

Do Social Information Help Book Search ?

Ludovic Bonnefoy*, Romain Deveaud* and Patrice Bellot**

* LIA – University of Avignon

** LSIS – Aix-Marseille University

Introduction

- a lot of « social » data available
 - user-generated content, dedicated to others
 - ratings, reviews and tags from Amazon corpus
 - friends, groups, ratings, reviews and tags from user profiles

- are they usefull for book search?

Outline

- Introduction
- Using social information for book search
 - Modeling book likeliness
 - Book thematic relatedness
 - All Amazon's social
- Conclusions and future work

Markov Random Field for IR

– weighting query terms [Metzler2005]

– unigram matches $f_T(q, D) = \log P(q_i|D)$

– bigram exact matches $f_O(q_i, q_{i+k}, D) = \log P(\#1(q_i, \dots, q_{i+k})|D)$

– bigram matches within an unordered window of 8 terms $f_U(q_i, q_{i+k}, D) = \log P(\#uw8(q_i, \dots, q_{i+k})|D)$

$$score_{SDM}(Q, D) = \lambda_T \sum_{q \in Q} f_T(q, D) + \lambda_O \sum_{i=1}^{|Q|-1} f_O(q_i, q_{i+1}, D) + \lambda_U \sum_{i=1}^{|Q|-1} f_U(q_i, q_{i+1}, D)$$

– $\lambda_T = 0.85, \lambda_O = 0.10, \lambda_U = 0.05$

Book likeliness

- documents are Amazon/LibraryThing pages
 - users comment and review and rate products
- intuition #1: a high **reviewed** product must be relevant
 - or at least popular...
 - PageRank-like [Bao2007]
- intuition #2: a high **rated** product must be relevant

$$\mathcal{L}(D) = \log(\#reviews(D)) \times \frac{\sum_{r \in \mathcal{R}_D} r}{\#reviews(D)}$$

Book likeliness (2)

- given a « book need » Q , the score of an Amazon/LibraryThing document D is:

$$s(Q, D) = \mathcal{L}(D) \times score_{SDM}(Q, D)$$

Book Thematic Relatedness

- book search results expected by a user share latent thematic
- these latent thematic are expressed through **users tags**
- given a user « book need », infer the thematic and retrieve books that share the same topic

Book Thematic Relatedness (2)

- build a tag profile for each book
- a tf-idf weighted vector of tags
- tf : number of times the tag is associated to the book
- df : number of books which have this tag associated

Book Thematic Relatedness (3)

- construct a tag profile for the query with a PRF approach
- standard relevance model with all query words
- top x documents used for building the profile (x = 30 for our official runs)

Book Thematic Relatedness (4)

- weight of tags in query tag profile :

$$w(t_i) = \sum_{b \in Top_x} tf.idf(t_i, b)$$

- books may have to contribute to the weight of a tag according to their relevance to the query

$$w(t_i) = \sum_{b \in Top_x} tf.idf(t_i, b) \times score(b, Q)$$

Book Thematic Relatedness (5)

- the top 1000 books retrieved are reranked
- cosine similarity between the tag profile of the query and the tag profile of each book

Book Thematic Relatedness : other experiments

- using the tags from user profiles
- users add books to their catalogue and set their own tags
- experiments on the training set (211 topics)
 - performance decreased by 50%

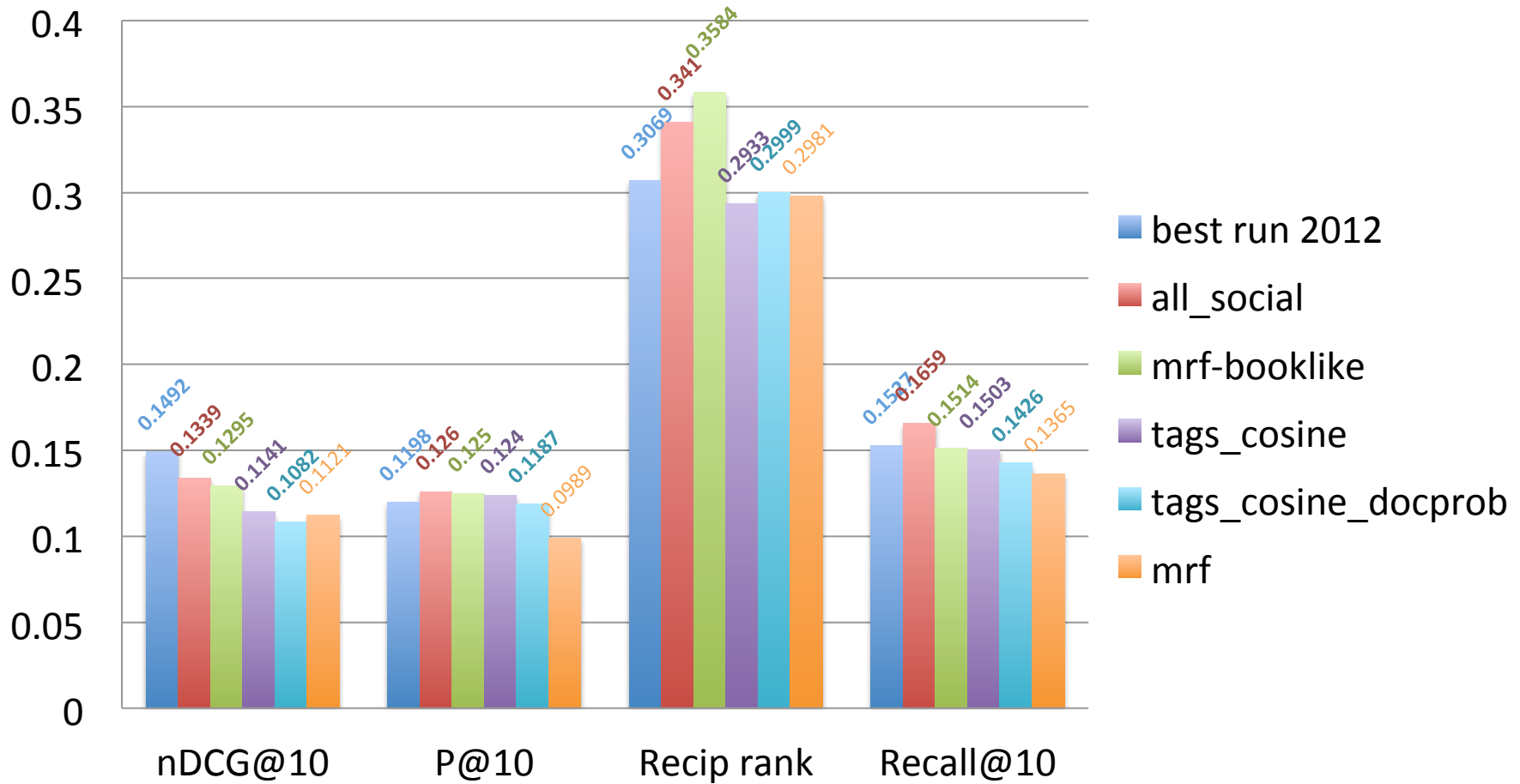
Book Thematic Relatedness : other experiments (2)

- Dewey classification : xxx.yyy
- Dewey profiles instead of tag profiles
 - a book dewey profile : {xxx = 1, yyy=1}
- experiments on the training set (211 topics)
 - performance decreased by 70%

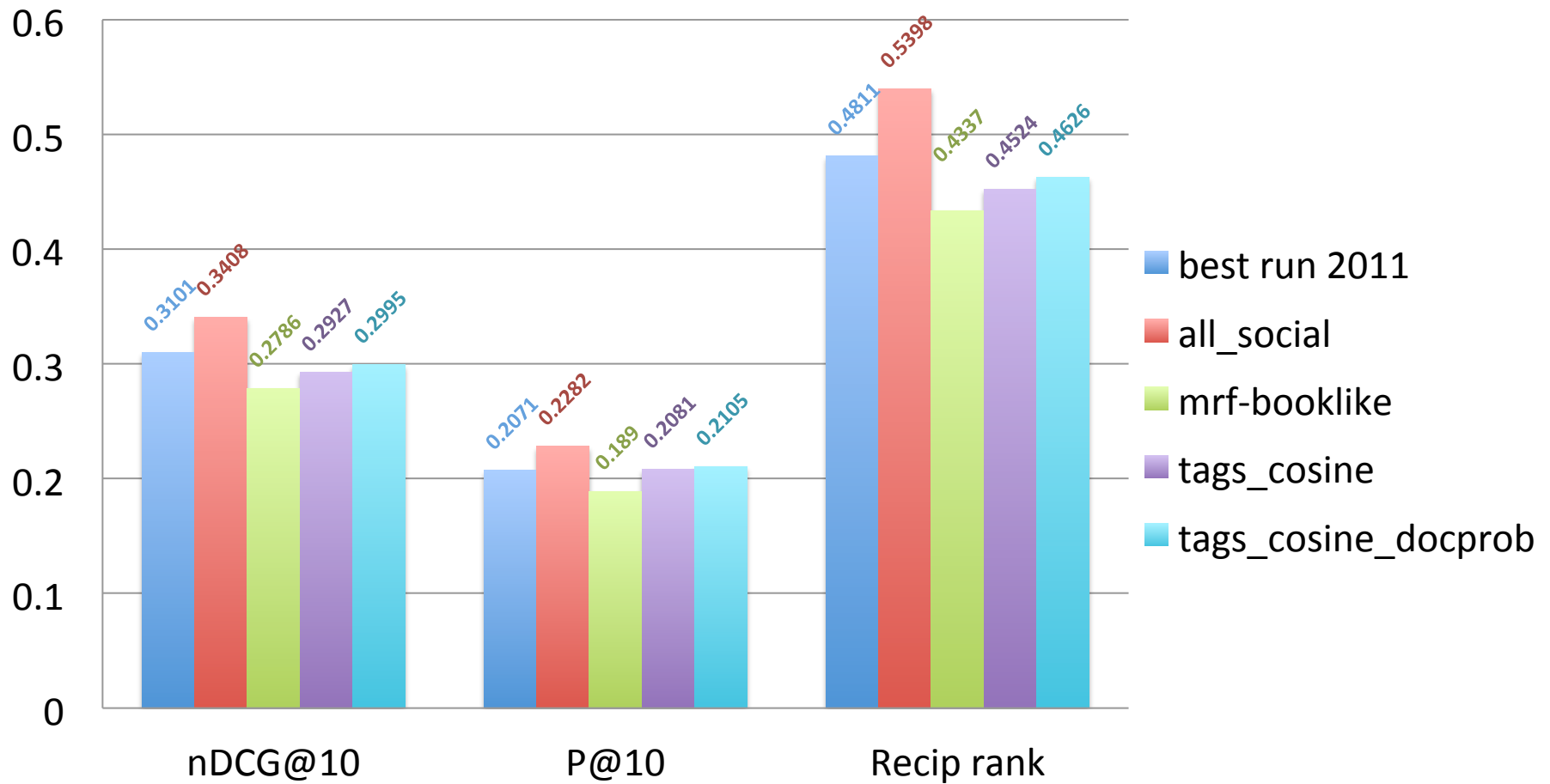
All Amazon's social

- using ratings, reviews and tags from Amazon corpus
- combining Thematic and Book Likelihood
 - logistic regression :
 - the two scores as features
 - two classes : relevant or not
 - training set : top 30 results returned by both approaches
 - relevance judgements from 2011 *qrels*

Official 2012 Evaluation



Unofficial Evaluation on 2011 LibraryThing judgements



Outline

- Introduction
- Using social information for book search
 - Modeling book likeliness
 - Book thematic relatedness
 - All Amazon's social
- **Conclusions and future work**

Conclusions and future work

- Do social information help book search?
 - yes but how much is difficult to estimate
 - the answer seems to depend a lot of the evaluation dataset used
- social ratings and reviews are good indicators of the interest of a book
- books tags can be used to model latent thematics of a user query and to guide book search
 - social : 1 / dewey : 0

Conclusions and future work (2)

- a lot of « social » data remains unexploited
 - ratings, reviews and tags from Amazon corpus
 - **friends, groups, ratings, reviews** and tags from user profiles

thank you for your attention