Scrutinizing Mobile App Recommendation: Identifying Important App-Related Indicators

Jovian Lin,
Kazunari Sugiyama,
Min-Yen Kan,
and Tat-Seng Chua

National University of Singapore

Introduction

Basic Recommendation Systems

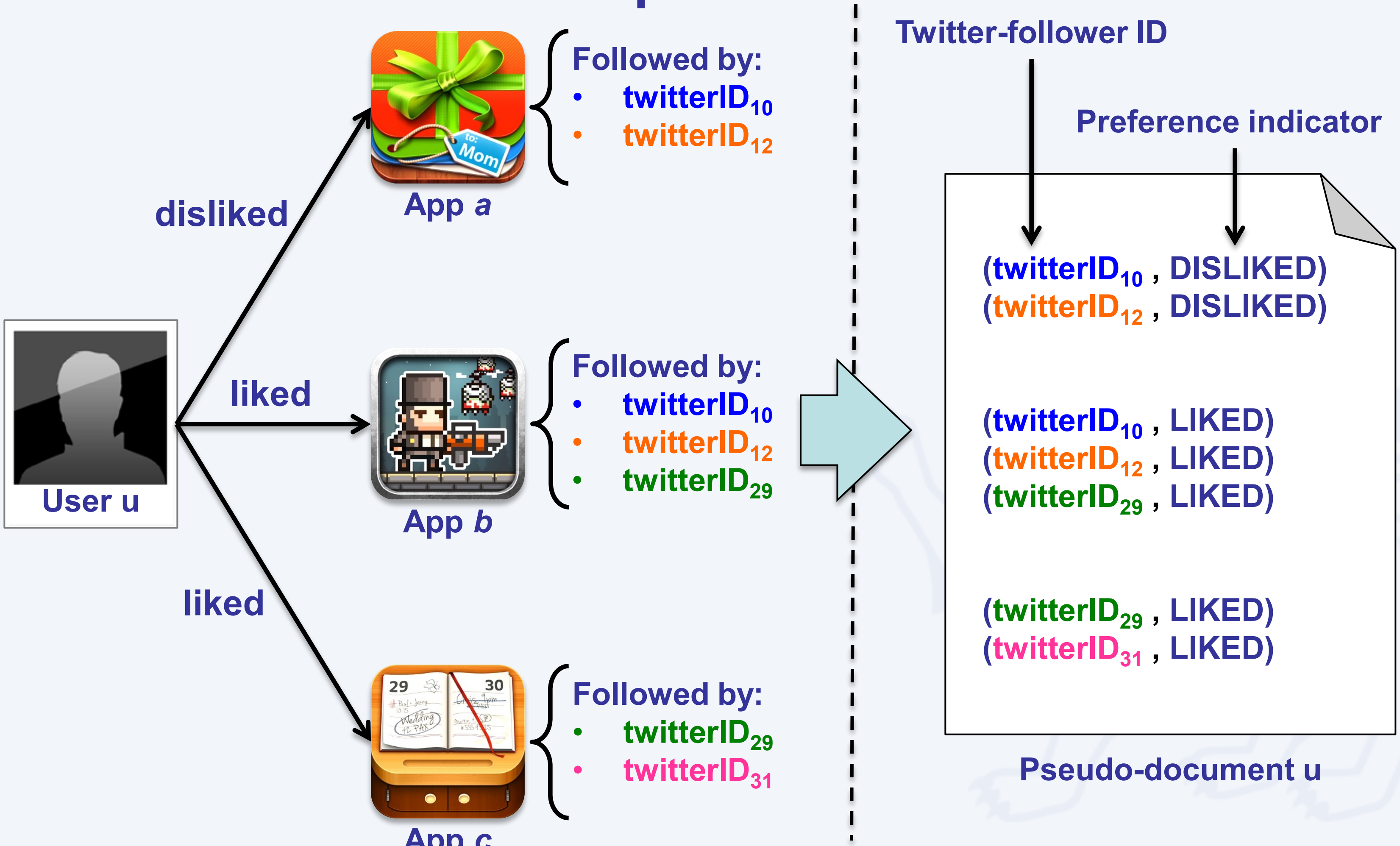
- **Collaborative Filtering (CF)**
  - User's ratings
- **Content-based Filtering (CBF)**
  - Contents of items

We have proposed mobile app recommendation systems:

- **Twitter followers (TWF)** [Lin et al., SIGIR'13]
- **Version sensitive recommendation (VSR)** [Lin et al. SIGIR'14]

Twitter Followers (TWF) [Lin et al., SIGIR'13]

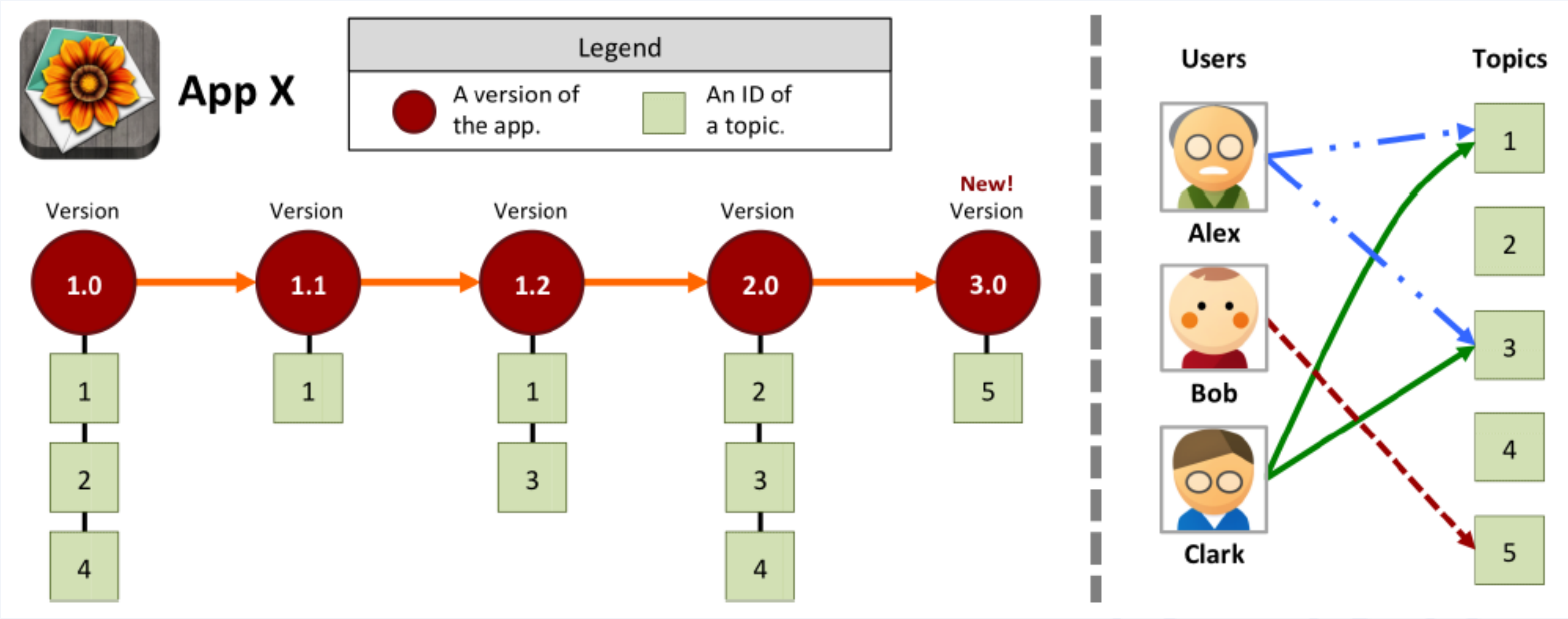
Pseudo-documents and pseudo-words



Version Sensitive Recommendation (VSR)

[Lin et al., SIGIR'14]

Relationship between version of apps and users



Introduction

How about unifying the followings?

- Collaborative filtering (CF)
- Content-based filtering (CBF)
- Twitter followers (TWF)
- Version sensitive recommendation (VSR)

To achieve this,

1. Unify the strengths of the four recommender techniques.

2. Propose a set of specific features (in the app domain) for the unifying framework.

3. Perform in-depth analysis of these features and uncover interesting connections.

Methodology

Feature set

1. The app's marketing-related metadata (M)
2. The user's history-related information (H)
3. The recommendation scores of different recommender systems (R)

Each app's feature vector $X_{u,a}$ is composed of the above three types of information.

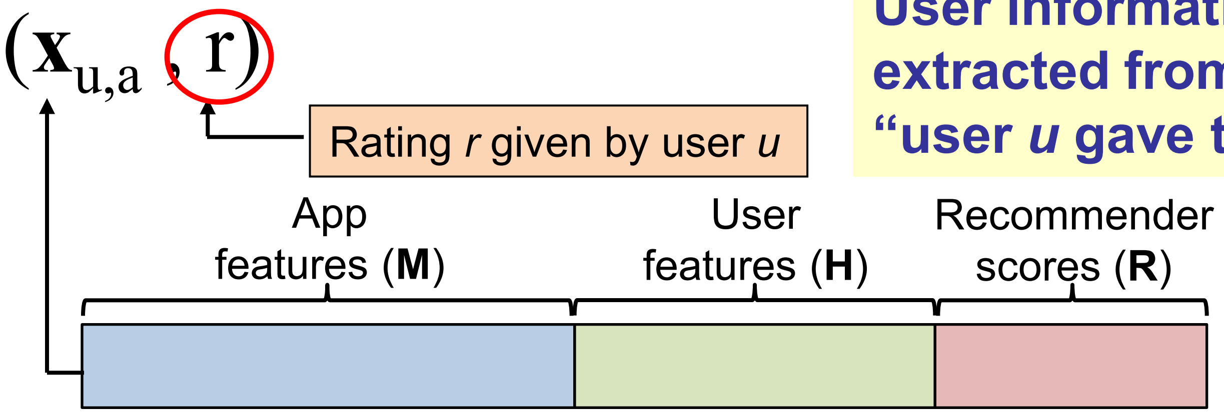
$$X_{u,a} = \{X^{M}_{a}, X^{H}_{ua}, X^{R}_{ua}\}$$

App's metadata

User's history

Recommendation scores

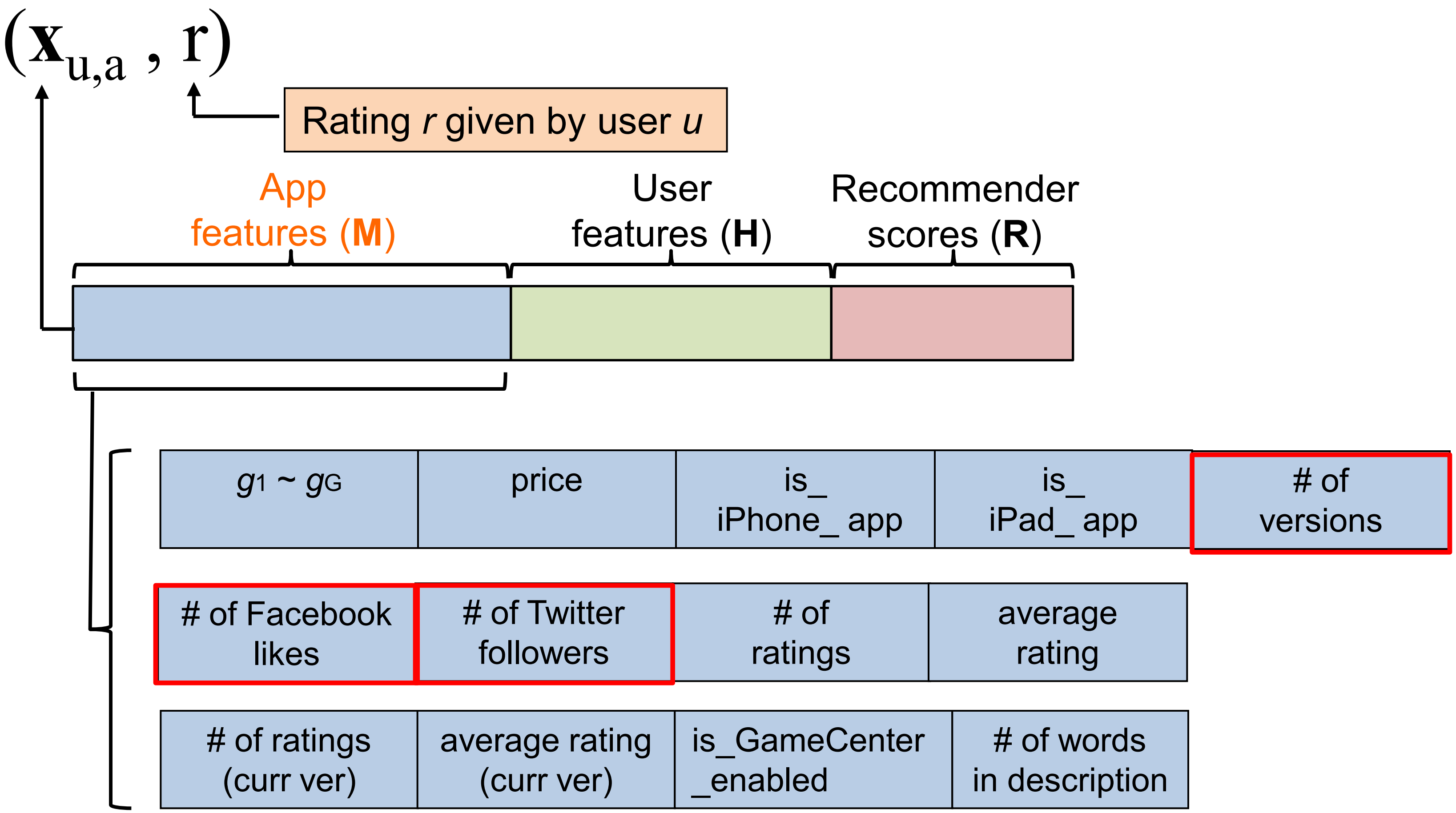
Feature set



User information is extracted from ratings. Namely, "user *u* gave the rating *r* to app *a*.

Feature set

App’s Marketing-Related Metadata (M)

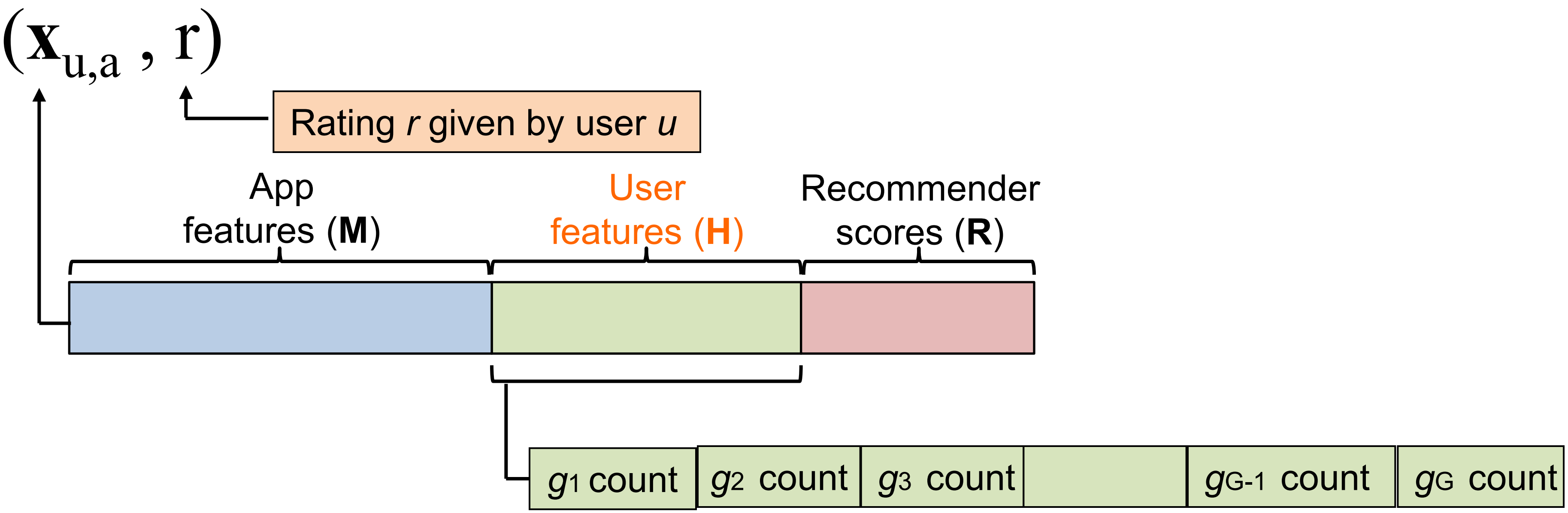


Enrich the app features by including

- **# of versions,**
- **# of Facebook likes,**
- **# of Twitter followers**

Feature set

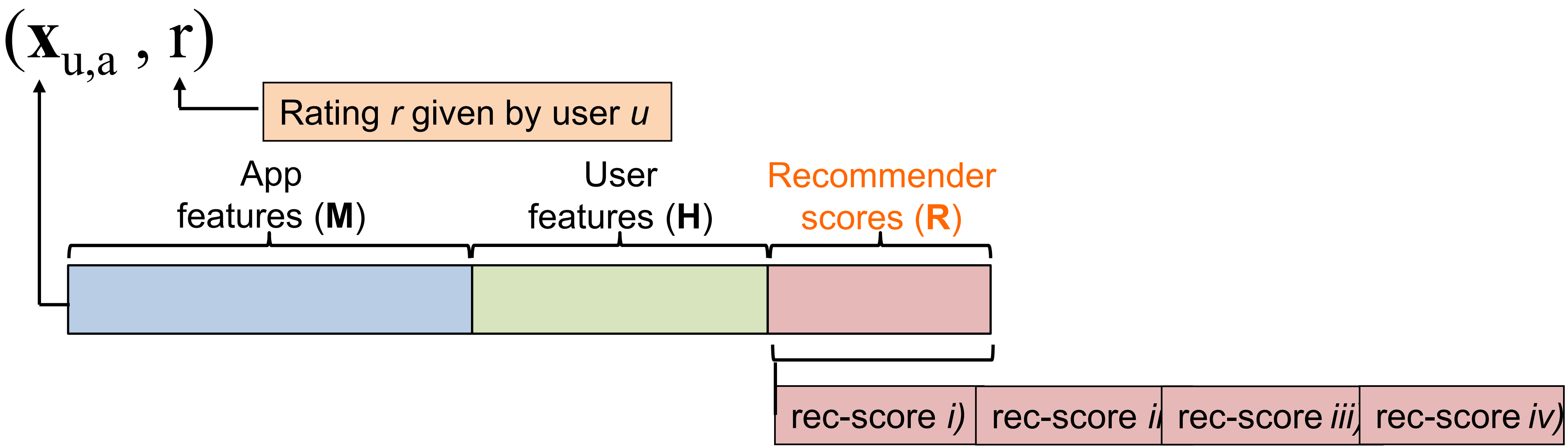
User's History-Related Information (H)



The number of times that apps in genre *g* were consumed by user *u*.

Feature set

Recommendation Scores from Different Recommender Techniques (R)

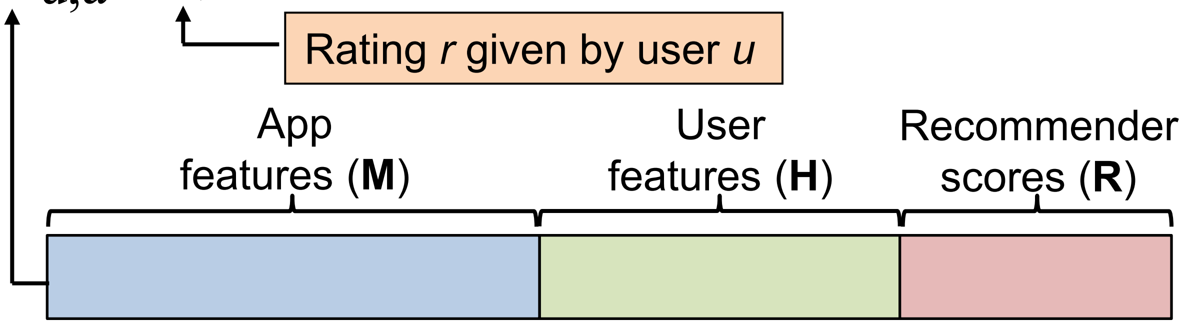


Include the recommendation scores from the individual recommendation algorithms:

i) Collaborative filtering (CF)
ii) Content-based filtering (CBF)
iii) Twitter followers (TWF)
iv) Version sensitive recommendation (VSR)

Combining App Features

$$(\mathbf{x}_{u,a}\ ,\ r)$$



Employy Gradient Tree Boosting (GTB) to train the model

(via “scikit-learn”).

Training Phase:

- Give
  - feature vector “$X_{u,a}$ “
  - rating “*r*”

  to GTB

- GTB constructs an ensemble of decision tree learners

Test Phase:

- Given a (testing) feature vector (*i.e.*, $X_{u,a}$), the learned model predicts “*r*”

Experiments

Experimental Data

After retaining only unique users who give at least 30 ratings, we obtain the following data:

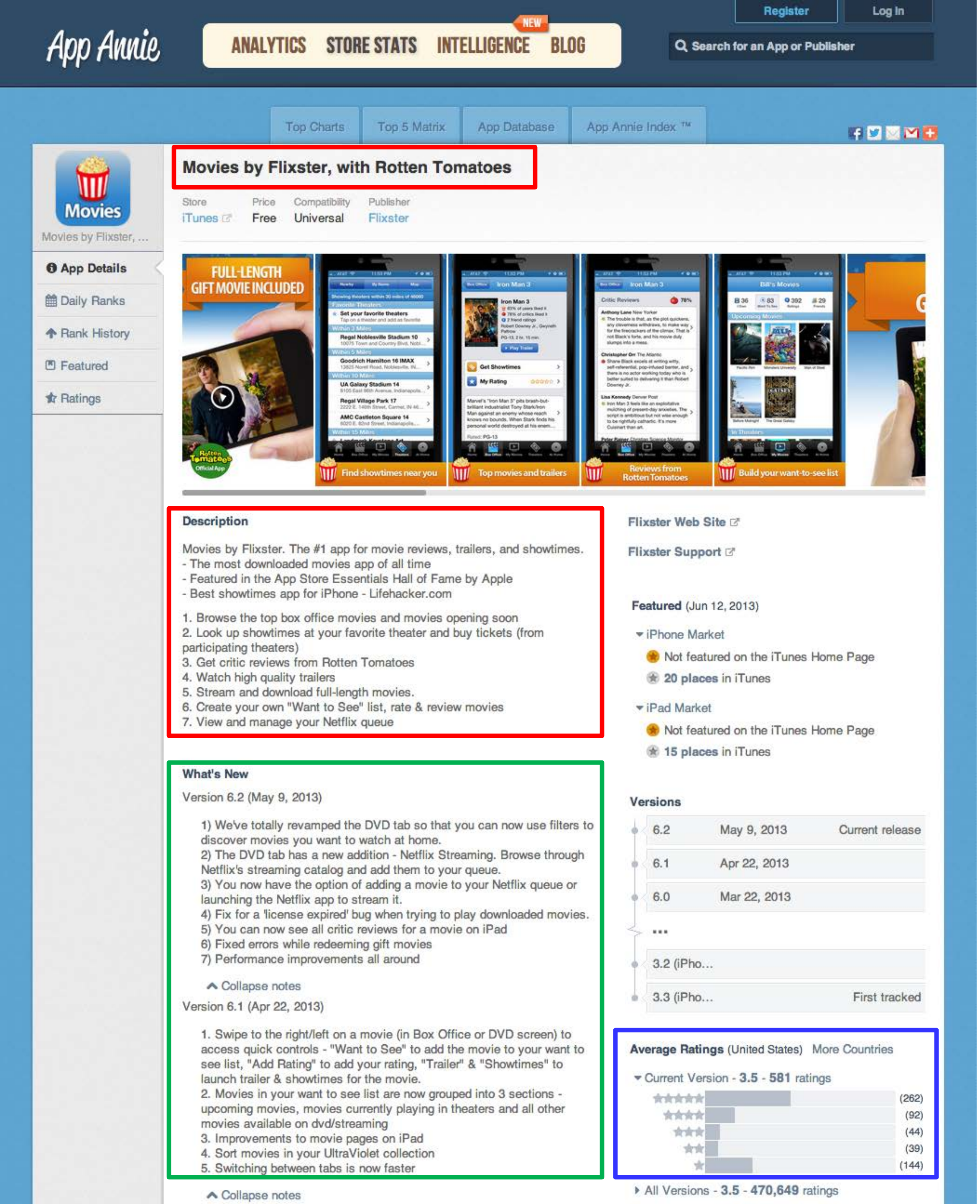
- 33,802 apps
- 16,450 users
- 3,106,759 ratings

Collected from

- iTunes App Store
- App Annie
- Twitter
- Facebook

Among about 33.8K apps,

- 7,124 (21.1%) have Twitter accounts
- 9,288 (27.5%) have Facebook accounts,
- 10,520 (31.1%) have at least 5 versions.
- 678 (2%) have both Twitter and Facebook accounts.



: App metadata

: Version information

: Rating information

Experiments

Comparative Approaches

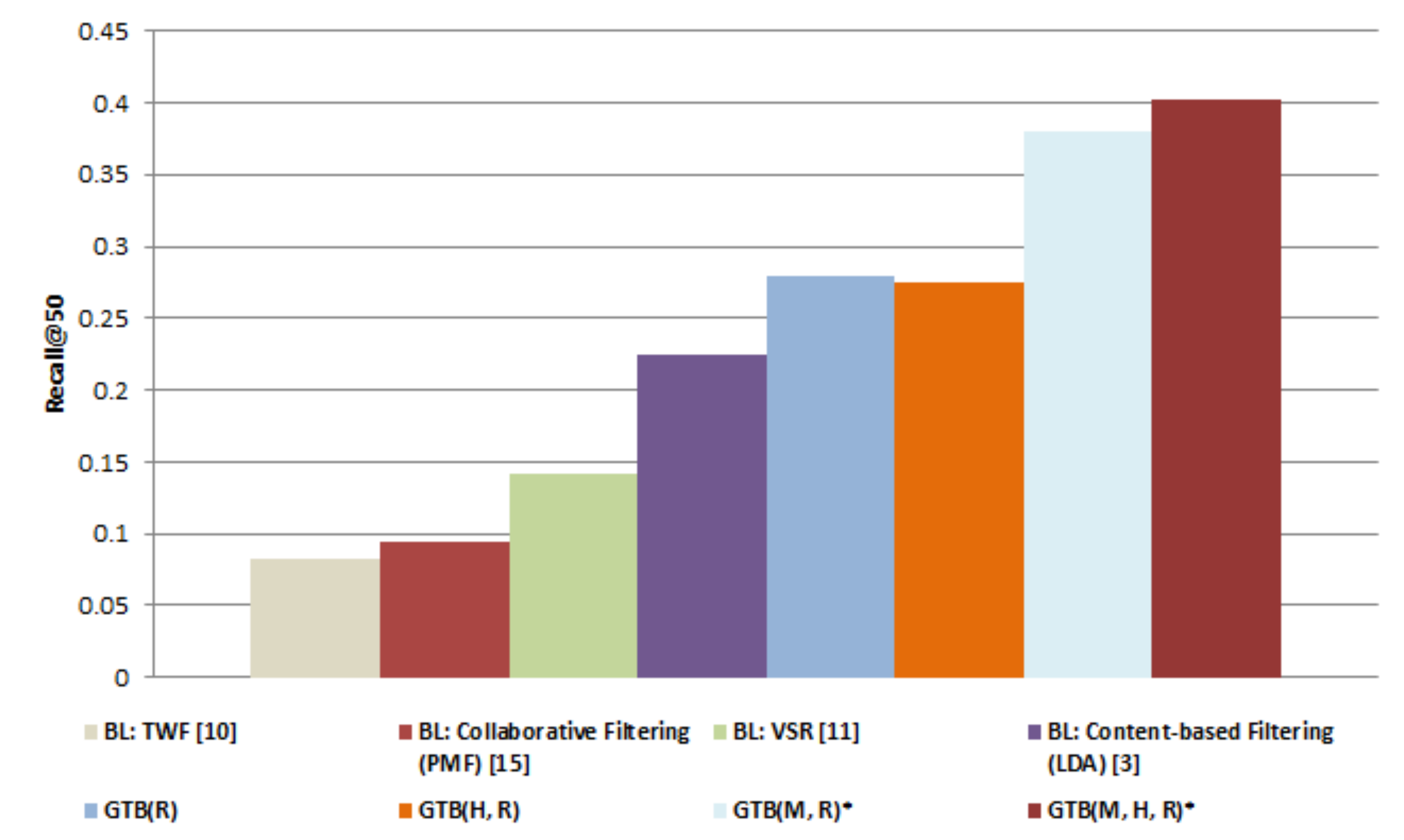
- **Individual recommendation techniques**
  - i) Collaborative filtering (PMF) [Salakhutdinov and Mnih, ICML'08']
  - ii) Content-based filtering (LDA) [Blei et al., JMLR'03]
  - iii) Twitter followers (TWF) [Lin et al., SIGIR'13]
  - iv) Version sensitive recommendation (VSR) [Lin et al., SIGIR'14]

  with $X_{u,a} = \{X^{H}_{ua}\}$
- **Hybrid recommendation techniques**
  - GTB(R)
  - GTB(H, R)
  - GTB(M, R)
  - GTB(M, H, R)

M: App's Marketing-related metadata
H: User's history-related information
R: Recommendation score

- **Evaluation measure: Recall@50**

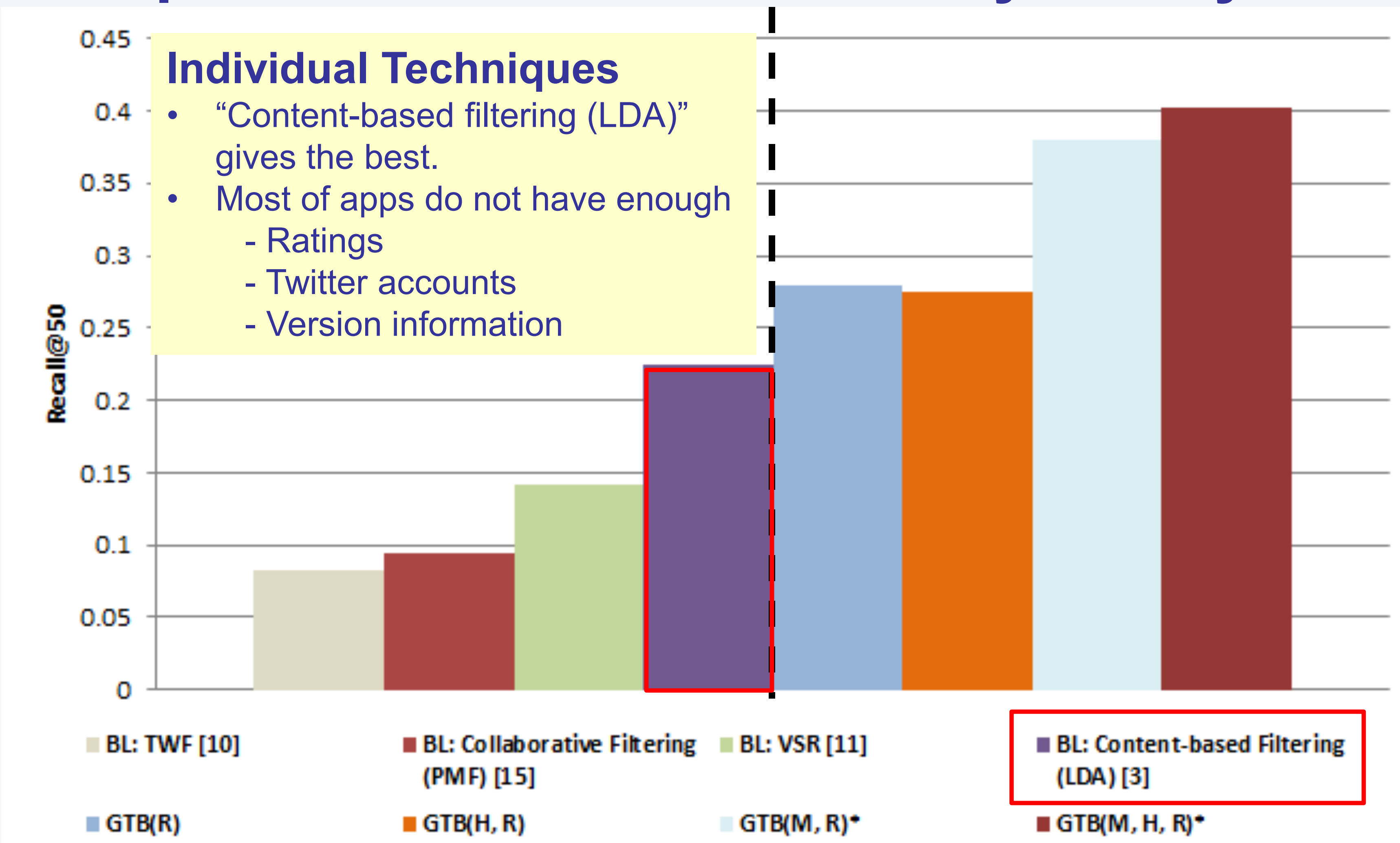
Comparison of Individual and Hybrid Systems



Comparison of Individual and Hybrid Systems

Individual Techniques

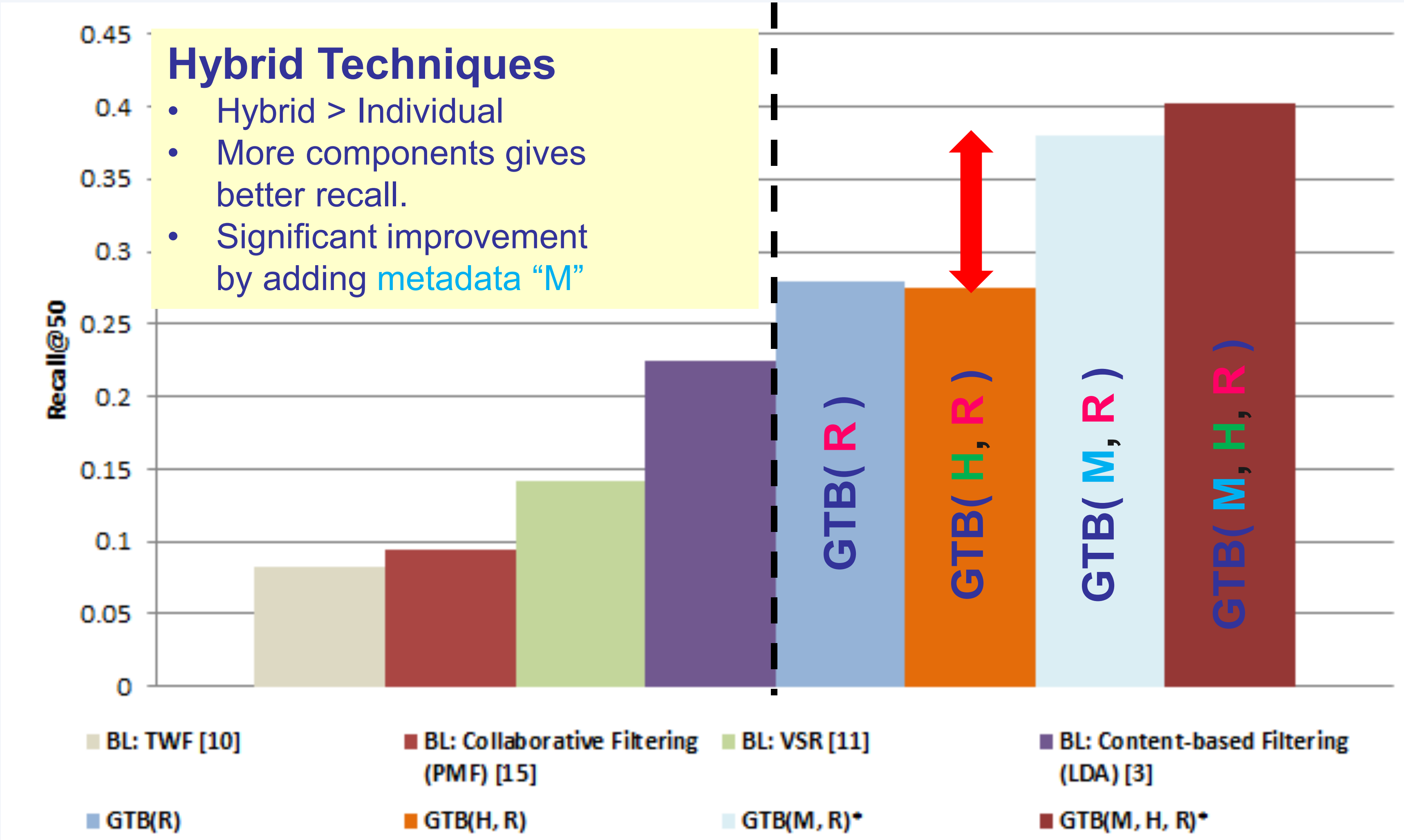
- "Content-based filtering (LDA)" gives the best.
- Most of apps do not have enough
  - Ratings
  - Twitter accounts
  - Version information



Comparison of Individual and Hybrid Systems

Hybrid Techniques

- Hybrid > Individual
- More components gives better recall.
- Significant improvement by adding metadata "M"



Ablation Testing for Hybrid Recommendation

Feature	Recall@50
GTB(M, H, R)	**0.403**
GTB(M, H, R), excluding TWF	0.363
GTB(M, H, R), excluding VSR	0.346
GTB(M, H, R), excluding collaborative filtering (PMF)	0.292
GTB(M, H, R), excluding content-based filtering (LDA)	0.237
TWF	0.082
VSR	0.141
Collaborative filtering (PMF)	0.094
Content-based filtering (LDA)	0.225

Improve by 16.5%

Ablation Testing for Hybrid Recommendation

Feature	Recall@50
GTB(M, H, R)	**0.403**
GTB(M, H, R), excluding TWF	0.363
GTB(M, H, R), excluding VSR	0.346
GTB(M, H, R), excluding collaborative filtering (PMF)	0.292
GTB(M, H, R), excluding content-based filtering (LDA)	0.237
TWF	0.082
VSR	0.141
Collaborative filtering (PMF)	0.094
Content-based filtering (LDA)	0.225

Improve by 11.0%

Ablation Testing Using Sufficient Twitter Information

Feature	Recall@50
$GTB_{TWF}(M, H, R)$	**0.446**
$GTB_{TWF}(M, H, R)$, excluding VSR	0.412
$GTB_{TWF}(M, H, R)$, excluding collaborative filtering (PMF)	0.390
GTB_{TWF} M, H, R), excluding content-based filtering (LDA)	0.386
$GTB_{TWF}(M, H, R)$, excluding TWF	0.338

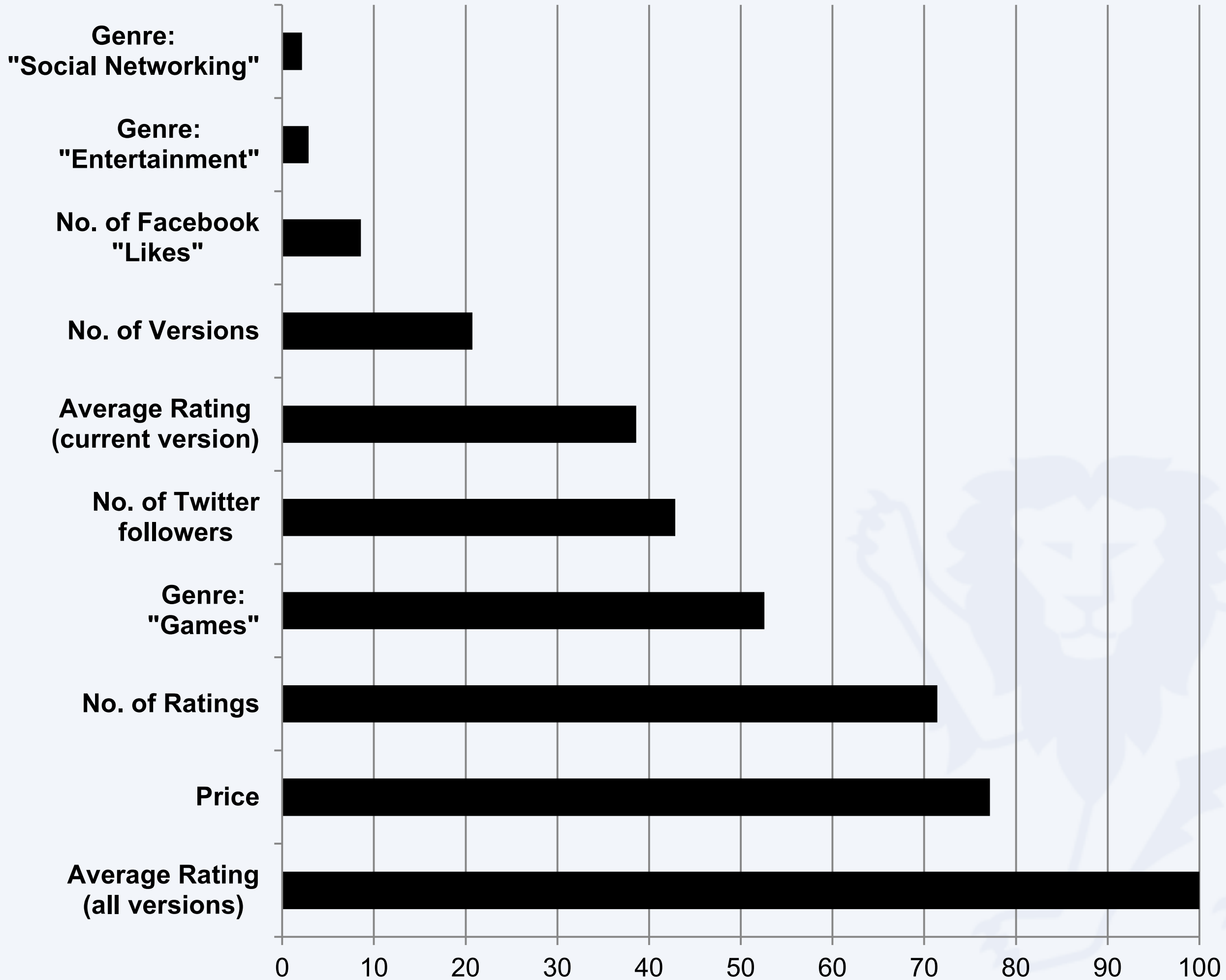
Ablating Twitter information results in largest dip, indicating Twitter followers are important by using full twitter information.

Ablation Testing Using Sufficient Version Information

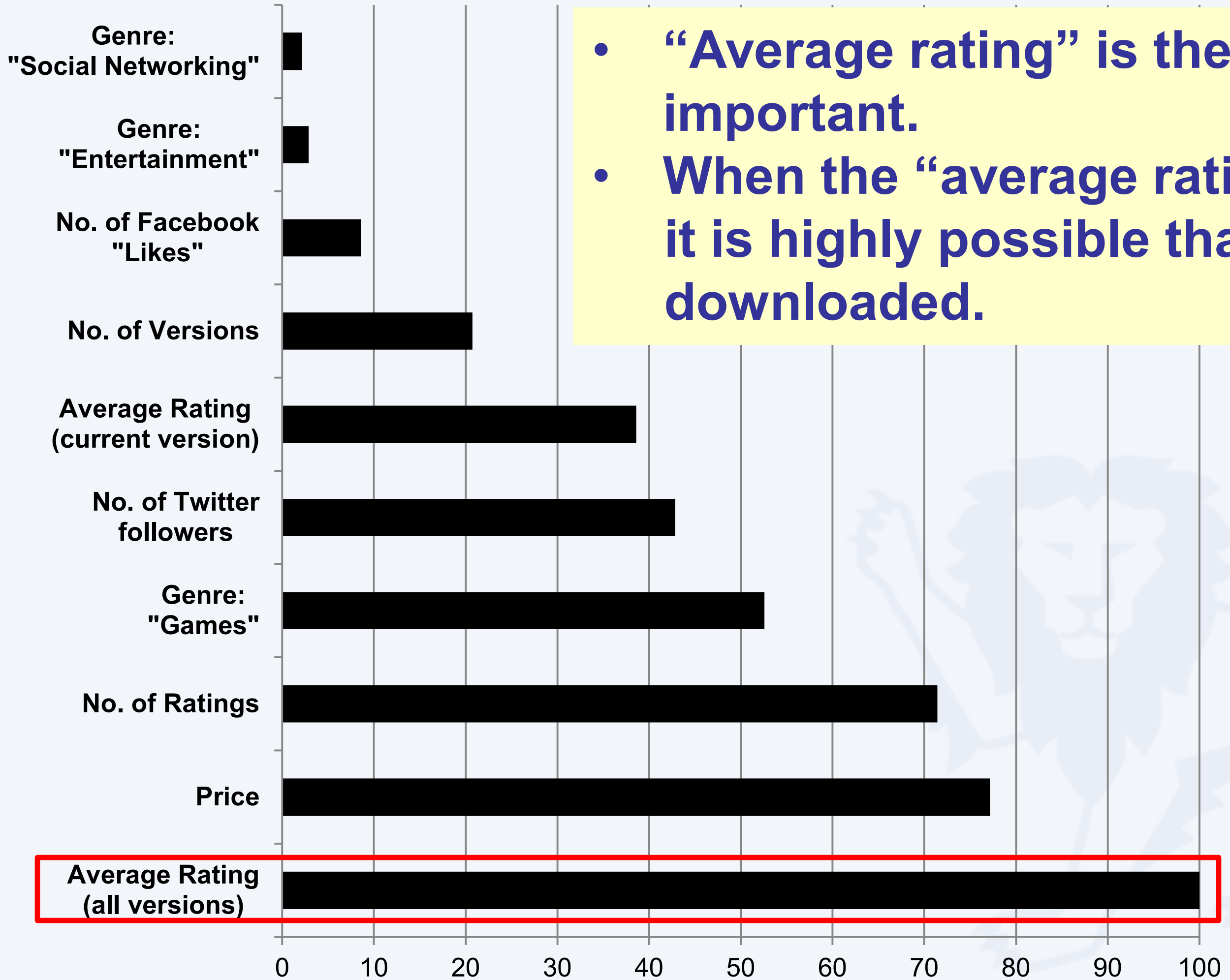
Feature	Recall@50
GTB_{VSR}(M, H, R)	**0.418**
GTB_{VSR}(M, H, R), excluding TWF	0.396
GTB_{VSR}(M, H, R), excluding content-based filtering (PMF)	0.361
GTB_{VSR}(M, H, R), excluding VSR	0.344
GTB_{VSR}(M, H, R), excluding collaborative filtering (PMF)	0.335

Although VSR does not displace CF, it still results in the second largest dip in recall scores.

Feature Importance

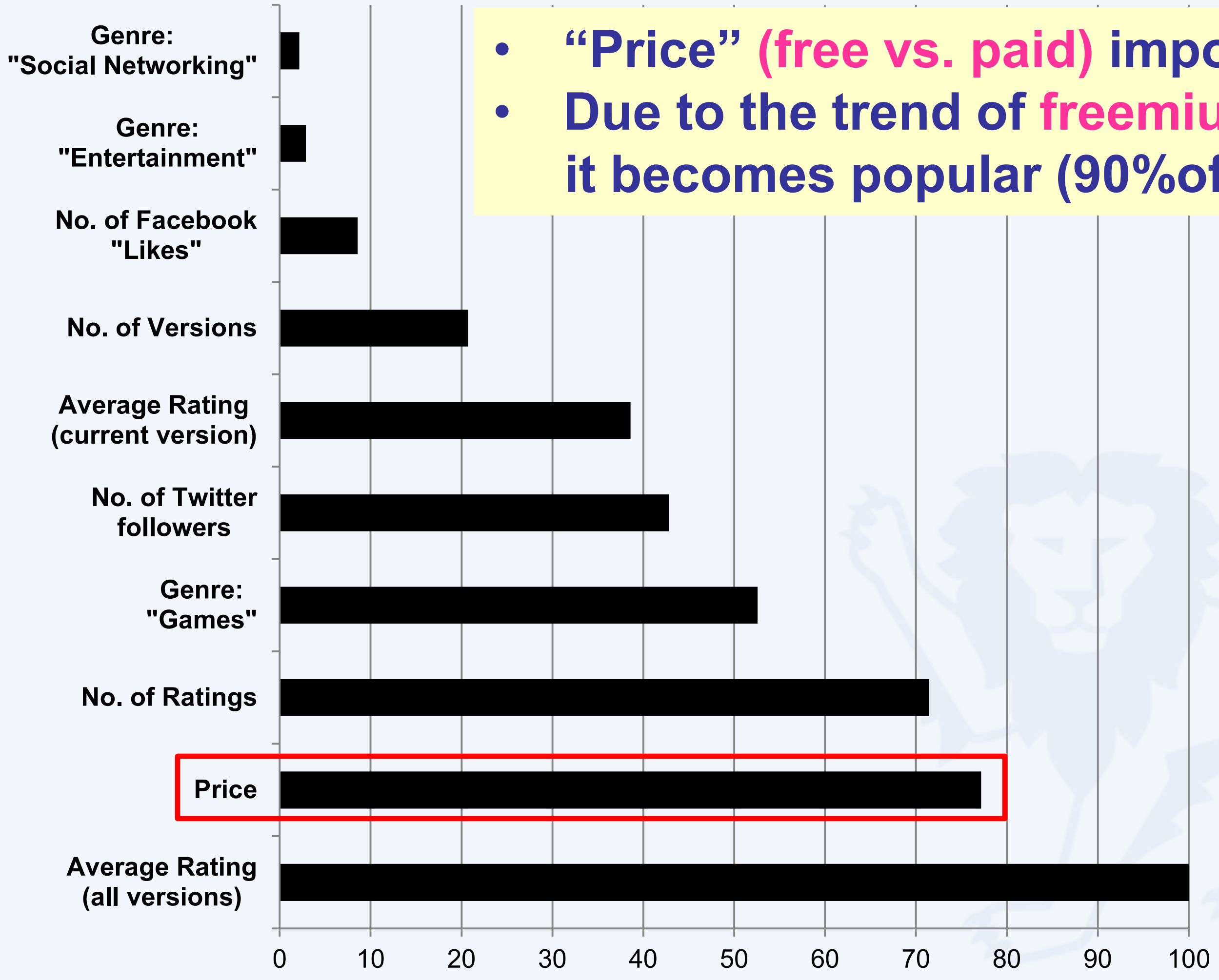


Feature Importance

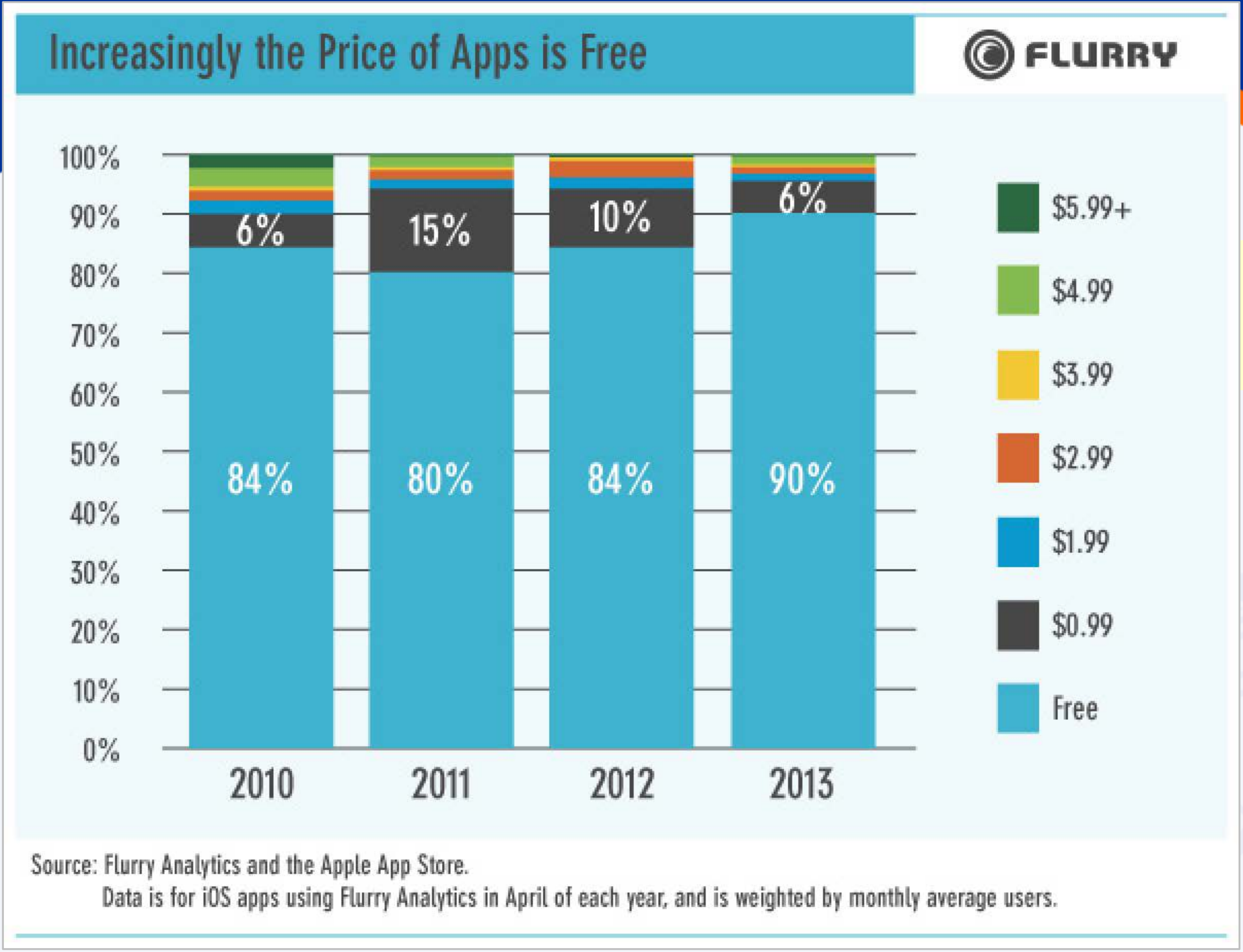


- "Average rating" is the most important.
- When the "average rating" is high, it is highly possible that the app is downloaded.

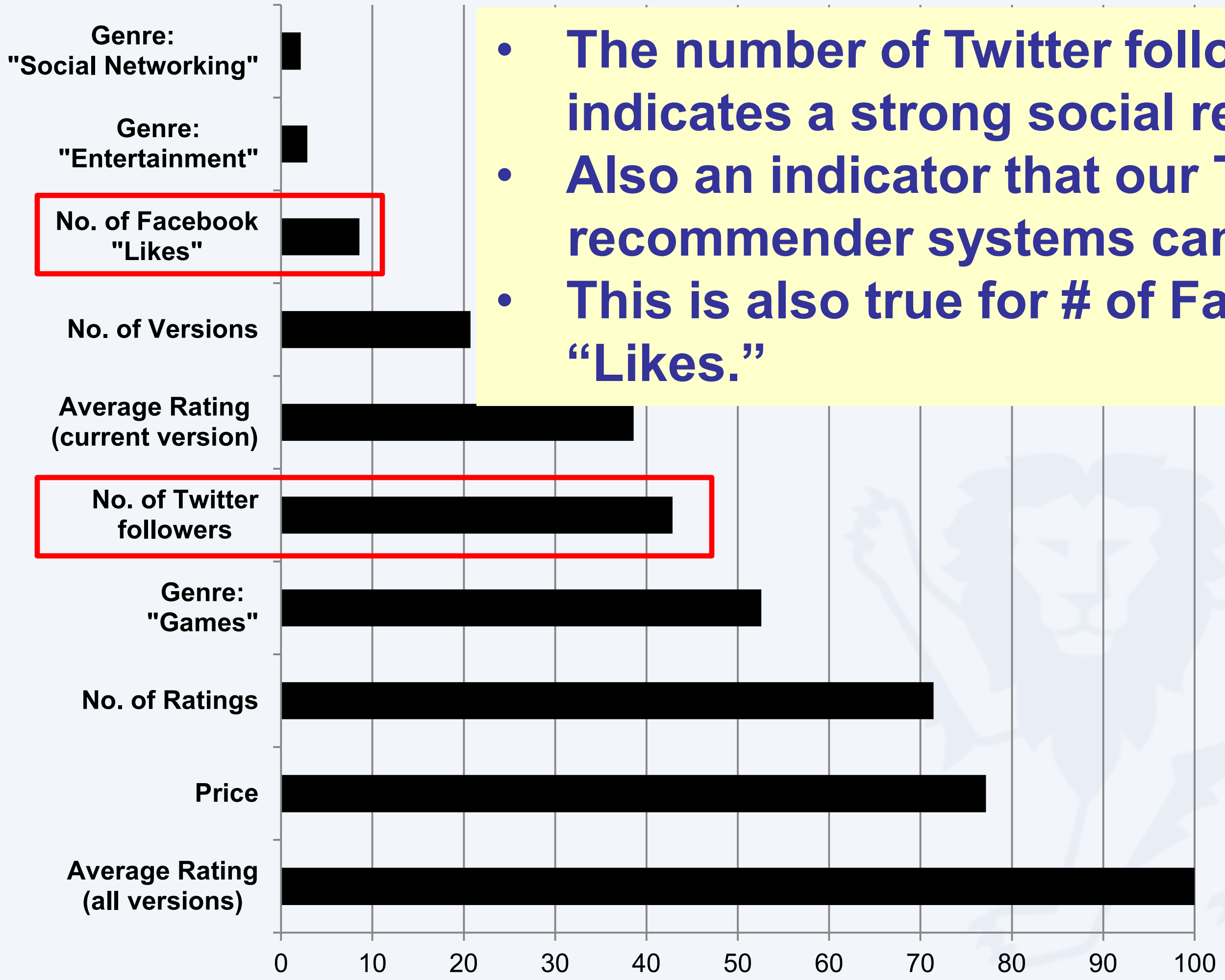
Feature Importance



- "Price" (free vs. paid) important.
- Due to the trend of freemium apps, it becomes popular (90%of app store).

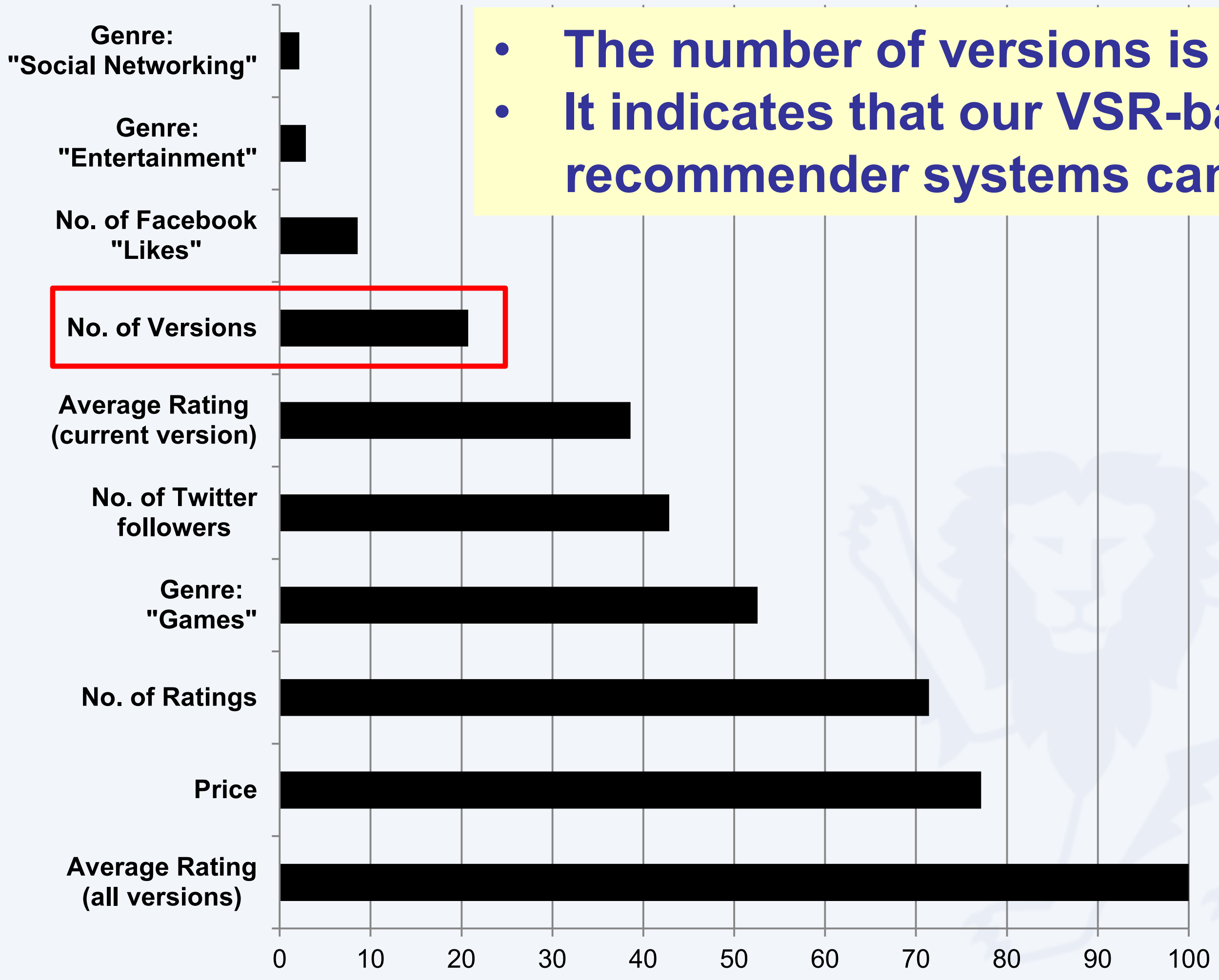

Increasingly the Price of Apps is Free
FLURRY
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
6%
15%
10%
6%
84%
80%
84%
90%
2010
2011
2012
2013
$5.99+
$4.99
$3.99
$2.99
$1.99
$0.99
Free
Source: Flurry Analytics and the Apple App Store.
Data is for iOS apps using Flurry Analytics in April of each year, and is weighted by monthly average users.

Feature Importance



- The number of Twitter followers (TWF) indicates a strong social reach.
- Also an indicator that our TWF-based recommender systems can be helpful.
- This is also true for # of Facebook "Likes."

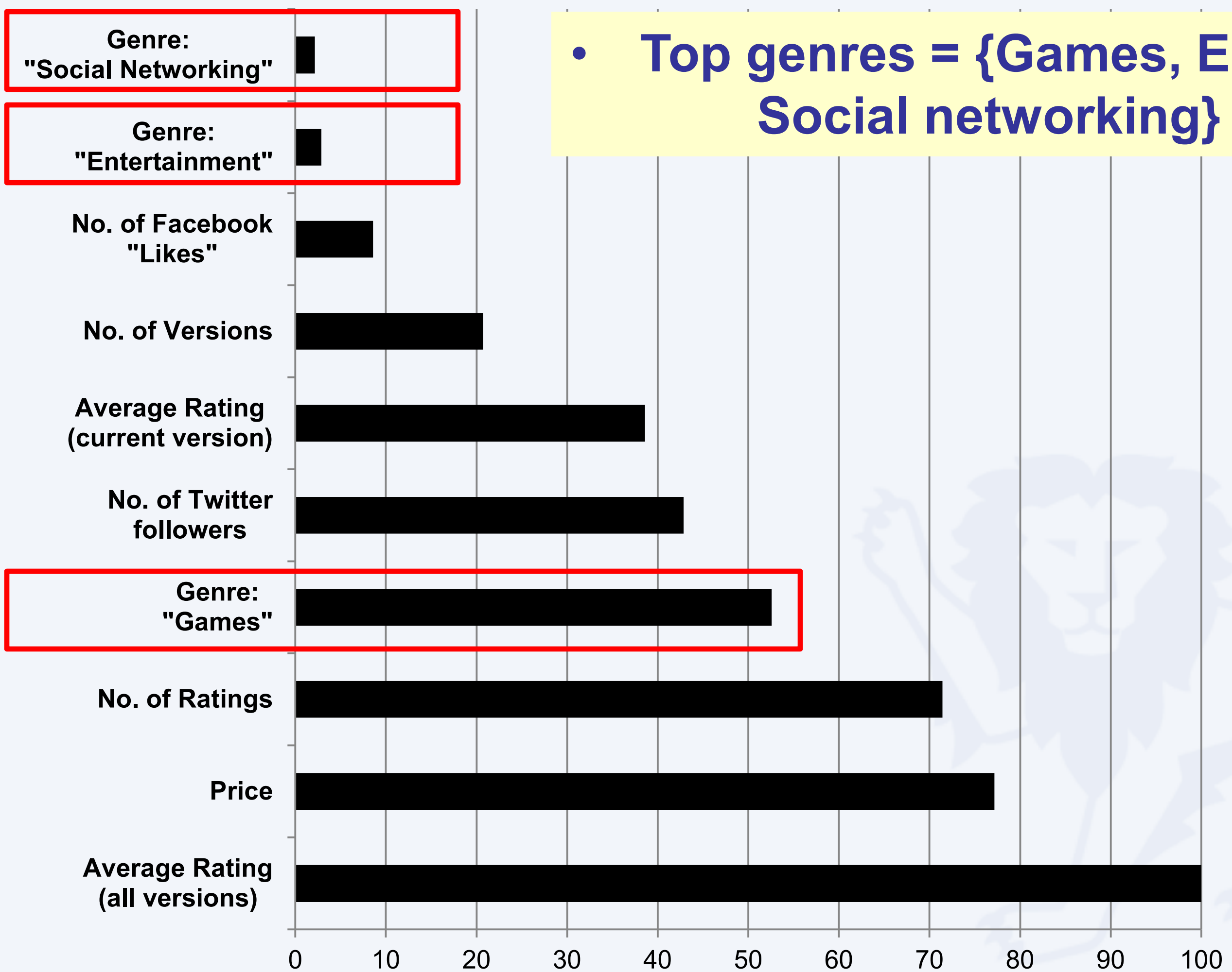
Feature Importance



- The number of versions is important.
- It indicates that our VSR-based recommender systems can be helpful.

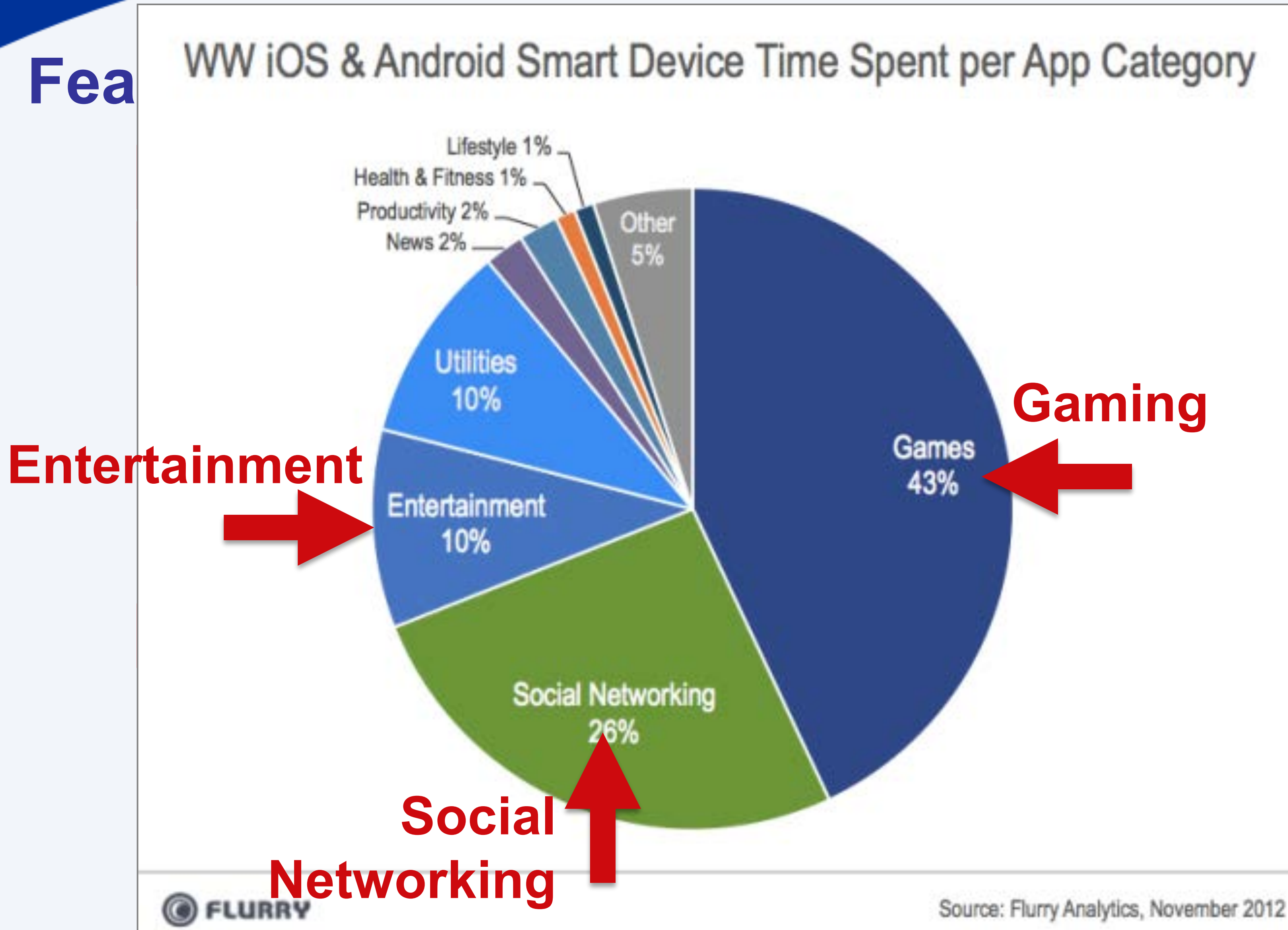
Feature Importance

- Top genres = {Games, Entertainment, Social networking}



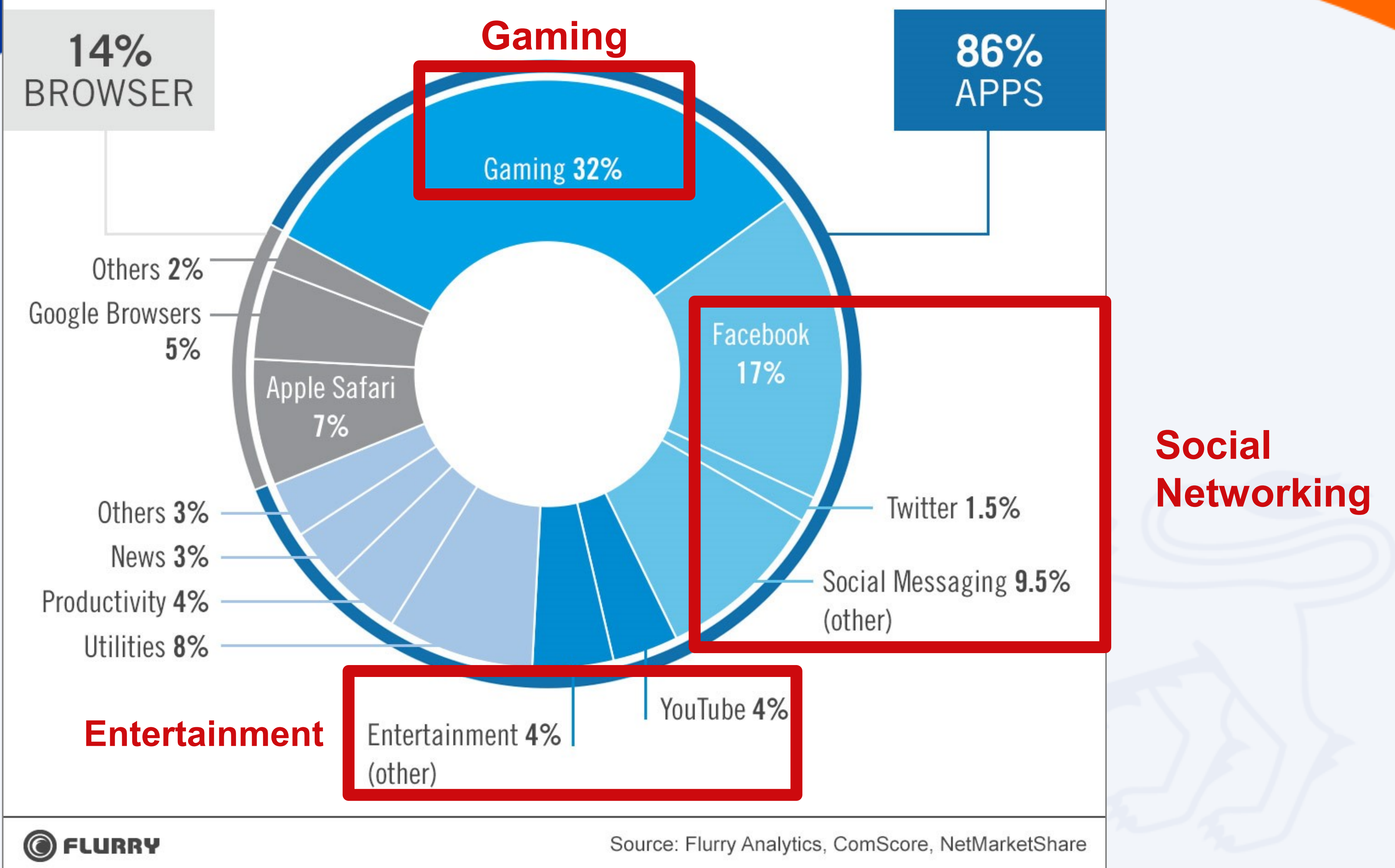
Fea

ment,



WW iOS & Android Smart Device Time Spent per App Category
Lifestyle 1%
Health & Fitness 1%
Productivity 2%
News 2%
Other 5%
Utilities 10%
Entertainment 10%
Social Networking 26%
Games 43%
Gaming
Entertainment
Social Networking
FLURRY
Source: Flurry Analytics, November 2012

Time Spent on iOS and Android Connected Devices



Conclusion

- **Employ GTB to integrate the features for the unifying recommendation techniques.**
- **Observe interesting correlations between important feature components and studies from third-party app analytics**
- **Our studies indicate that mobile app recommendation systems need to**
  - Further focus on user and trend analysis in social networks
  - Treat genre information with more importance

Thank you very much!