NTNU
Norwegian University of
Science and Technology

Adaptive Aggregation of Recommender Systems

Olav Bjørkøy

Department of Computer and Information Science

October 3rd, 2011

Terminology

u	user
i	item (article, website, movie, product, other user...)
r	rating, relevance, connection (domain specific)
$p(u, i)$	predicted rating for a user and an item.
m	a method for predicting ratings.

NTNU
Norwegian University of
Science and Technology

2006: The Netflix Recommender Challenge

- USD 1MM prize for a 10% accuracy improvement.
- Breakthrough: Combining methods from many teams.
- Team BellKor finally achieved a 10.06% improvement by combining **107** different recommender algorithms.

Today: Web Search

Google

"Today we use more than 200 signals, including PageRank, to order websites, and we update these algorithms on a weekly basis." (google.com/corporate/tech.html)

Bing

"We use over 1,000 different signals and features in our ranking algorithm." (bing.com/community/site_blogs/b/search/archive/2011/02/01/thoughts-on-search-quality.aspx)

NTNU
Norwegian University of
Science and Technology

Why multiple algorithms?

— Use more data.
— Capture more predictive aspects.
— Disjoint predictors.

Bell, R., Koren, Y., and Volinsky, C. (2007) (Netflix)

"Quite frequently we have found that the more accurate predictors are less useful within the full blend."

NTNU
Norwegian University of
Science and Technology

$$\hat{r}_{u,i} = \sum_{m \in M} w_m \times p(m, u, i)$$

NTNU
Norwegian University of Science and Technology

The Problem: Latent Subjectivity

- Generalized weights.
- Treats all users and items the same.
- Varying accuracy across users.
- Varying accuracy across items.
- A case of misplaced subjectivity.

NTNU
Norwegian University of
Science and Technology

Systems that insist on being adaptive in a certain way are not really adaptive at all.

Adaptive Recommenders

— Combine multiple disjoint recommender systems.
— Use content-based, collaborative filtering and hybrids.
— Automatically predict their contextual accuracy.

$$\hat{r}_{u,i} = \sum_{m \in M} p_w(m,u,i) \times p_r(m,u,i) \tag{1}$$

NTNU
Norwegian University of Science and Technology

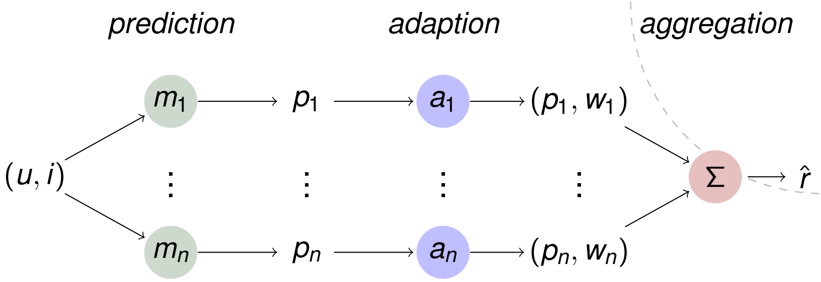

Figure: Layers of recommenders: The predictive layer consists of ordinary recommender systems. This produces a set of predicted ratings (p). The adaptive layer estimates how well each modeling method will perform for the current user and item. The aggregation weighs the predictions into a final score $\hat{r}$.

Training phase

- Split into two sets using bootstrap aggregation: (t_1, t_2).
- Use t_1 to train the basic recommenders.
- Create an error matrix using the basic recommenders and t_2.
- The error matrix values are predicted errors for (u, i, m).
- Train adaptive recommenders on the error matrix.

$$\forall(u, i, r) \in (d_e - d_m) : E(m)_{u,i} = |r - p(m, u, i)| \quad (2)$$

NTNU
Norwegian University of
Science and Technology

Prediction phase

- Calculate each prediction $\hat{r}_{(u,i,m)}$.
- Calculate each predicted error $\hat{e}_{(u,i,m)}$.
- The adaptive weights are the inverses of the normalized error.
- Sum the weighted predictions to get the final $\hat{r}$.

$$\hat{r}_{u,i} = \sum_{(m_e,m_r)\in M} \left(1 - \frac{p(m_e,u,i)}{error(u,i)}\right) \times p(m_r,u,i) \quad \text{where} \quad error(u,i) = \sum_{m_e \in M} p(m_e,u,i) \tag{3}$$

NTNU
Norwegian University of
Science and Technology

Results

- Calculate RMSE values for basic recommenders, simple aggregations and adaptive aggregation.
- Used the Movielens movie rating dataset (see paper for more details).

$$\mathrm{RMSE}(\hat{R}, R) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{R}_i - R_i)^2}{n}} \tag{4}$$

NTNU
Norwegian University of Science and Technology

(a) RMSE values for the five disjoint subsets:

	method	d_1	d_2	d_3	d_4	d_5
S	svd1	1.2389	1.1260	1.1327	1.1045	1.1184
S	svd2	1.2630	1.1416	1.1260	1.1458	1.1260
S	svd3	1.0061	0.9825	0.9830	0.9815	0.9797
S	svd4	1.0040	0.9830	0.9849	0.9850	0.9798
S	slope_one	1.1919	1.0540	1.0476	1.0454	1.0393
S	item_avg	1.0713	0.9692	0.9662	0.9683	0.9725
S	baseline	1.0698	0.9557	0.9527	0.9415	0.9492
S	cosine	1.1101	0.9463	0.9412	0.9413	0.9382
S	knn	1.4850	1.1435	1.1872	1.2156	1.2022
A	median	0.9869	0.8886	0.8857	0.8857	0.8855
A	average	0.9900	0.8536	0.8525	0.8525	0.8519
A	adaptive	**0.9324**	**0.8015**	**0.7993**	**0.8238**	**0.8192**

(b) Statistics for the methods:

	method	min	max	mean	σ	Δ
S	knn	1.1435	1.4850	1.2467	0.3487	-
S	svd2	1.1260	1.2630	1.1605	0.2277	6.9%
S	svd1	1.1045	1.2389	1.1441	0.2197	1.4%
S	slope_one	1.0393	1.1919	1.0756	0.2415	5.9%
S	item_avg	0.9662	1.0713	0.9895	0.2023	8.0%
S	svd4	0.9798	1.0040	0.9873	**0.0924**	2.2%
S	svd3	0.9797	1.0061	0.9865	0.0991	0.1%
S	cosine	0.9382	1.1101	0.9754	0.2595	1.1%
S	baseline	0.9415	1.0698	0.9738	0.2196	1.6%
A	median	0.8855	0.9865	0.9065	0.2005	6.9%
A	average	0.8519	0.9900	0.8801	0.2344	2.9%
A	adaptive	**0.7993**	**0.9324**	**0.8352**	0.2225	5.1%

NTNU
Norwegian University of
Science and Technology

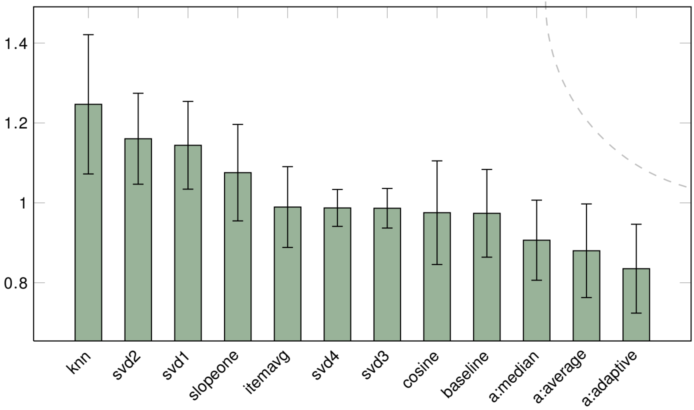

Figure: Average RMSE plot.

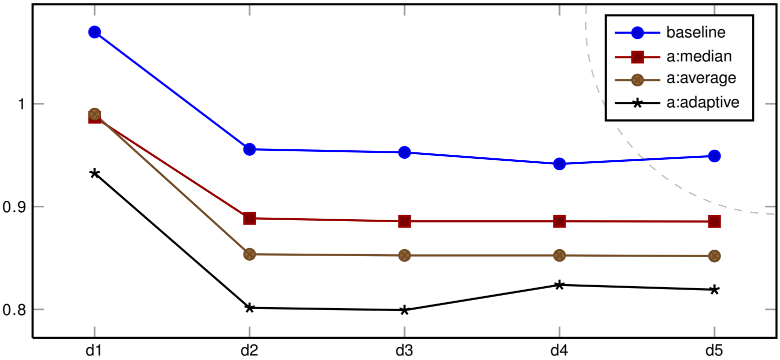

Figure: RMSE Variations: This plot shows that, while the standard deviation of each method may be high, this has more to do with the selected dataset than with their performance in comparison with each other.

Limitations

- Lots of added complexity for fairly unknown improvement.
- Only tested on a few datasets, no real world situations.
- Only compared to simple aggregation methods.
- Neither the aggregators nor the basic recommenders were heavily optimized to the domain of the dataset.

NTNU
Norwegian University of
Science and Technology

Adaptive Recommenders

- Combine disjoint recommenders
- Weight recommenders by predicted accuracy.
- Accuracy predictions are contextually dependent on (u, i, m).
- *Any* applicable recommender becomes a worthy addition.
- See paper for references and more results.

NTNU
Norwegian University of
Science and Technology