ONLINE RECOMMENDER SYSTEM FOR RADIO STATION HOSTING: EXPERIMENTAL RESULTS REVISITED

Dmitry I. Ignatov[1], Sergey Nikolenko[1,2], Taimuraz Abaev[1], and Natalia Konstantinova[3]

[1]National Research University Higher School of Economics, Russia

[2]Steklov Institute of Mathematics at St. Petersburg of the RAS, Russia

[3]University of Wolverhampton, UK

IEEE/WIC/ACM International
Conference on Web Intelligence
August 11-14, 2014
Warsaw, Poland

OUTLINE

- FMhost Online Radio Hosting
- Recommender Model
  - Data
  - Model and Algorithms
- Quality of Service Evaluation (QoS)
  - User and Radio Station Activity Analaysis
  - Results for the proposed methods
  - Comparison with SVD-based recommender
- Conclusion

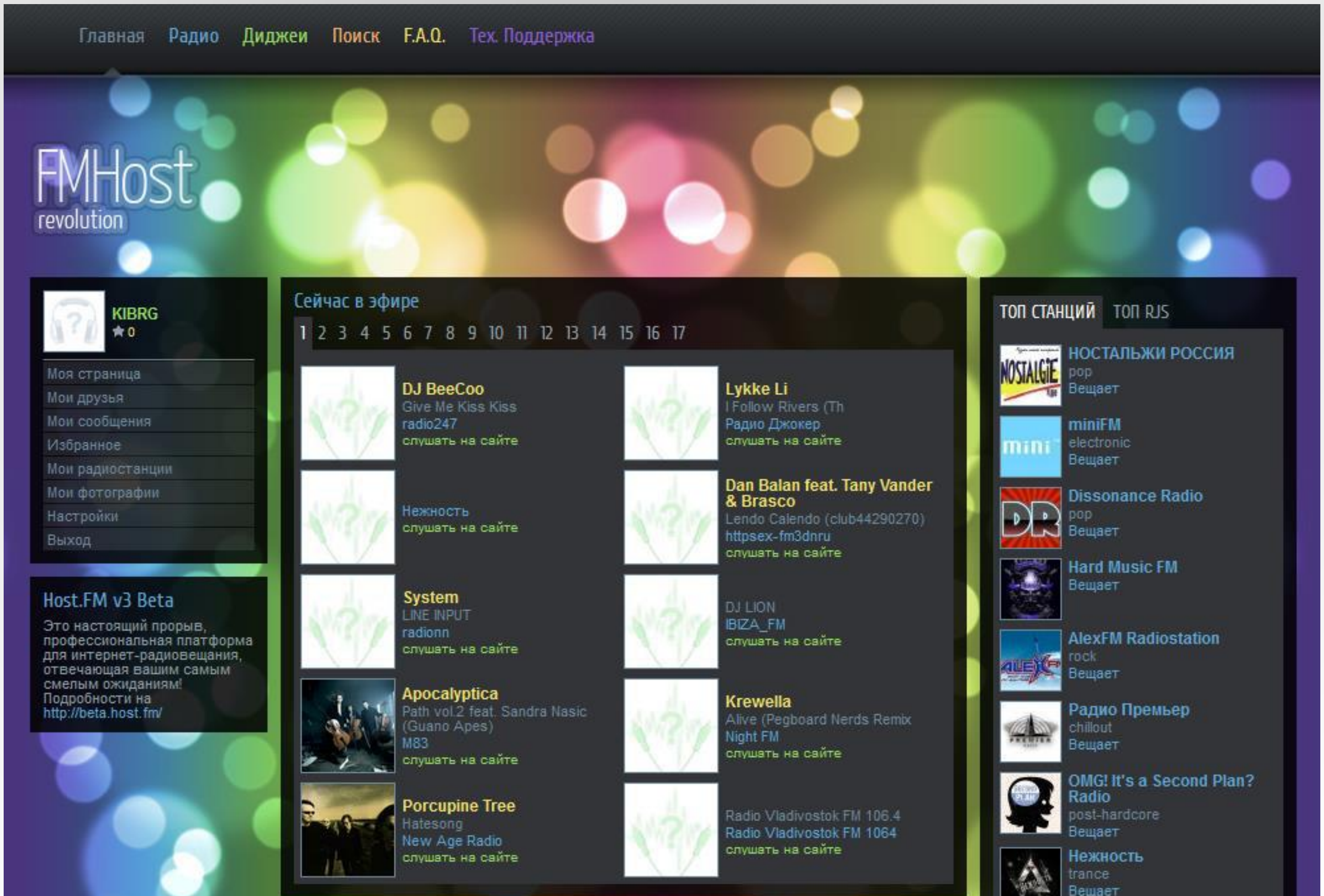
ONLINE RADIO HOSTING FMHOST

- FMhost.me or Host.fm
- Real radio, not a streamer
- Social network
- Lives
- New features
- Listener oriented
- Likes
- Favorites

USERS

- Unauthorized
- Listeners
- DJs
- Station owners

ONLINE RADIO HOSTING FMHOST IN 2012



MUSIC RECOMMENDATION

Conferences and workshops:

- International Society for Music Information Retrieval Conference (ISMIR)
- Recommender Systems Conference (RecSys)
- Workshop on Music Recommendation and Discovery (WOMRAD)

Web services:

- Last.fm
- Pandora
- iTunes

PREVIOUS WORK

Usually methods for music recommendation use quite limited data sources:

- Collaborative filtering exploits only users' ratings
- Acoustic methods relies on acoustic information
- Hybrid approaches combine different methods

PREVIOUS WORK

- B. Logan. Music recommendation from song sets. In Proc. the 5th International Conference on Music Information Retrieval, Barcelona, Spain, 2004.
- O. Celma. Foafing the music: Bridging the semantic gap in music recommendation. In Proc. the 5th International Semantic Web Conference, Athens, Georgia, 2006.
- K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences. In Proc. the 7th International Conference on Music Information Retrieval, Victoria, Canada, 2006.
- S. Pauws, W. Verhaegh, and M. Vossen. Fast generation of optimal music playlists using local search. In Proc. the 7th International Conference on Music Information Retrieval, Victoria, Canada, 2006.
- **Dmitry I. Ignatov, Andrey V. Konstantinov, Sergey I. Nikolenko, Jonas Poelmans, Vasily Zaharchuk: Online Recommender System for Radio Station Hosting. BIR 2012: 1-12.**

THE PREVIOUS ALGORITHM

- Ignatov et al. 2011

Table 1. FMhost's recommender system satisfaction survey.

User opinion	Number of respondents (%)
I like it very much, all recommendations were relevant	54 (49%)
Good, I like most of the radio stations	22 (20%)
Sometimes there are interesting stations	16 (14%)
I like only few recommended radio stations	9(8%)
None of the recommended stations was satisfactory	10 (9%)

MOTIVATION

- It is rare case when different approaches to recommendations are used together (e.g. history of listening and tags)
- Too few research activity in radiostation recommendation

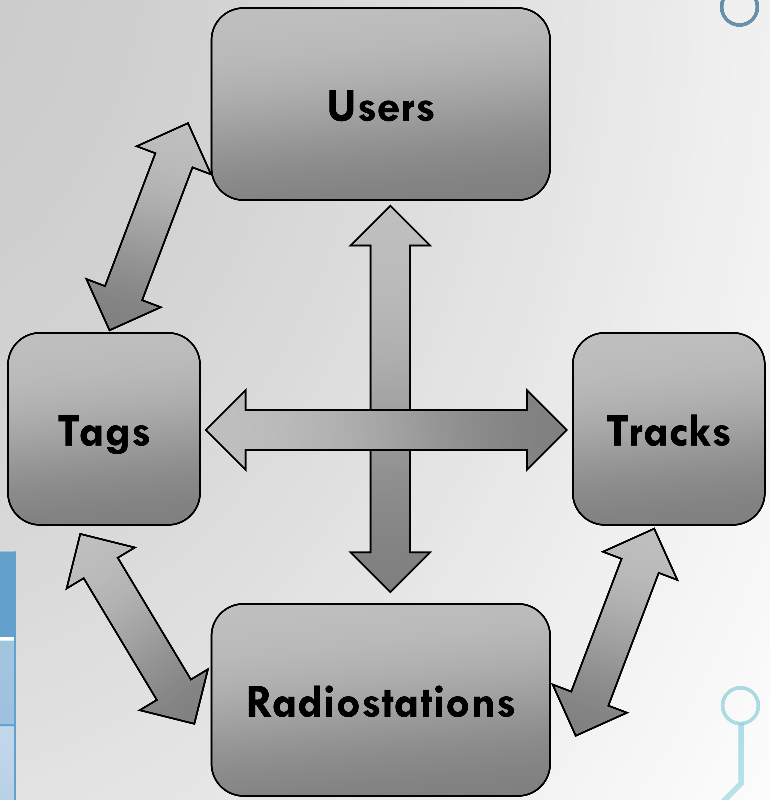
PROBLEM SETTING

- To propose models and algorithms for radiostation (and music) recommendation
- To implement the proposed algorithms, test and compare them on real data of radio hosting FMHost

FMHOST DATA

Entity	Count
User	4266
Tag	3618
Radiostations	2209
Tracks	4165

Relation	Count
User-tag	38504
Radiostation-tag	18539
User-Radiostation	24803
Track-tag	18781
Radiostation-track	22525



THE MODEL: DATA

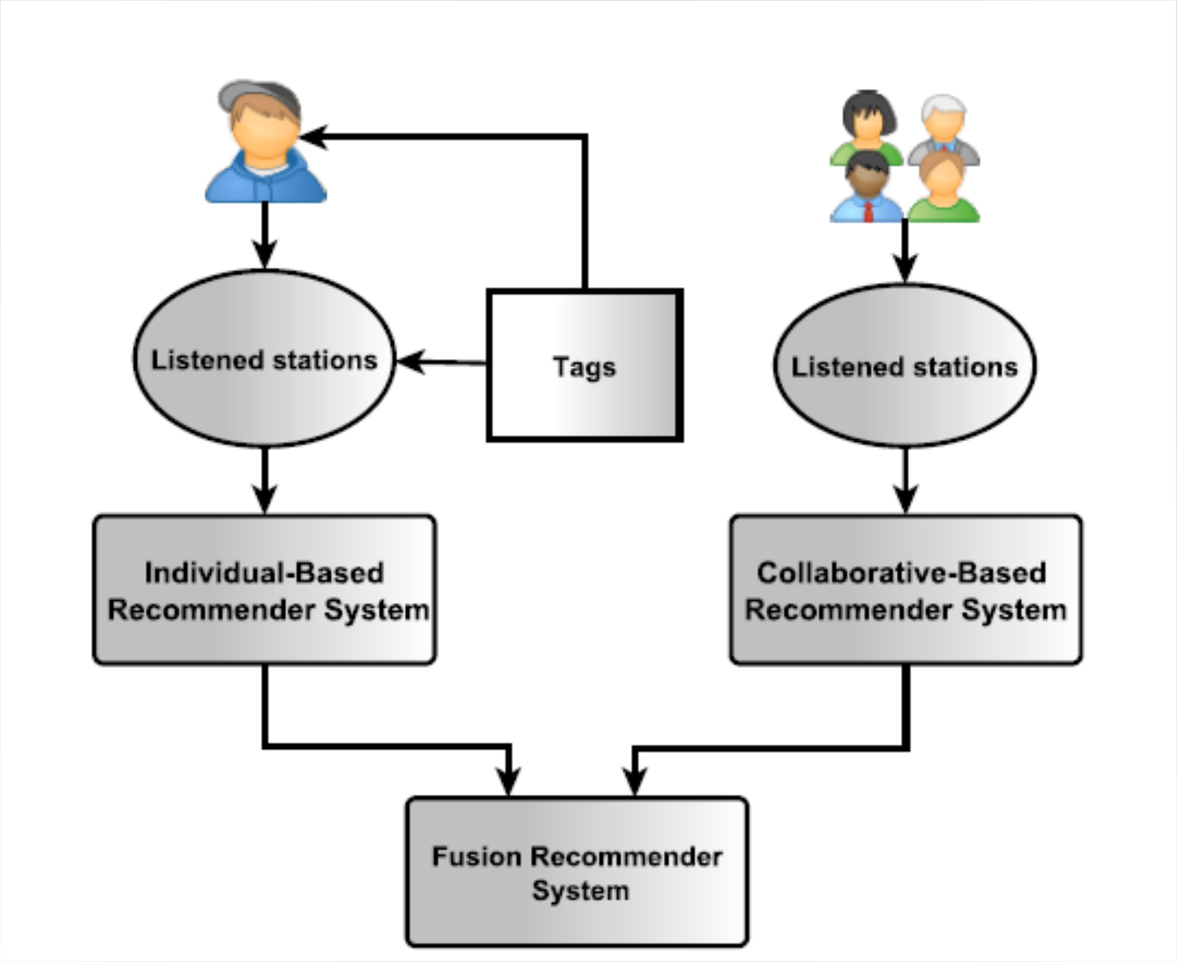
- U is a set of users, R is a set of radio stations, T is a set of tags
- $A=(a_{ut})$, $B=(b_{rt})$, $C=(c_{ur})$, $X=(x_{st})$
- frequency vectors

$$v^A = \sum_{t \in T} a_{ut},\ v^B = \sum_{t \in T} b_{rt},\ \text{and}\ v^C = \sum_{r \in R} a_{ur}$$

- Normalized matrices, e.g.

$$A_f = (a_{ut} \cdot (v_u^A)^{-1})$$

THE MODEL: ARCHITECTURE



METHODS: INDIVIDUAL-BASED RECOMMENDER SYSTEM (IBRS)

Distance between user and radiostation:

$$d(u_0, r) = \sum_{t \in T} |a_{u_0 t} - b_{rt}|$$

Relevance of radiostation r_i for user u_0:

$$score(r_i) = 1 - d(u_0, r_i) / \max_{r_j \in R} d(u_0, r_j)$$

METHODS: COLLABORATIVE-BASED RECOMMENDER SYSTEM (CBRS)

Users' similarity: $\mathrm{sim}(u_0, u) = \frac{\sum_{t \in T} u_{0t} u_t}{\sqrt{\sum_{t \in T} u_{0t}^2 \cdot \sum_{t \in T} u_t^2}}$

$$sim(u_0, u) = 1 - d_{u_0 u} / \max_{u' \in U} d_{u_0 u'}$$

Relevance of radiostation for the target user u_0:

$$score(r) = sim(u^*) \cdot c_{fu^*r}$$

$$u^* = argmax_{u \in U_{u_0}, r \in R / L(u_0)} sim(u) \cdot c_{fur}$$

c_{fur} – frequency of visits of radiostation r by user u

$L(u_0)$ – the set of radiostations listened by user u_0

U_{u_0} – the set of *k* most similar users with the target user u_0

METHODS: FUSION RECOMMENDER SYSTEM (FRS)

For each recommendation list of size *n*:

$$\beta^* \cdot score^C(r) + (1 - \beta^*) \cdot score^I(r)$$

we maximize β by a chosen quality measure:

$$\beta^* = argmax_{\beta \in [0,1]} F\text{−}measure$$

$$\beta^* = argmax_{\beta \in [0,1]} NDCG$$

METHODS: SVD-BASED RECOMMENDER

- $\log(c_{ur} + 1) \sim \mu + b_u + b_r + v_{u.}{}^T v_{r.}$

c_{ur} is the number of times user u listened to radio station r

μ is the general mean, b_u and b_r are the baseline predictors for the user u and the station r

$v_{u.}$ and $v_{r.}$ are the vectors of the user and station features

METHODS: MUSIC RECOMMENDATION TO USERS & REPERTOIRE RECOMMENDATION FOR RADIOSTATIONS

Similarity of users and songs:

$$sim(u_0, s) = \frac{\sum_{t \in T} u_{0t} s_t}{\sqrt{\sum_{t \in T} u_{0t}^2 * \sum_{t \in T} s_t^2}}$$

Relevance of song s_i for user u_0 via distance:

$$score(s_i) = 1 - d(u_0, s_i) / \max_{s_j \in S} d(u_0, s_j)$$

Similarity radiostations and songs:

$$sim(r_0, s) = \frac{\sum_{t \in T} r_{0t} s_t}{\sqrt{\sum_{t \in T} r_{0t}^2 * \sum_{t \in T} s_t^2}}$$

Relevance of song s_i for radiostation r_0 via distance:

$$score(s_i) = 1 - d(r_0, s_i) / \max_{s_j \in S} d(r_0, s_j)$$

QoS: DISTRIBUTION ANALSYSIS

- Looking for Power Law $P(x)=Cx^{-\alpha}$

Table 2. Basic parameters of the user and radio visits datasets, along with their power-law fits and the corresponding *p-value* .

Dataset	n	$\langle x \rangle$	σ	x_{max}	$\hat{x}_{min}$	$\hat{\alpha}$	n_{tail}	*p-value*
User dataset	4187	5.86	12.9	191	12 ± 2	2.46(0.096)	117	**0.099**
Radio dataset	2209	11.22	60.05	1817	46 ± 11	2.37(0.22)	849	**0.629**

QoS: DISTRIBUTION ANALYSIS

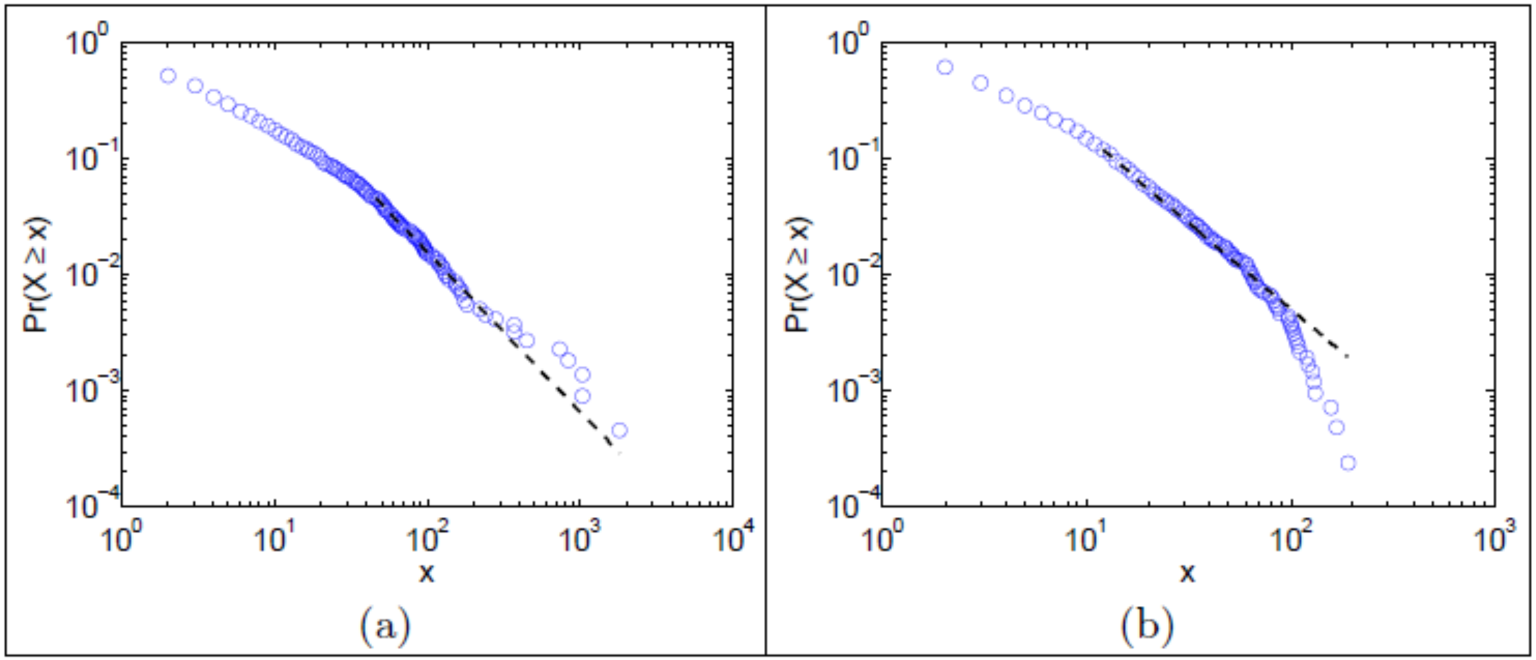


Fig. 2. Cumulative distribution functions $P(x)$ and their maximum likelihood power-law fits for the FMhost two empirical data sets. (a) The frequency distribution of radio station visits. (b) The frequency of visits of unique users.

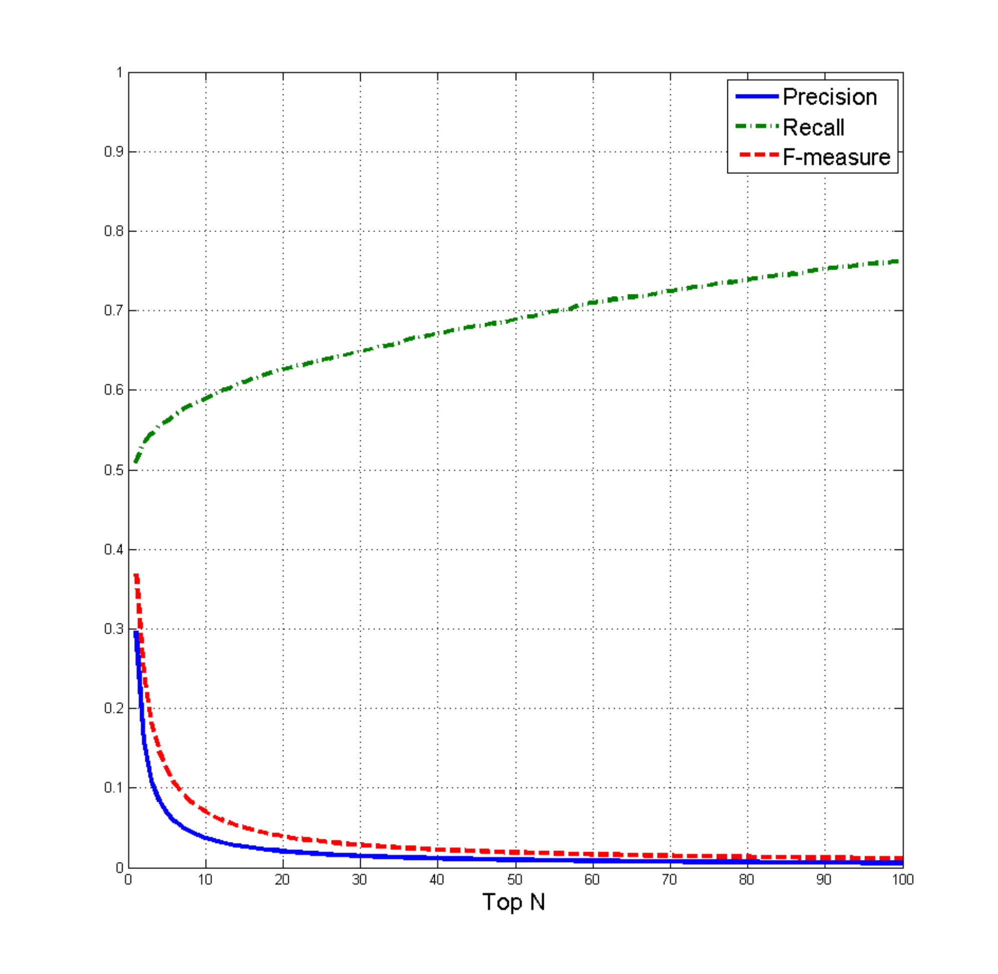
QoS: DISTRIBUTION ANALYSIS

- Pareto Principle (20%:80%)

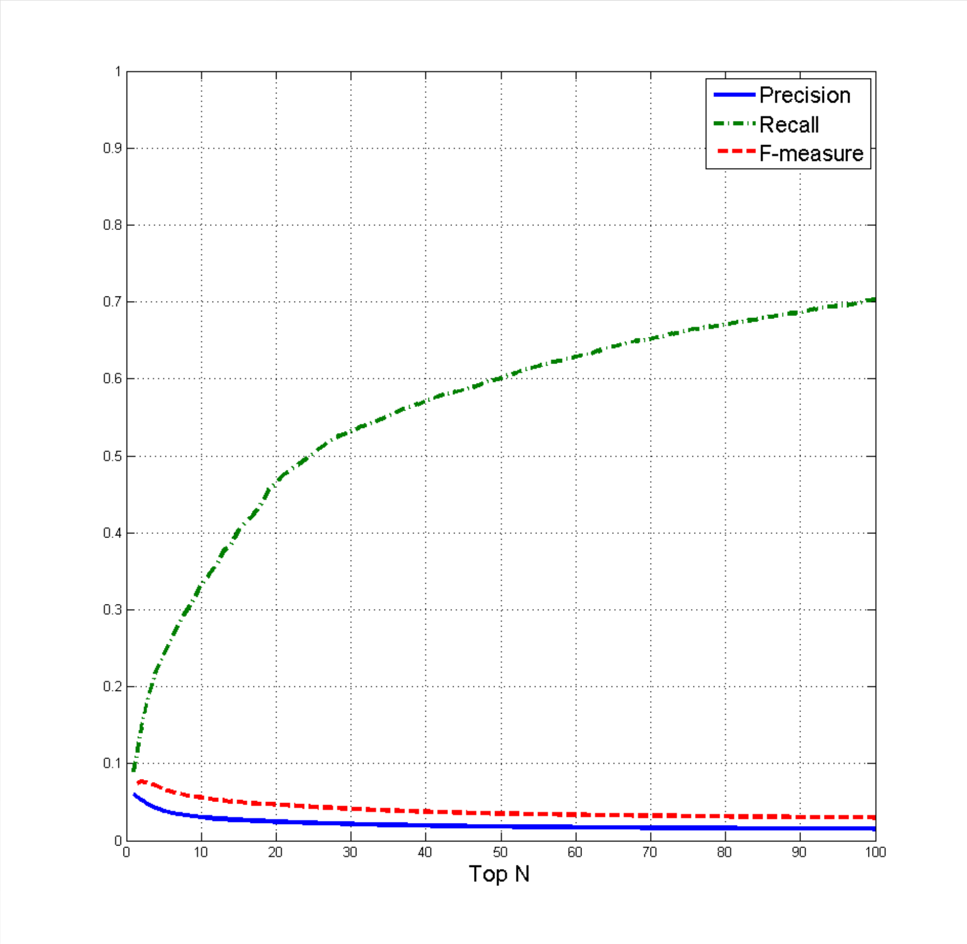
$$W = P^{(\alpha-2)/(\alpha-1)}$$

- 50%:80% for radio stations
- 50%:83% for user visits

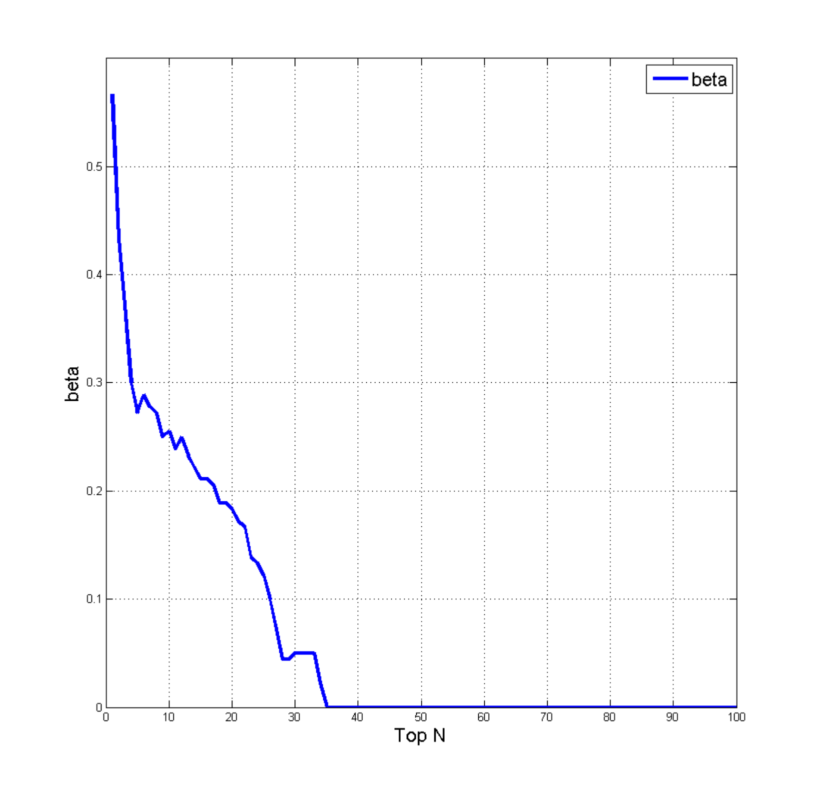
RESULTS: IBRS

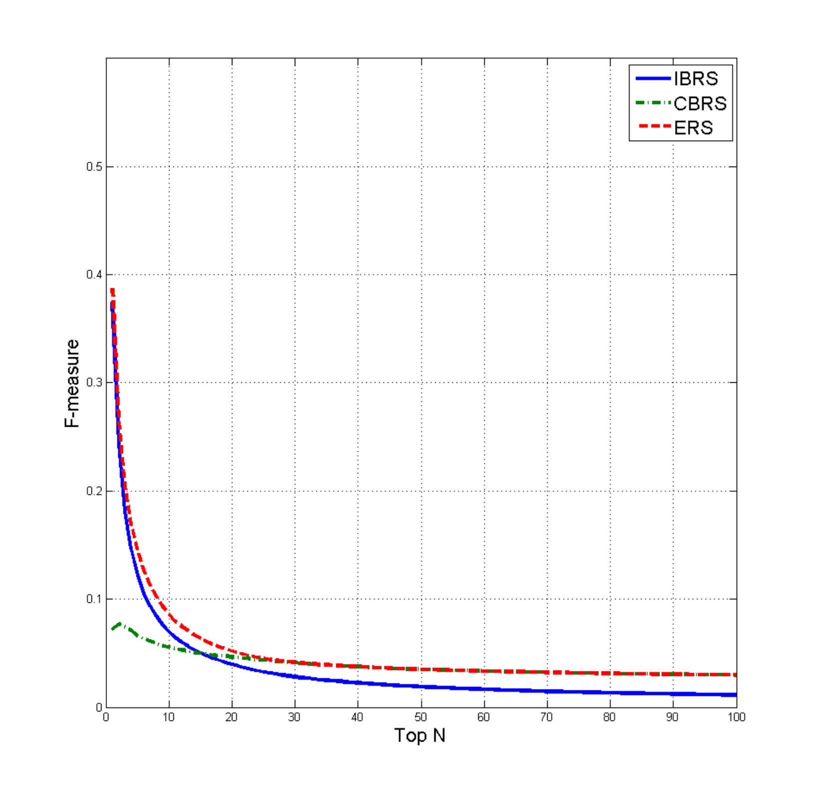


RESULTS: CBRS



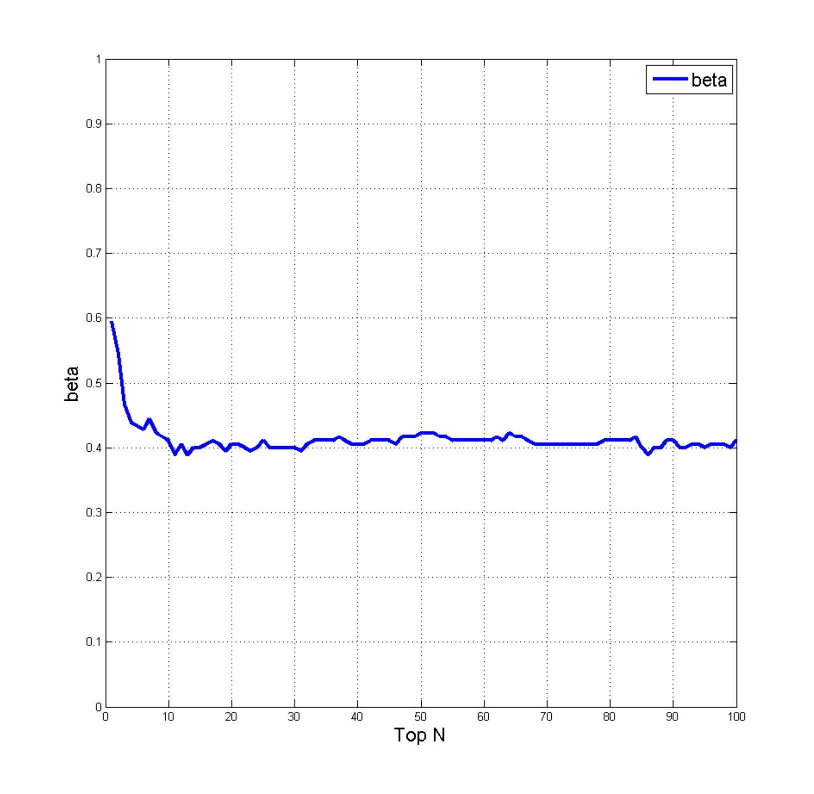
RESULTS: FRS
MAXIMIZATION OF F-MEASURE

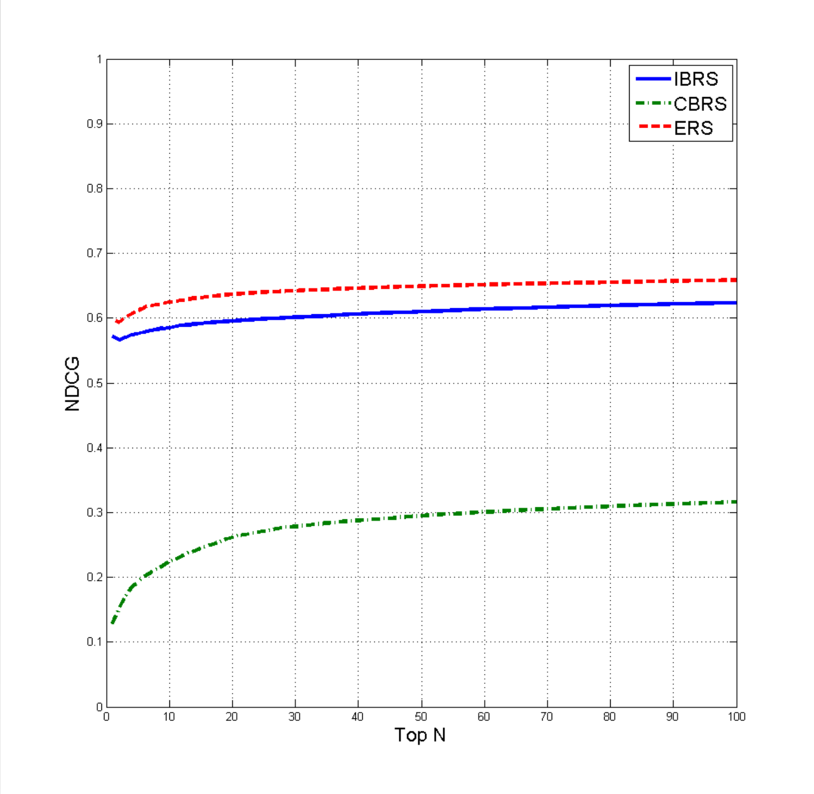




$$\beta \cdot score^{CBRS}(r) + (1-\beta) \cdot score^{IBRS}(r)$$

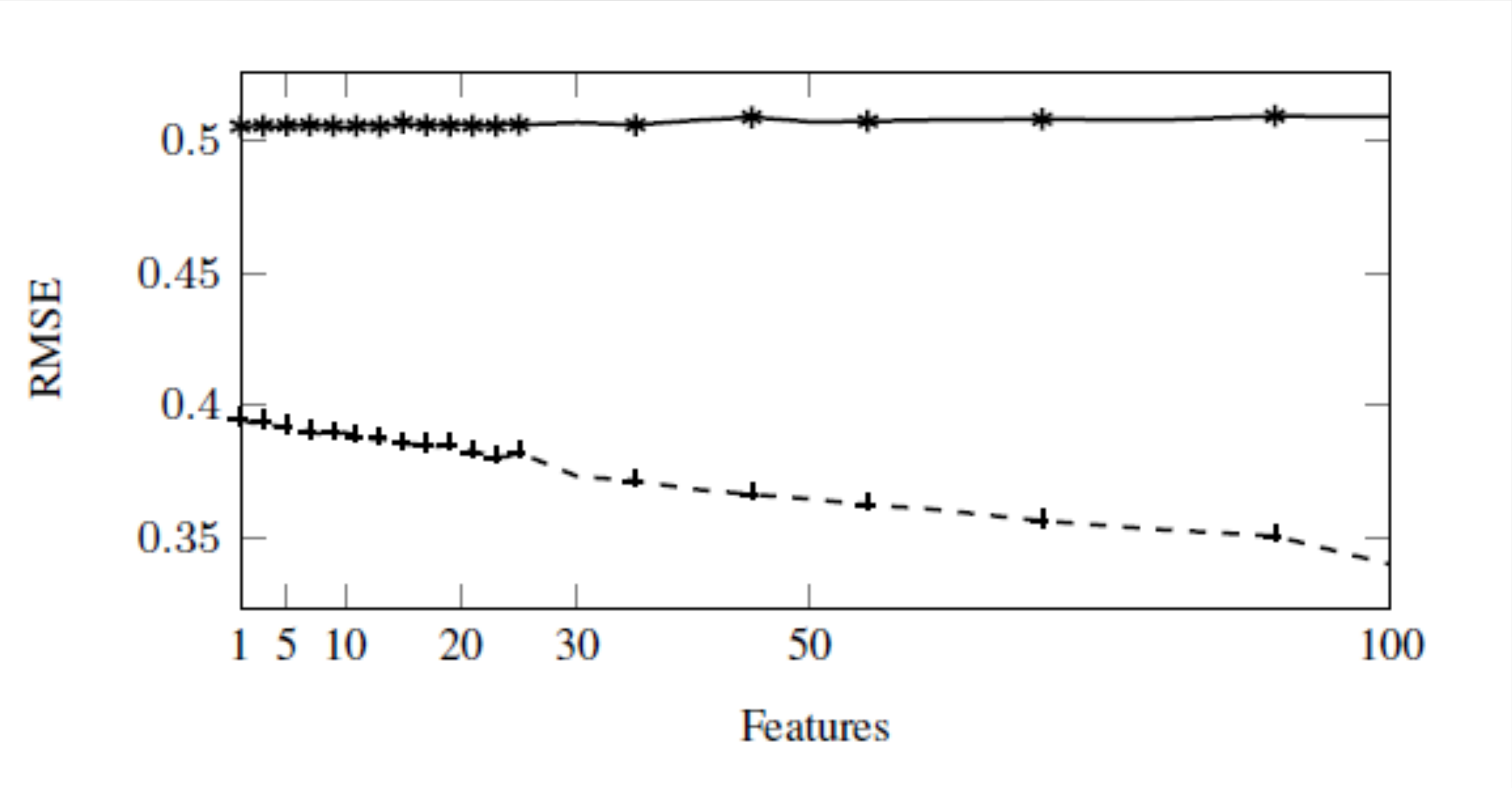
RESULTS: FRS MAXIMIZATION OF NDCG





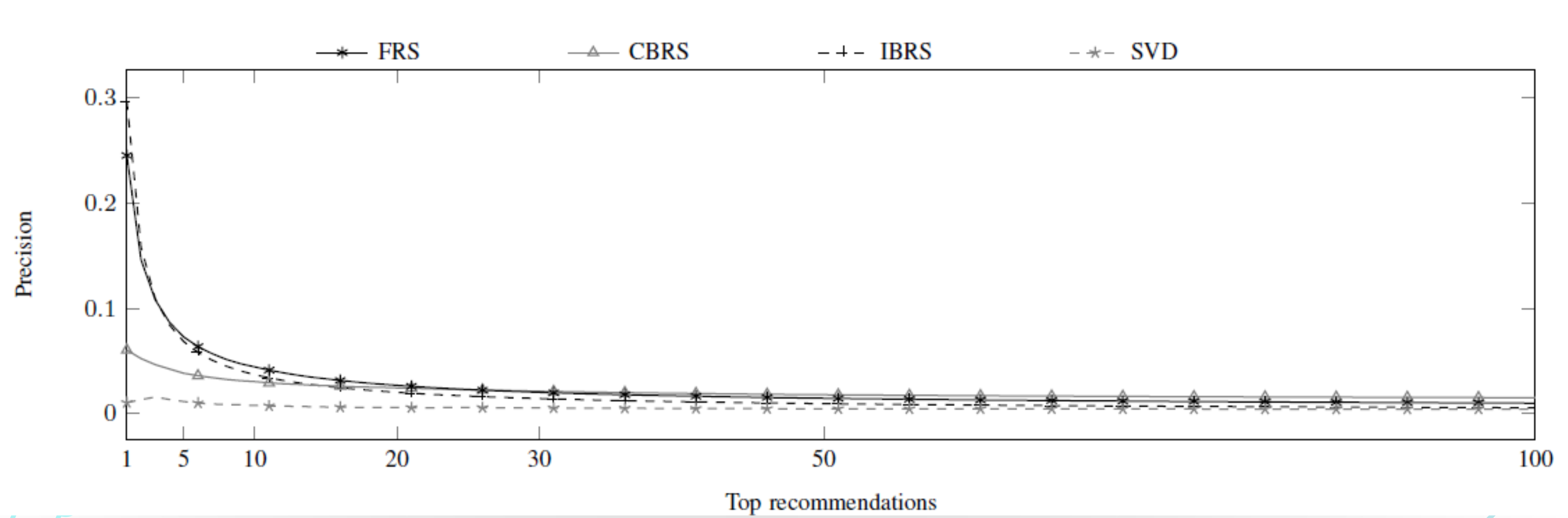
$$\beta \cdot score^{CBRS}(r) + (1-\beta) \cdot score^{IBRS}(r)$$

RESULTS: SVD

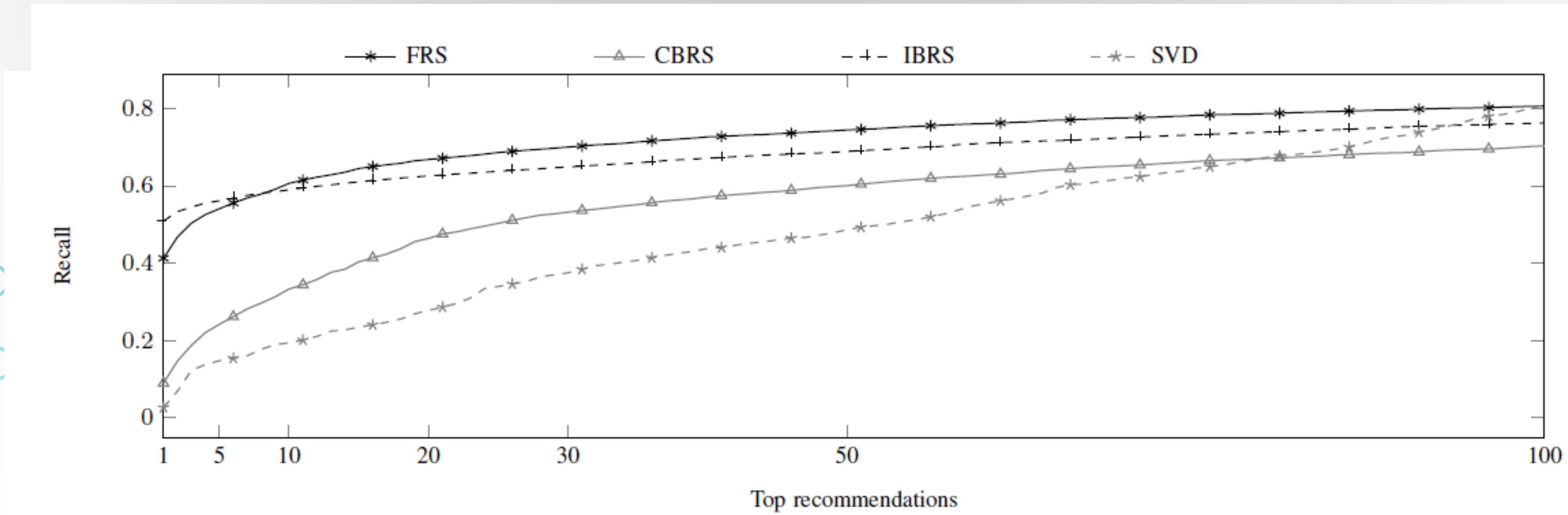


Solid line denotes error on the validation set; dashed line, error on the training set.

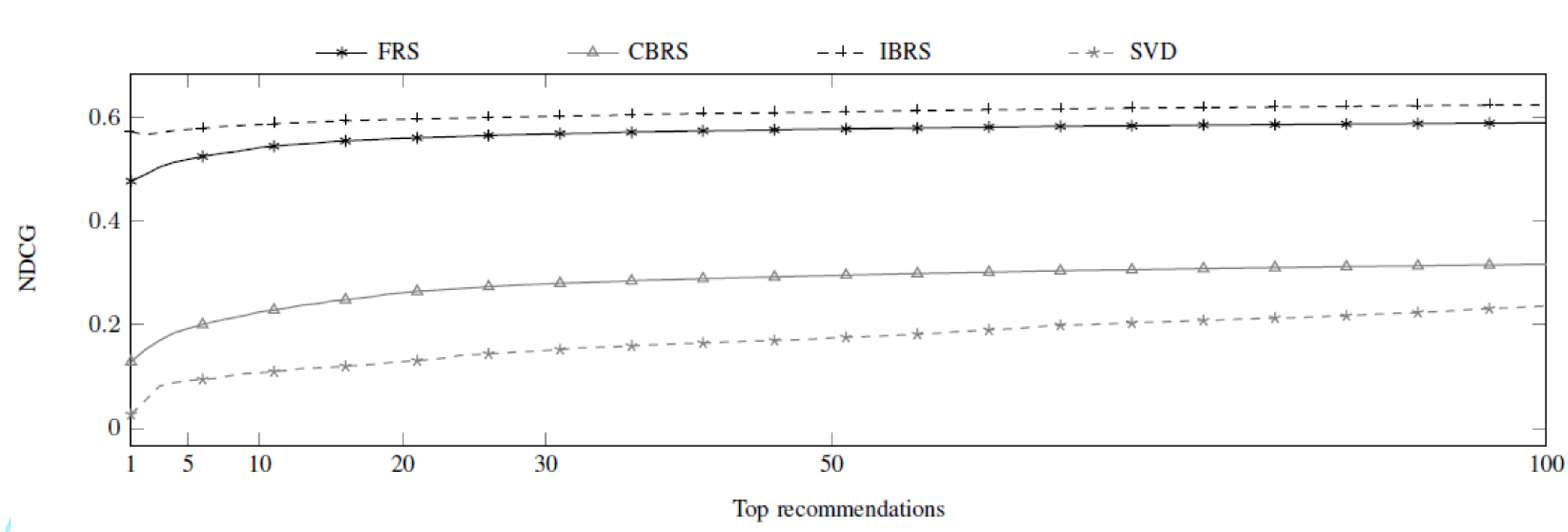
RESULTS: COMPARISON



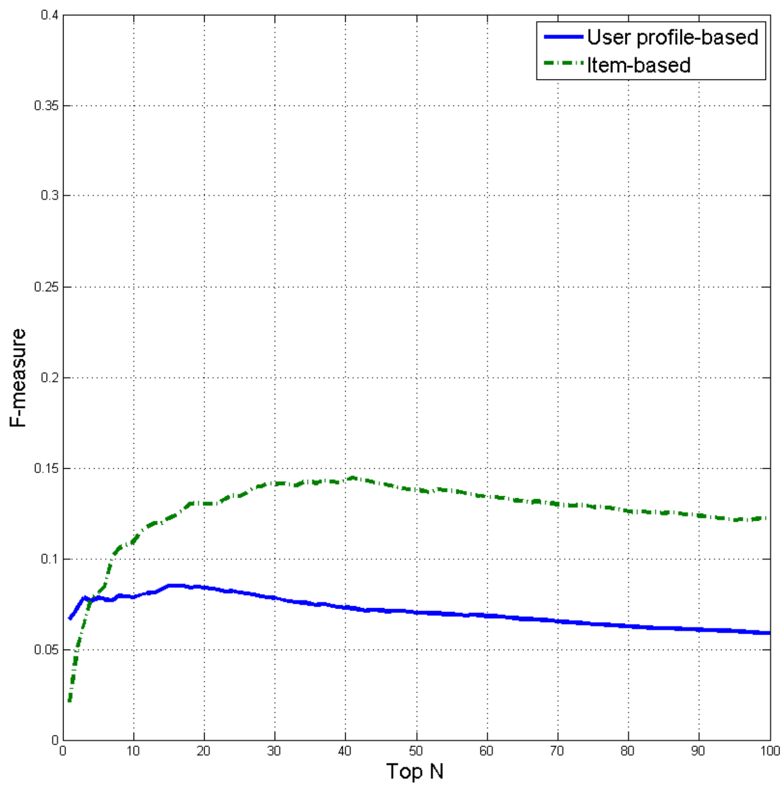
RESULTS: COMPARISON

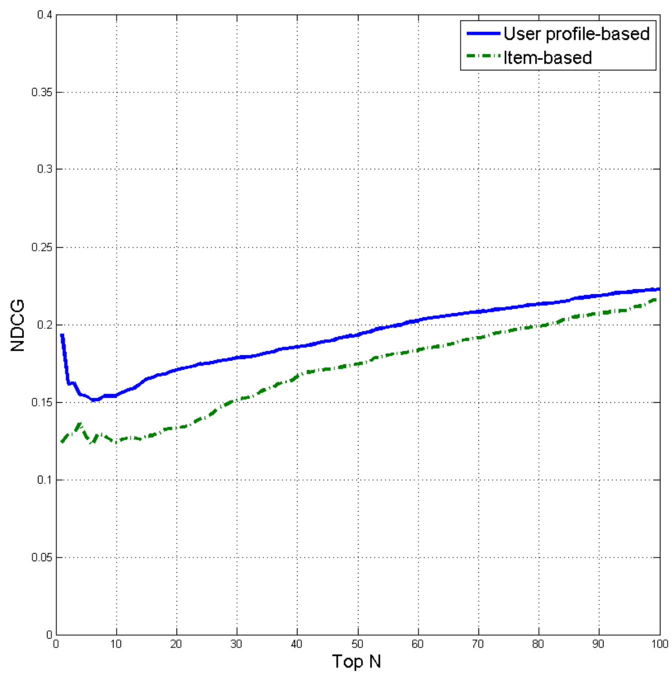


RESULTS: COMPARISON

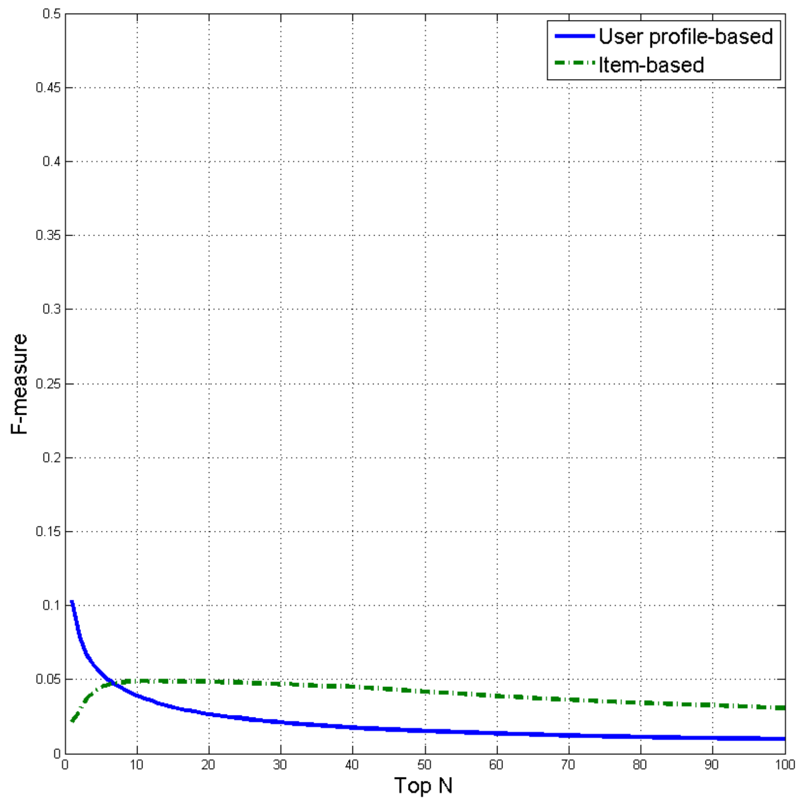


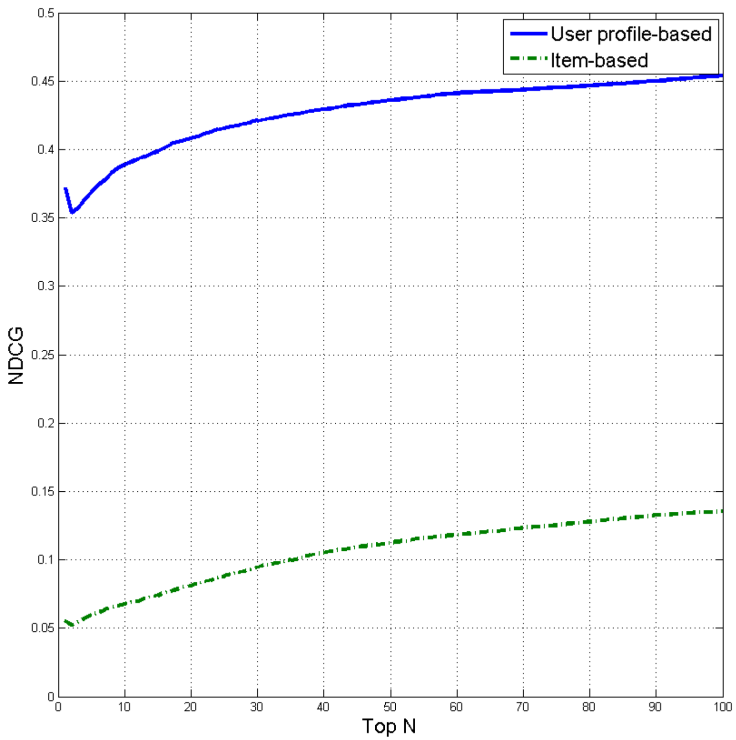
RESULTS: MUSIC RECOMMENDATION TO USERS





RESULTS: RECOMMENDATION OF REPERTOIRE FOR RADIOSTATIONS





CONCLUSION

- We have described the underlying models, algorithms, and system architecture of the new improved FMHost service and tested it on the available real dataset.
- By using bimodal cross-validation, we have built a hybrid algorithm FRS tuned to maximize either F-measure or NDCG for various values of N and β. The FRS algorithm performs better than the three other approaches, namely IBRS, CBRS, and SVD, both in terms of F-measure and in terms of NDCG.
- According to the NDCG@n measure, IBRS is strictly better than CBRS, so the former one is a better ranker.
- Surprisingly, in our experiments the state-of-the-art SVD-based technique performed poorly in comparison to our proposed algorithms. This can be explained by the small size and sparseness of our dataset.

Thank you!

Question?