
- Οι λύσεις των 2 αλγορίθμων διαφοροποίησης είναι προσεγγιστικές, καθώς το γενικότερο πρόβλημα του «facility dispersion» είναι NP-hard.
- Δεν υπάρχει κάποια διάταξη των αξιωμάτων ως προς τη σημαντικότητά τους, αλλά μας ενδιαφέρει μόνο το ποιά ικανοποιούνται.

Επειτάσεις

- Έλεγχος ικανοποίησης και των 8 αξιωμάτων, στην περίπτωση που η απόσταση είναι «metric».
- «Χαλάρωση» κάποιων αξιωμάτων(π.χ. stability), προκειμένου να διευκολύνουμε την εύρεση κατάλληλων συναρτήσεων διαφοροποίησης.
- Εύρεση νέων συναρτήσεων, που θα ανάγονται στο «facility dispersion» και θα είναι βέλτιστες για συγκεκριμένες περιοχές(π.χ. web search).

Επειτάσεις(συνέχεια...)

- Εισαγωγή βαρών σημαντικότητας στα αξιώματα που ορίζονται. Έτσι, προτιμάται η ικανοποίηση αξιωμάτων με μεγαλύτερα βάρη
 - Μεγάλο βάρος σημαίνει ότι το αξίωμα ικανοποιεί καλύτερα τους στόχους συστήματος διαφοροποίησης.
- Αξιοποιώντας τα novelty και relevance, εύρεση κατάλληλων μέτρων με συνδυασμό τους. Τα νέα αυτά μέτρα θα στοχεύουν στην καλύτερη αξιολόγηση της απόδοσης των αλγορίθμων.

Ερώτηση

- Το MonoObjective μεγιστοποιεί την $f(S) = \sum_{u \in S} w'(u)$,
όπου $w'(u) = w(u) + \frac{\lambda}{|U|-1} \sum_{v \in U} d(u, v)$

Πώς προκύπτει η βέλτιστη λύση-σύνολο S ?

- Αποδεικνύεται ότι δεν υπάρχει συνάρτηση που να ικανοποιεί και τα 8 αξιώματα. Πώς σχετίζεται αυτό το γεγονός με την επιλογή συναρτήσεων διαφοροποίησης?

Ευχαριστώ για την προσοχή σας!

ΕΡΩΤΗΣΕΙΣ???