

Part I: Designing a Measurement and Evaluation Plan (joint)

This work will show how to make recommender system for internet shop.

Translation of Business Goals into Metrics

To identify quality metrics of recommender system we need clearly understand business goal and restrictions. First of all, we will speak not about just increasing selling but about increasing of profit. In my opinion, we can sell many cheap goods but it doesn't get big profit to company while if you buy one expensive product with a few cheap products it can bring more money with the same satisfaction of customers or much more. So we should recommend combination of expensive and cheap products to customer. Second, we are making user oriented recommender system it means we should give recommendation individually for customers based on their past experience if it's possible. These recommendations should be from different categories to increase profit selling products not only related with school (it can be office products). Third, internet shop has big advantage it's availability of different products, which not exist in local shops. So recommend products from this category we realize selling deep products from catalog.

Constraints:

1. On landing page we can recommend only 5 products.
2. On page with products we can show only 5 recommendations (it can be cross sale)

Target is to rise selling in back-to-school period with rising profit. It can be explained with formulation below:

$$P = \sum n C$$

where n – number of selling for one product, C – cost for one product. P – profit. Using this parameter in combination with parameters, which will be explained below, we will find the best algorithm for recommender system.

Based on prediction of ratings and their real values we can understand how adequate our algorithms using RMSE or MAE. It will show us accuracy of algorithms on known values, but it's not enough because ranking is very important for recommender system. It means that on test our algorithm should set up existing results on top of the ranking, because customer bought it and he likes it, so we think if he did not buy other products it means he likes them less or dislikes. For this can be used MAP (Mean Average Precision) because we need have good result in first 5 recommendations so position in rating is very important for us. From this five recommended products will be calculated profit (P). Additionally will be used precision-recall metrics, for future explanation I will use term precision-recall for top 5 products ($P@5$, $R@5$). We ignore Rank Correlation because for us most interest top of the list, but this metric shows just trend.

To make our recommender system more attractive to our goal we need additional metrics. To recommend products from different categories and price we need diversity metric (set up price gap). Next, it can be interesting to know personalization of our recommendations but we should be careful because it is incorrect to apply this approach to customers with small number of ratings. If we do not know anything about customer, we can't give personalized recommendation. One of the interesting metric is serendipity, it shows us how customer will be excited in good think of this word. To measure how deep products from catalog we recommend we can use parameter coverage.

Hard metrics: RMSE, MAE, MAP (MAP@5), P@5, R@5

Soft metrics: diversity, personalization, serendipity, coverage

We have task with a lot of parameters for optimization further we will look on them precisely.

Plan for evaluating base algorithms

Let's look on models used for prediction of ratings: TFIDF Content Based algorithm, Item-Item and User-User collaborative filtering, Matrix Factorization and personally-scaled average predictions. Based on our requirements for personal recommendations we prefer algorithms founded on users and they parameters. So for the first step good choice will be User-User collaborative filtration, Matrix Factorization and in minor TFIDF Content Based algorithm. To give answer on question which algorithm is better for the task we need make exploration of our metrics (Hard metrics and Soft metrics).

For exploration data and its visualization has been used Python. This is not honor task so there is no tuning of model. Further we will look on our metrics precisely.

RMSE, MAE and MAP@5 calculates using formulation from lessons.

$$MAE = \frac{\sum |r_i - s_i|}{n}$$

$$RMSE = \sqrt{\frac{\sum (r_i - s_i)^2}{n}}$$

$$MAP@5 = \frac{1}{|U|} \sum AP(O(u))$$

Precision and recall calculates based on top 5 elements.

$$P@5 = \frac{\text{\#relevant returnd items in Top - 5}}{\text{\#returnd items (5 elements)}}$$

$$R@5 = \frac{\text{\#relevant returnd items in Top - 5}}{\text{\#relevant items}}$$

To calculate diversity among price we will subtract one price from other for each recommended product from Top-5. After that we rescale it using algorithm min-max for all users it's needed for comparing different people in one scale (from 0 to 1). Diversity calculate for each customer. To get one number we just average this number for all users, its will be averaged diversity for algorithm.

$$Diversity_{user} = \frac{1}{\#combinations} \sum_i^4 \sum_{j>i} \frac{|price_i - price_j| - \min(|price_k - price_d|)}{\max(|price_k - price_d|) - \min(|price_k - price_d|)}$$

$$Averaged\ Diversity = \frac{1}{|U|} \sum_i Diversity_i$$

To add category diversity we need add coefficient of diversity based on exploration of data. From there we can see that we have LeafCat and FullCat, first name show last level, second – hierarchy. So we drop one repeated element from hierarchy its Office Products, after that we have four categories.

Example Name: Office & School Supplies/Desk Accessories & Workspace Organizers/Mouse Pads & Wrist Rests/Mouse Pads

Level 1: Mouse Pads

Level 2: Mouse Pads & Wrist Rests

Level 3: Desk Accessories & Workspace Organizers

Level 4: Office & School Supplies

Level	Diversity
Different 1 level	0.25
Different 2 level	0.5
Different 3 level	0.75
Different 4 level	1.0

Personalization metric is very important in this task but calculate it is very difficult and correct only for customers with many ratings. Compare two vectors with each other and delete on count of users to fined coefficient of personalization.

To calculate serendipity we will use formulation from course:

$$serendipity = \frac{1}{n} \sum_{i=1}^n \max(pred(s_i) - prim(s_i), 0) is_relevant(s_i)$$

$isrelevant(s_i) = 1$ if the user rely likes the item and 0 if not. 1 if rating more than 3

$prim(s_i) = popularity\ of\ item$

Is relevant means customer likes this product, rating more than 3.

Availability show us how deep we are in catalog, if the number small it means we are deep in catalog. To get one number we will use averaged number.

Coverage@5 will show us how many products we use in top-5 recommendations, we will use percentages from all products using recommendation for all users.

In the future we will see how each of these metrics effective.

Hybridization

As we know we have two places for recommendations so we will use two different algorithm for them it will be first hybridization. It helps us split cold start from other.

Second, If we know something about previous experience of client it will be better to give recommendation using combination of CBF, Item-Item, MF, PersBias and User-User algorithms

$$R = \alpha_1 CBF + \alpha_2 ItemItem + \alpha_3 MF + \alpha_4 PersBias + \alpha_5 UserUser$$

α – coefficient. This coefficient from 0 to 1. Sum of these coefficients must be 1. It gave us possibility to use superposition of algorithms based on rating count. Because some algorithms works better with known customers and some algorithms work better with people with small number of ratings.

The quality of recommendation algorithm will be evaluated using metrics explained before.

Part II: Measurement

In this section will be analyzed provided data using metrics above (RMSE, MAE, MAP@5, P@5, R@5, diversity among price, diversity among categories, personalization, serendipity, mean availability, catalog coverage). For RMSE and MAE we analyze they distribution on customers and averaged value for algorithm. For MAP@5, P@5, R@5 we use threshold for items relevance greater 3. For diversity we took calculation for each customers and average it for algorithm. For serendipity we used threshold like for MAP@5, P@5, R@5. The $\text{prim}()$ calculates like popularity of product normalized to space from 0 to 5.

Continue we will show different metrics and analyze them, in the end we will summarize results in one table.

RMSE and MAE

On pictures bellow we show changing of metrics for customers. In the table, we see error on different algorithms. We can see the biggest error has PersBias algorithm.

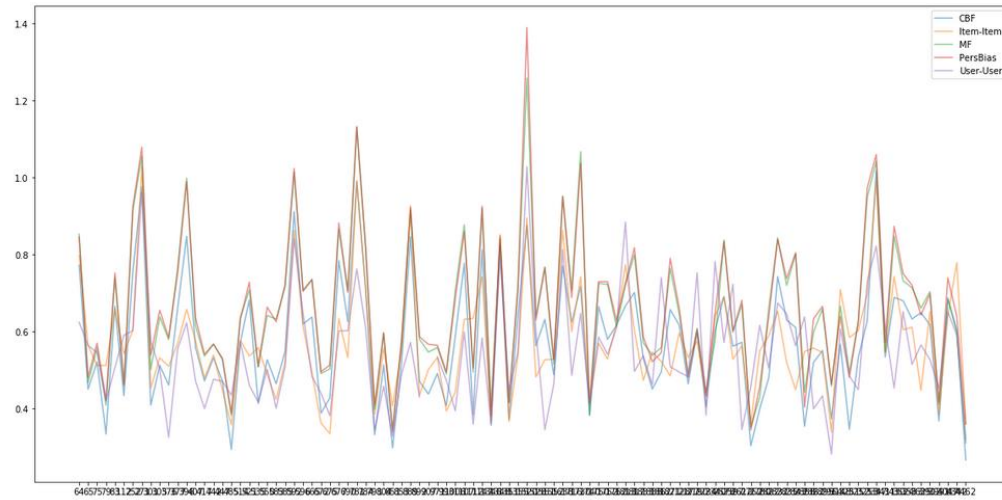


Figure 1. Ratings RMSE for users using different algorithms

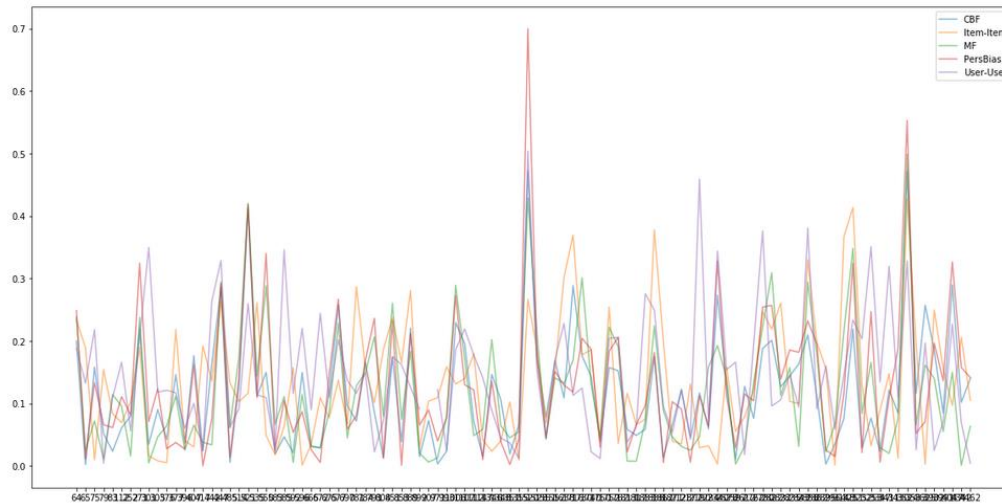


Figure 2. Ratings MAE for users using different algorithms

Mean RMSE and MAE for algorithm

Algorithm	Mean RMSE	Sdt RMSE	Mean MAE	Std MAE
CBF	0.572387	0.162112	0.119093	0.095137
Item-Item	0.574672	0.147212	0.135706	0.100839
MF	0.659029	0.188677	0.121366	0.103932
PersBias	0.666273	0.192842	0.133248	0.117852
User-User	0.545130	0.142089	0.149009	0.105061

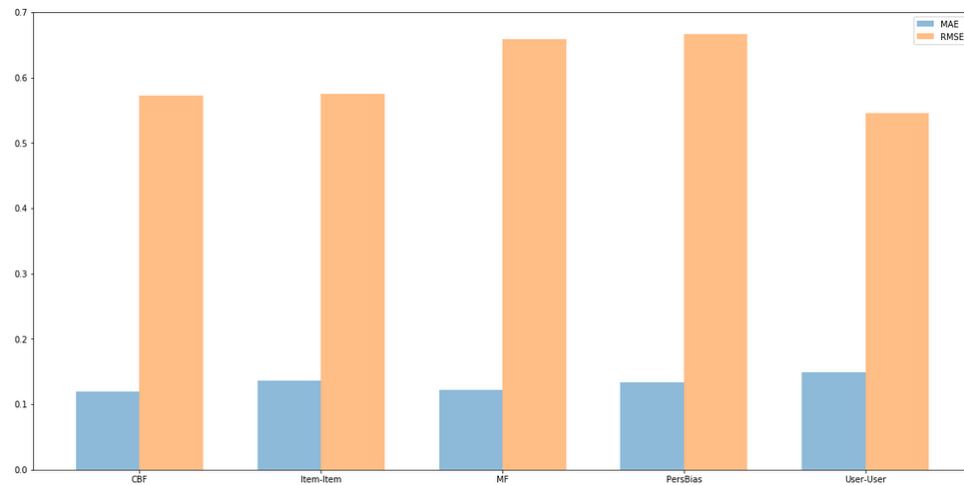
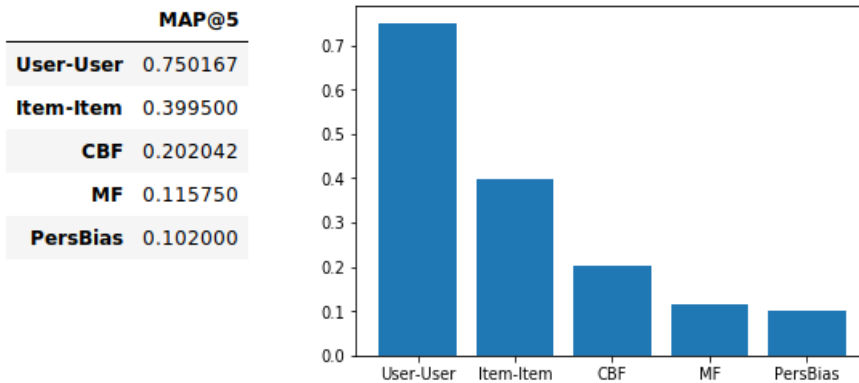


Figure 3. Comparing of algorithms errors

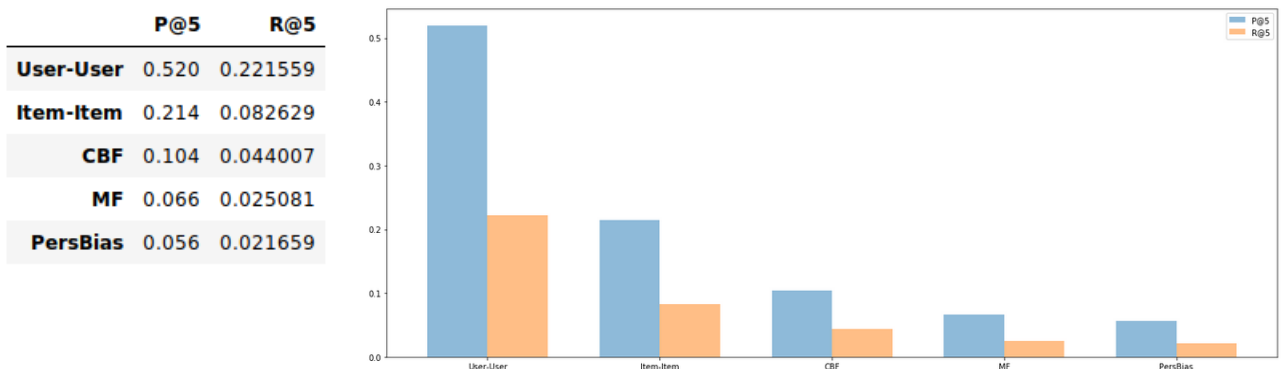
MAP@5

MAP@5 shows us how many relevant data exist in top 5 elements. User-User algorithm shows the best result, it work better than MF it looks very strange for this type algorithms.



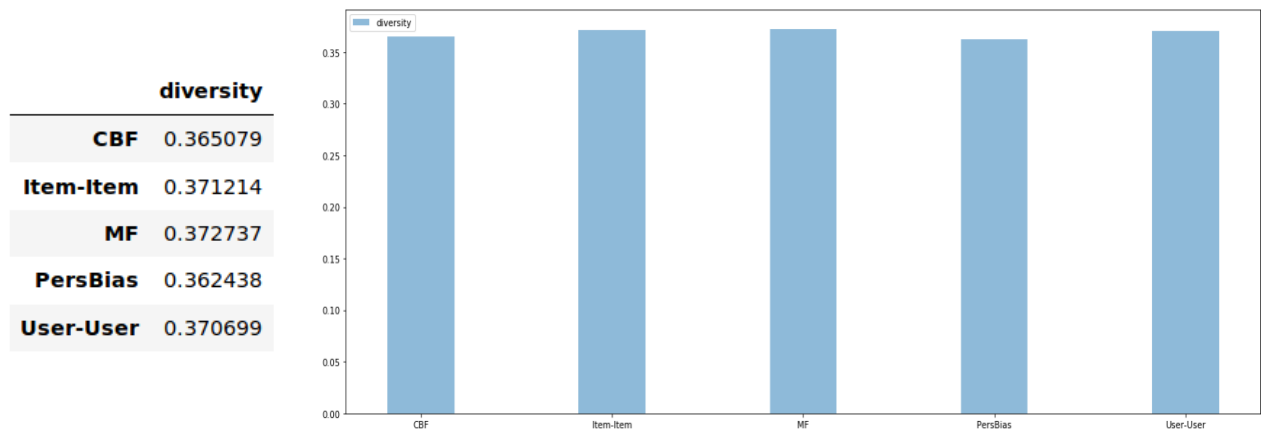
P@5, R@5

On hard metrics User-User algorithm shows the best results, it means we have enough clients with ratings to make similarity in user-user dimension.



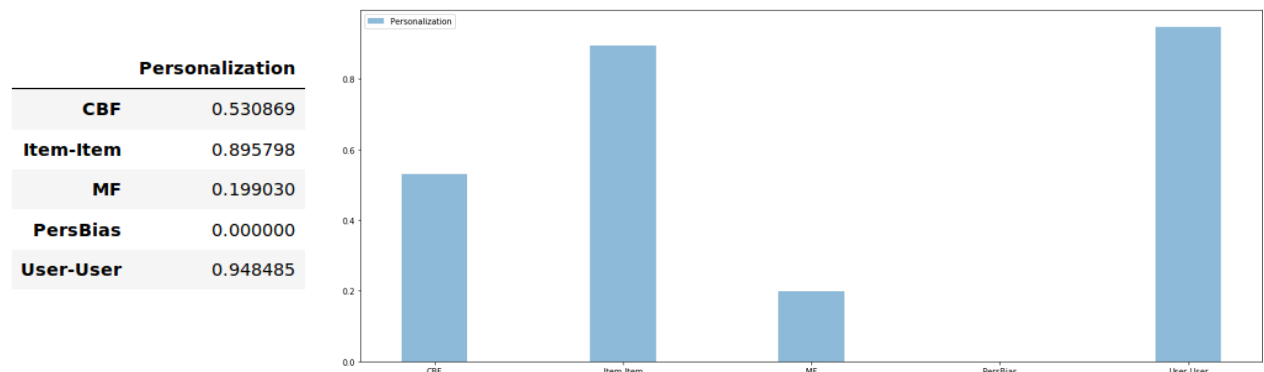
Diversity among prices

All algorithms have the same values.



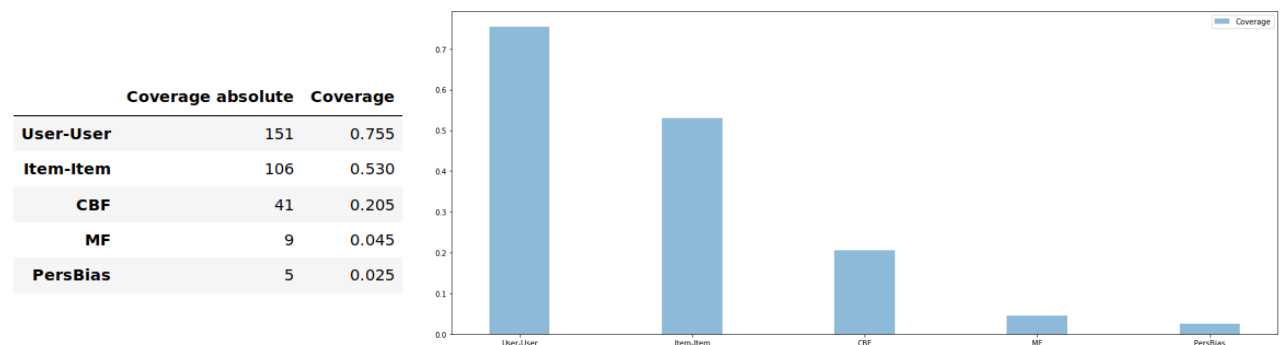
Personalization

We have zero result for PersBias algorithm, because it is constant approach. And again User-User algorithm shows the biggest result.



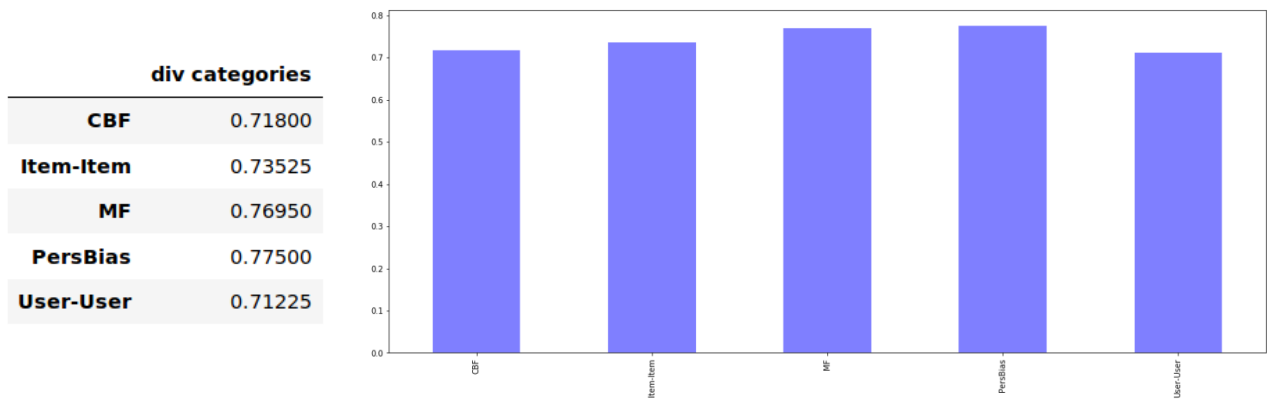
Catalog coverage

Catalog coverage shows how deep in catalog we are so it is important parameter. So based on user-user dependency we have the best result.



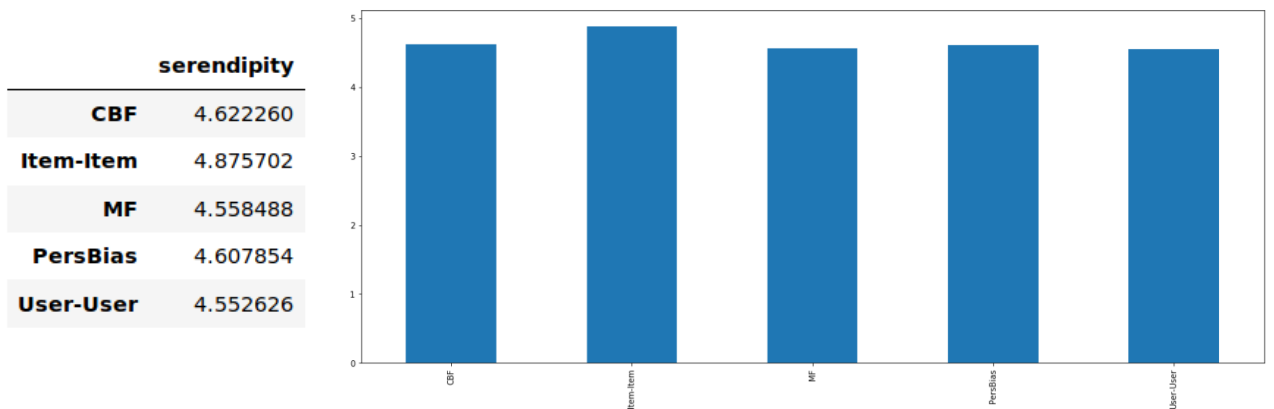
Diversity among categories

Here we can see that all algorithms give as the same result and diversity not very big.



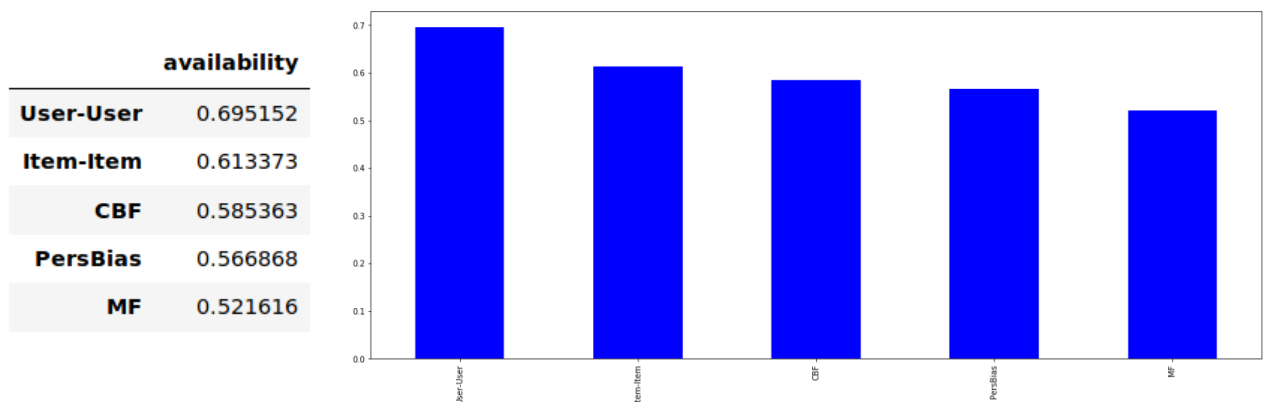
Serendipity

So serendipity is not too different for each algorithm.



Mean availability

Mean availability should be less in our task because it rise probability to buy product in internet.



Summary

	MAE_ mean	MAE_ std	RMSE_ mean	RMSE_ std	MAP@5	P@5	R@5	diversity (price)	Pers	Cover absolute	Cover	Div (cat.)	serend	availability
User-User	0.149	0.105	0.149	0.105	0.750	0.520	0.222	0.371	0.948	151.000	0.755	0.712	4.553	0.695
Item-Item	0.136	0.101	0.136	0.101	0.400	0.214	0.083	0.371	0.896	106.000	0.530	0.735	4.876	0.613
CBF	0.119	0.095	0.119	0.095	0.202	0.104	0.044	0.365	0.531	41.000	0.205	0.718	4.622	0.585
MF	0.121	0.104	0.121	0.104	0.116	0.066	0.025	0.373	0.199	9.000	0.045	0.770	4.558	0.522
PersBias	0.133	0.118	0.133	0.118	0.102	0.056	0.022	0.362	0.000	5.000	0.025	0.775	4.608	0.567

Table above shows aggregated results. From there we can see that User-User algorithm shows the best result and we will base on it. Second is Item-Item. Item-Item is pretty close to User User in many parameters. It is better in serendipity.

Coverage of MF is too small, just 9 products, its very small for coverage and personalization. MF will be ineffective. MF shows good result only for diversities, but it is too weak parameters.

PersBias has no personalization.

So in the future, it will be good approach to combine strong algorithms to get better results.

Part III: Mixing

Linear combination of User-User and Item-Item (UU&II)

As we saw there is two algorithms User-User and Item-Item with the best results, so first of all we will make linear combination of them:

$$UU\&II = \alpha UU + (1 - \alpha) II$$

In table below we can see metrics from alpha value. From there we can see that with alpha = 0.8 we have big diversity price and MAP@5, enough small MAE and RMSE values. In the end we got smaller hard metrics then in User-User (alpha = 1) and not strongly changed soft metrics.

	RMSE	MAE	MAP@5	P@5	R@5	div_p	pers	cov_abs	cov	div_c	serend	availability
0.0	0.57541	0.00148043	0.3895	0.21	0.0809626	0.370674	0.897212	106	0.53	0.73175	4.82314	0.612677
0.1	0.555624	0.000932532	0.432958	0.232	0.0906948	0.369927	0.901212	104	0.52	0.735	4.77206	0.62288
0.2	0.53888	0.000384638	0.49	0.272	0.108113	0.364949	0.905616	103	0.515	0.7235	4.72456	0.631969
0.3	0.525559	-0.000163256	0.535333	0.31	0.124163	0.368486	0.909091	104	0.52	0.7235	4.67702	0.644489
0.4	0.516031	-0.000711151	0.607528	0.362	0.146211	0.367386	0.91604	108	0.54	0.717	4.63556	0.65425
0.5	0.510599	-0.00125905	0.676569	0.376	0.153383	0.36785	0.914141	111	0.555	0.7145	4.60827	0.659097
0.6	0.509431	-0.00180694	0.687264	0.414	0.169333	0.367819	0.918626	117	0.585	0.71325	4.56907	0.677487
0.7	0.512518	-0.00235483	0.737264	0.444	0.183448	0.369669	0.925737	128	0.64	0.70575	4.54482	0.692617
0.8	0.519688	-0.00290273	0.763764	0.474	0.196594	0.377006	0.932404	141	0.705	0.70975	4.53531	0.695332
0.9	0.530668	-0.00345062	0.758667	0.5	0.210674	0.372474	0.941576	143	0.715	0.71125	4.53592	0.697036
1.0	0.54513	-0.00399852	0.750167	0.52	0.221559	0.370699	0.948485	151	0.755	0.71225	4.55263	0.695152

Diversification among price

Second approach based on diversification among price. Main steps:

- Take top 10 instead 5 elements
- Calculate price difference from first element
- Create new score adding result to main score.
- Re rank items using new score and take top 5

	MAP@5	P@5	R@5	div_p	pers	cov_abs	cov	div_c	serend	availability
User-User_price	0.511083	0.384	0.156223	0.373164	0.939758	160	0.8	0.73675	5.65341	0.68692
Item-Item_price	0.274736	0.176	0.0682815	0.379403	0.900646	98	0.49	0.7735	6.22911	0.606561
CBF_price	0.204861	0.1	0.0409972	0.368229	0.624848	46	0.23	0.752	6.52206	0.612983
MF_price	0.128167	0.066	0.0244844	0.360167	0.418141	15	0.075	0.788	6.58353	0.562655
PersBias_price	0.188833	0.068	0.0277949	0.348951	0	5	0.025	0.8	7.27279	0.606945

Diversification among category

Third approach based on diversification among category. Main steps:

- Take top 10 instead 5 elements
- Calculate category difference from first element
- Create new score adding result to main score.
- Rerank items using new score and take top 5

	MAP@5	P@5	R@5	div_p	pers	cov_abs	cov	div_c	serend	availability
User-User_category	0.60525	0.414	0.171034	0.375843	0.950465	156	0.78	0.73275	8.41178	0.690322
Item-Item_category	0.346722	0.164	0.0624101	0.382475	0.913939	106	0.53	0.75225	8.85527	0.626048
CBF_category	0.164056	0.086	0.0351239	0.364749	0.619636	40	0.2	0.721	8.65954	0.591222
MF_category	0.138917	0.066	0.0259349	0.364879	0.228121	10	0.05	0.841	8.52181	0.545483
PersBias_category	0.147583	0.076	0.0309301	0.370688	0	5	0.025	0.775	8.48557	0.657933

Combination of UU&I (alpha 0.8) and diversification among price and category

Last approach is combination of three before. Steps:

1. Linear combination of User-User and Item-Item
2. Diversification among price
3. Diversification among category
4. Calculation of parameters

Result table

	MAP@5	P@5	R@5	div_p	pers	cov_abs	cov	div_c	serend	availability
User-User_price	0.511083	0.384	0.156223	0.373164	0.939758	160	0.8	0.73675	5.65341	0.68692
User-User_category	0.60525	0.414	0.171034	0.375843	0.950465	156	0.78	0.73275	8.41178	0.690322
UU&I_price_category	0.517403	0.382	0.154192	0.376734	0.933253	136	0.68	0.72975	7.69105	0.693117
Item-Item_category	0.346722	0.164	0.0624101	0.382475	0.913939	106	0.53	0.75225	8.85527	0.626048
Item-Item_price	0.274736	0.176	0.0682815	0.379403	0.900646	98	0.49	0.7735	6.22911	0.606561
CBF_price	0.204861	0.1	0.0409972	0.368229	0.624848	46	0.23	0.752	6.52206	0.612983
CBF_category	0.164056	0.086	0.0351239	0.364749	0.619636	40	0.2	0.721	8.65954	0.591222
MF_price	0.128167	0.066	0.0244844	0.360167	0.418141	15	0.075	0.788	6.58353	0.562655
MF_category	0.138917	0.066	0.0259349	0.364879	0.228121	10	0.05	0.841	8.52181	0.545483
PersBias_price	0.188833	0.068	0.0277949	0.348951	0	5	0.025	0.8	7.27279	0.606945
PersBias_category	0.147583	0.076	0.0309301	0.370688	0	5	0.025	0.775	8.48557	0.657933

From table we can see that User-User_category and price results have similar results in many parameters. The biggest difference we have on MAP@5, P@5 and serendipity, on it the best result give us User-User_category approach.

Table of recommended products for user 1730 by User-User_category

Item	Availability	ASIN	Price	Brand	Title	LeafCat	FullCat
2326	0.905137	B00G411O8G	9.99	NaN	Swingline Thermal Laminating Pouch, Letter Siz...	Laminating Supplies	Office Products/Office Electronics/Presentatio...
1938	0.454502	B005YGLA5Y	5.55	BIC	BIC For Her Fashion Retractable Ball Pen, Medi...	Ballpoint Pens	Office Products/Office & School Supplies/Writi...
2040	0.726813	B0085IPYO6	8.90	Smead	Smead MO File Box, Letter Size, 3 inch Expansi...	All-Purpose Labels	Office Products/Office & School Supplies/Label...
2247	0.406177	B00D4613Y4	8.01	NaN	Five Star Split Pencil Cup, Purple (72260)	Modular Storage Systems	Office Products/Office & School Supplies/Offic...
2041	0.726813	B0085IPZGS	11.45	Smead	Smead MO Records Filing Starter Kit, Letter Si...	All-Purpose Labels	Office Products/Office & School Supplies/Label...

Table of recommended products for user 1862 by User-User_category

Item	Availability	ASIN	Price	Brand	Title	LeafCat	FullCat
1032	0.644513	B001DPSUUI	23.00	Fellowes	Fellowes Hot Laminating Pouches, Letter, 5 mil...	Laminating Supplies	Office Products/Office Electronics/Presentatio...
670	0.981560	B000MFHX3U	7.67	3M	3M Whiteboard Eraser for Whiteboards, 2-Pack	Dry Erase Boards	Office Products/Office & School Supplies/Prese...
1217	0.991766	B0027CTFBO	21.79	Fellowes	Bankers Box SmoothMove Moving and Storage Boxe...	Box Mailers	Office Products/Office & School Supplies/Envel...
1297	0.945595	B002K9M7MS	NaN	NaN	3M Permanent Adhesive Shipping Labels, 2 x 4 I...	Shipping Labels	Office Products/Office & School Supplies/Label...
1317	0.901367	B002M7VZXA	15.16	NaN	Avery NoteTabs, 2 x 1.5 Inches, Neon Blue and ...	Removable Labels	Office Products/Office & School Supplies/Label...

Part IV: Proposal and Reflection

Back to school period has a huge lift in office products sales. Therefore, in this time we can sell many products from different categories and price, we just need recommend correct products to customers.

We present a project for constructing the recommendations that best fit our business needs.

We considered many parameters to make our model:

1. Good recommendation – we show people items they really like, we think so. Our “hard” metrics show us how often our algorithm makes mistakes. We can show just 5 items.
2. Catalog coverage, diversity of prices and categories – we want to show items from different categories, recommend expensive product with cheap. We want to show items not available in offline shops. And we want to sell many items from our catalog.
3. Personalization – we want to show personalized items, it should not be the same for different customers.

Based on these criterions we have made experiment with different recommender systems.

The best result shows User-User recommender algorithm with diversification by category. We recommend items to a user; we find a set of users with similar. To do so, we evaluate the ratings they have given and search for users with similar ratings for the same items. So we find people who like the same items like target user. We added some tweaks to this base model to increase the variety of products (diversification by category).

Making more than 10 models, we can say that User-User collaborative model has the best results based on metrics.

In the end I would like to make summarize. To estimate algorithms we used 5 **Hard metrics** and 4 Soft metrics. These metrics transform business goals to metrics. We had 5 main algorithms, based on them we made hybrids, they quality estimated by metrics. Project has been done on Python. Code can be looked on github: