Abstract

Introduction

Data Description

We built a Facebook App[1] to collect information about users, their interactions and preferences. Our dataset contains information about each App user, along with a subset of information about their friends visible to the App. The data collection is performed with full permission from the user and in accordance with an approved Ethics Protocol[2].

Over 200 users installed the Facebook App sometime during the evaluation period. At any time, around 100 users have actively used the App. From these core App users, the App has access to their detailed Facebook profiles and their interactions with a total of 39,850 friends. While we have complete interaction data for the App users with their friends, and profile data (including wall post data) for the App users and friends, we do not have complete interactions for the App users' friends (unless they themselves are App users). Hence in the forthcoming analysis, we limit our evaluation to App users for which we are assured to have full interaction data.

Our App tracks many user (and their friends') details and interactions on Facebook. Interactions that occur through wall posts provide a rich variety of content and interaction data. We distinguish four Facebook items from wall posts: general posts (e.g., status updates, activity updates such as new friends, and interactions such as the user liked these pages), links, photos and videos. Four main interactions on these items are permitted by Facebook: posting an item to a friend's wall, commenting, liking, and tagging[3]. The App does not track deletions of these items and interactions (e.g., unlike) for performance reasons and we found very few deletions during an initial testing stage.

We summarize relevant basic statistics of the data in Table 1-3 below. The tables distinguish the data from the App users and from all App users and friends. Table 1 summarizes the number of records for each item (row) and interaction (column) combination. Table 2 shows some demographics from user profiles[4].

Methodology

Evaluation

- Page likes are most predictive followed by group and favourites
- Mutual information vs Size plot shows that the most predictive groups/page/favourites are spread around middle.

[1]Name and link omitted for anonymity.

[2]Link omitted for anonymity.

[3]Some Facebook interaction features such as liking comments were introduced after App user studies began and so are not tracked.

[4]Note that count of schools are not unique as each user can attend more than one degree of the same type.

App Users	Posts	Tags	Comments	Likes
Wall	27,955	5,256	15,121	11,033
Link	3,974	—	5,757	4,279
Photo	4,147	22,633	8,677	5,938
Video	211	2,105	1,687	710
App Users and Friends	**Posts**	**Tags**	**Comments**	**Likes**
Wall	3,384,740	912,687	2,152,321	1,555,225
Link	514,475	—	693,930	666,631
Photo	1,098,679	8,407,822	2,978,635	1,960,138
Video	56,241	858,054	463,401	308,763

Table 1: Number of records in Items and Interactions Tables. Rows are type of Facebook item and columns are type of Facebook interaction.

Table	#Records (App Users)	#Records (App User and Friends)
Users	103	39,850
Column	**#Non-empty (App Users)**	**#Non-empty (App User and Friends)**
Gender	102	36,401
Birthday	103	27,624
Breakdown	**Count (App Users)**	**Count (App User and Friends)**
Male	73	19,742
Female	29	16,659
High School	104	29,503
College	115	29,223
Graduate School	56	7733

Table 2: App user demographics.

Table	#Records (App Users)	#Records (App User and Friends)
Groups	3,469	373,608
Page Likes	10,771	825,452
Favourites	TBA	TBA

Table 3: Groups of Interests Tables

	Friend recommendation	Non-Friend recommendation
Like	1844	1393
Dislike	1359	2807

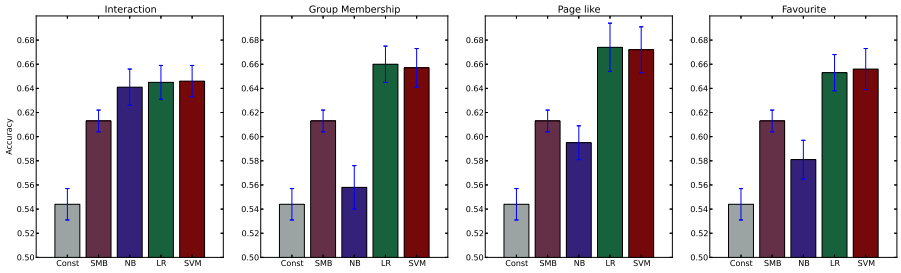

Figure 1: Accuracy plots

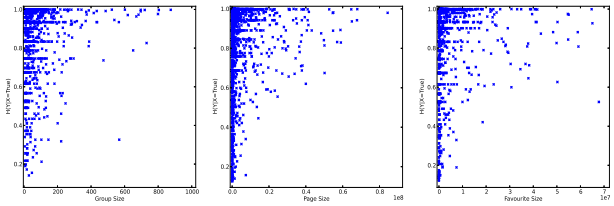

Figure 2: Conditional Entropy vs Size

This signifies that medium sized groups/pages/favourites are more predictive (!!!although not very strong signal)

- Collapsed mutual information ranking
  Video > Photo > Post > Link
  Comments > Like > Tags
  (No significant difference between Incoming and Outgoing)

- For Large social networks encoding user's membership(group/pages/favourites) as features and performing matrix factorization is not scalable (as there are millions of groups/pages/activities). This research shows that the memebership can be directly plugged into existing highly scalable algorithms to achieve better accuracy than state of art matrix factorization techniques.

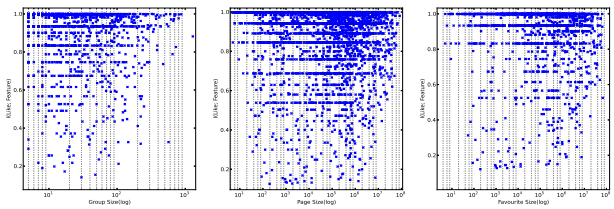

Figure 3: Conditional Entropy vs Size(log)

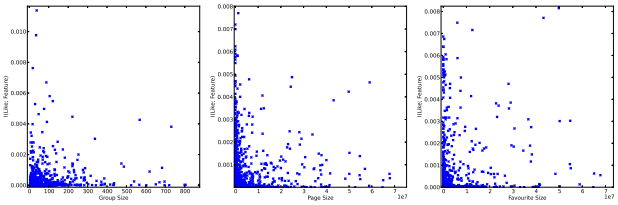

Figure 4: Mutual Information vs Size

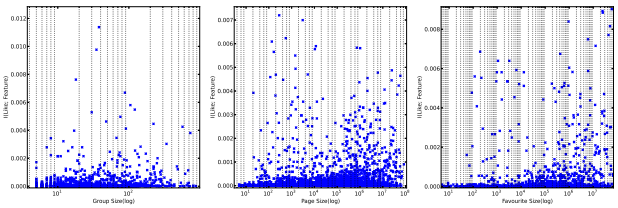

Figure 5: Mutual Information vs Size(log)

Modality	ConditionalEntropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
video	0.850116919421	117	44	2402	2962
link	0.914700608855	989	486	1530	2520
post	0.918160390299	1154	576	1365	2430
photo	0.9259513318	675	349	1844	2657
Modality	MutualInformation	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
post	0.0594688018722	1154	576	1365	2430
link	0.0490133947836	989	486	1530	2520
photo	0.0273700186713	675	349	1844	2657
video	0.0064494273283	117	44	2402	2962
Type	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
Tags	0.919971597996	809	407	1710	2599
Comments	0.921373670369	1009	511	1510	2495
Likes	0.924414408934	1135	583	1384	2423
Type	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
Likes	0.0553328447681	1135	583	1384	2423
Comments	0.0479360954681	1009	511	1510	2495
Tags	0.0361031572103	809	407	1710	2599
Direction	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
Outgoing	0.049470651248	1074	562	1445	2444
Incoming	0.0470981200647	1081	584	1438	2422
Direction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
Outgoing	0.928353525673	1074	562	1445	2444
Incoming	0.934921690705	1081	584	1438	2422

Interaction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
VIDEO_LIKES_OUTGOING	0.722397472066	31	7	2488	2999
VIDEO_LIKES_INCOMING	0.775086171837	43	12	2476	2994
POST_TAGS_INCOMING	0.802855788214	444	143	2075	2863
PHOTO_COMMENTS_OUTGOING	0.816036600382	135	45	2384	2961
PHOTO_TAGS_OUTGOING	0.818540125915	427	145	2092	2861
PHOTO_LIKES_INCOMING	0.853009353418	277	106	2242	2900
LINK_LIKES_OUTGOING	0.855368864256	728	282	1791	2724
POST_TAGS_OUTGOING	0.856163233945	505	196	2014	2810
LINK_COMMENTS_INCOMING	0.865533720447	530	213	1989	2793
POST_LIKES_OUTGOING	0.870618130159	834	342	1685	2664
VIDEO_COMMENTS_OUTGOING	0.883596675997	43	18	2476	2988
LINK_LIKES_INCOMING	0.888406184542	742	326	1777	2680
POST_LIKES_INCOMING	0.889444522106	929	410	1590	2596
POST_COMMENTS_INCOMING	0.890449138325	763	338	1756	2668
PHOTO_TAGS_INCOMING	0.890856994354	485	215	2034	2791
POST_COMMENTS_OUTGOING	0.894129257133	694	312	1825	2694
LINK_COMMENTS_OUTGOING	0.895134128697	543	245	1976	2761
PHOTO_LIKES_OUTGOING	0.89759612767	161	73	2358	2933
PHOTO_COMMENTS_INCOMING	0.906541518549	290	137	2229	2869
VIDEO_COMMENTS_INCOMING	0.92652449307	53	27	2466	2979
FRIENDS	0.965942963751	1392	895	1127	2111
VIDEO_TAGS_INCOMING	0.999999611934	17	17	2502	2989
VIDEO_TAGS_OUTGOING	0.999999611934	16	16	2503	2990

Interaction	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)
POST_LIKES_INCOMING	0.0524561740233	929	410	1590	2596
POST_LIKES_OUTGOING	0.050407410911	834	342	1685	2664
FRIENDS	0.0473539909731	1392	895	1127	2111
LINK_LIKES_OUTGOING	0.0457244144031	728	282	1791	2724
POST_COMMENTS_INCOMING	0.0405155210996	763	338	1756	2668
LINK_LIKES_INCOMING	0.0396004717225	742	326	1777	2680
POST_COMMENTS_OUTGOING	0.0352609340207	694	312	1825	2694
POST_TAGS_INCOMING	0.0315770377754	444	143	2075	2863
LINK_COMMENTS_INCOMING	0.0299391900579	530	213	1989	2793
POST_TAGS_OUTGOING	0.0296108833783	505	196	2014	2810
PHOTO_TAGS_OUTGOING	0.0286060075004	427	145	2092	2861
LINK_COMMENTS_OUTGOING	0.0261662648974	543	245	1976	2761
PHOTO_TAGS_INCOMING	0.0235803030398	485	215	2034	2791
PHOTO_LIKES_INCOMING	0.0155060725343	277	106	2242	2900
PHOTO_COMMENTS_INCOMING	0.0120447345005	290	137	2229	2869
PHOTO_COMMENTS_OUTGOING	0.00851951888057	135	45	2384	2961
PHOTO_LIKES_OUTGOING	0.00687365820792	161	73	2358	2933
VIDEO_LIKES_INCOMING	0.00309679794205	43	12	2476	2994
VIDEO_LIKES_OUTGOING	0.00259896191026	31	7	2488	2999
VIDEO_COMMENTS_OUTGOING	0.00197259388552	43	18	2476	2988
VIDEO_COMMENTS_INCOMING	0.0017837362825	53	27	2466	2979
VIDEO_TAGS_INCOMING	3.598735115e-05	17	17	2502	2989
VIDEO_TAGS_OUTGOING	3.39757297236e-05	16	16	2503	2990

Related Work

Conclusions