MOVIE RECOMMENDER SYSTEM

BRBSB BARUWATTA

16/ENG/016

Content

Problem

Background & Aim

Technology

Design & Implementation

Results

Challenges

Future improvements

Q & A

Problem

- Consumers Have many wants
- Choosing the most suitable item???
- Products are increasing rapidly
- Also problem in movie industry

Solution

Recommender System

Movie Recommender System

Background and Aim

Popular Videos Streaming and Movie Platforms

Increasing the Accuracy

Best recommender system perform 70% Netflix

"The BellKor Solution to the Netflix Grand Prize"

- Yehuda Koren

Aim

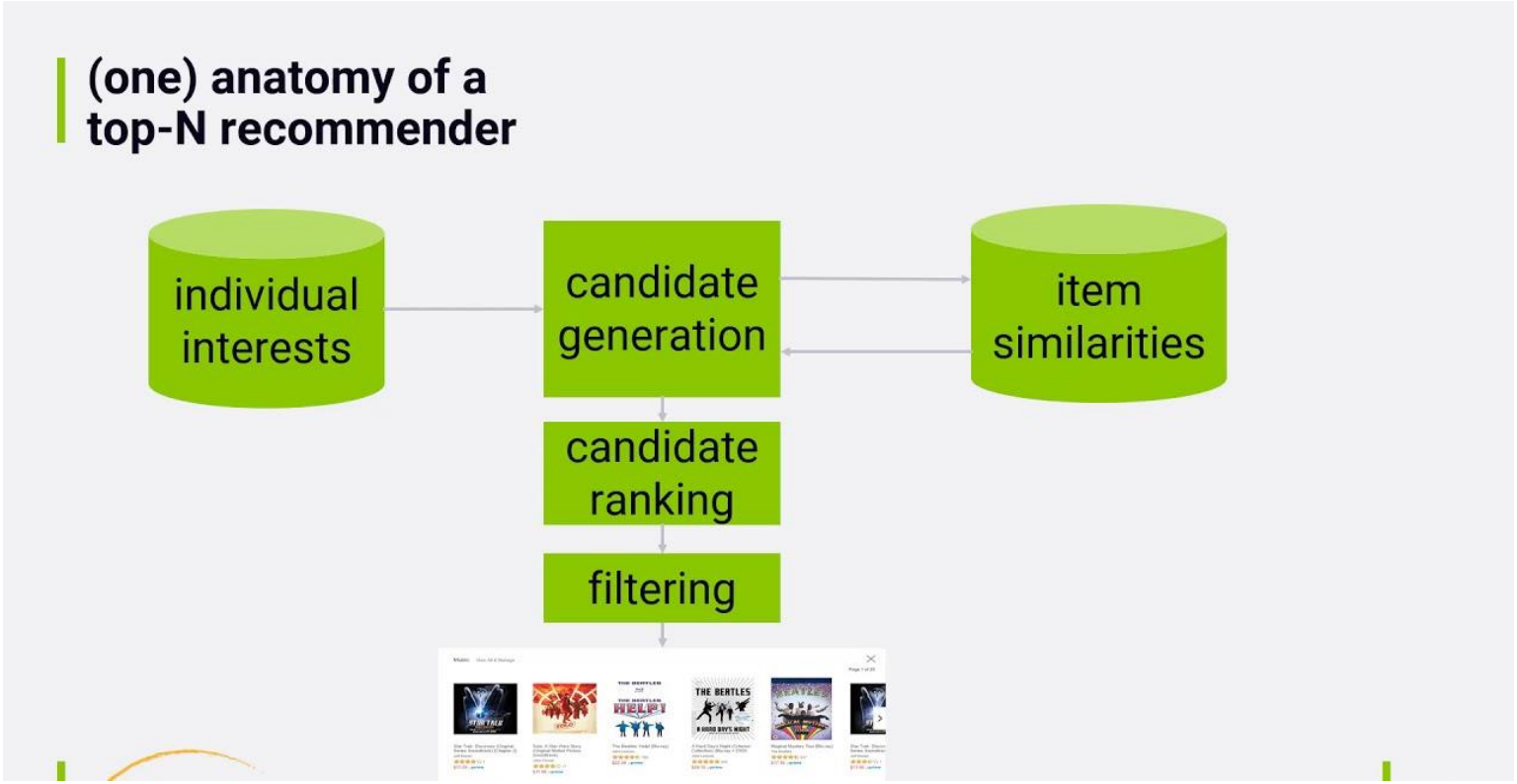
- Build a hybrid recommendation system which perform better

Technologies

- Movielens data set 100k data set
- Anaconda for data analysis
- Surprise Library

Methodology

- TOP N Architecture



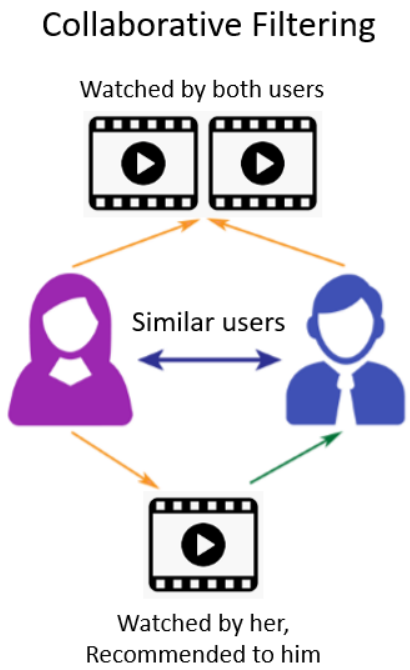
Methodology

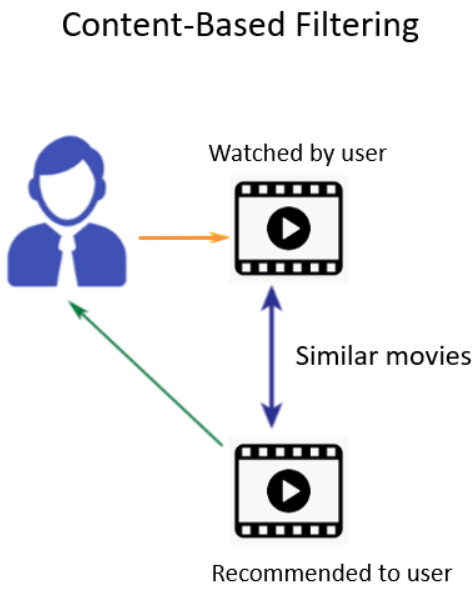
Theory

Find relationship between consumer – item based on action or attributes

Hybrid

Assign Weight Factor





Methodology

BENCHMARKS

Accuracy

MSE

RMSE

Hit Rate = Hits/Users

mean absolute error (MAE)

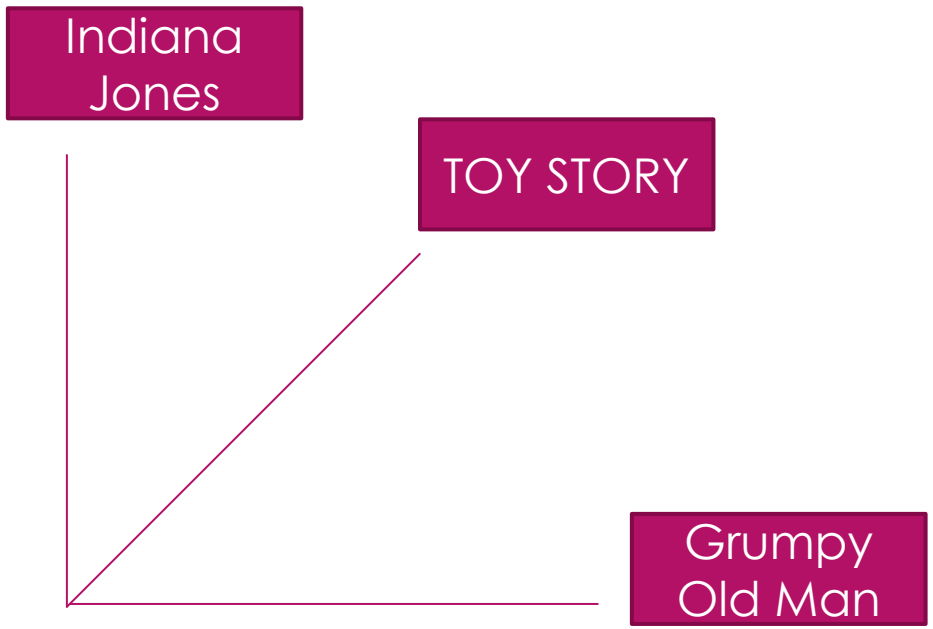
$$\frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

root mean square error (RMSE)

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}}$$

Methodology

- Content Based Filtering KNN



	Comedy	Adventure	Horror	Romance	....	action	scifi
Matrix	0	0	0	0		1	1
Doo little	1	1	0	0		0	0

Similarity Matrixes (Cosine Sim & Exponential Sim)

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Methodology

SVDpp

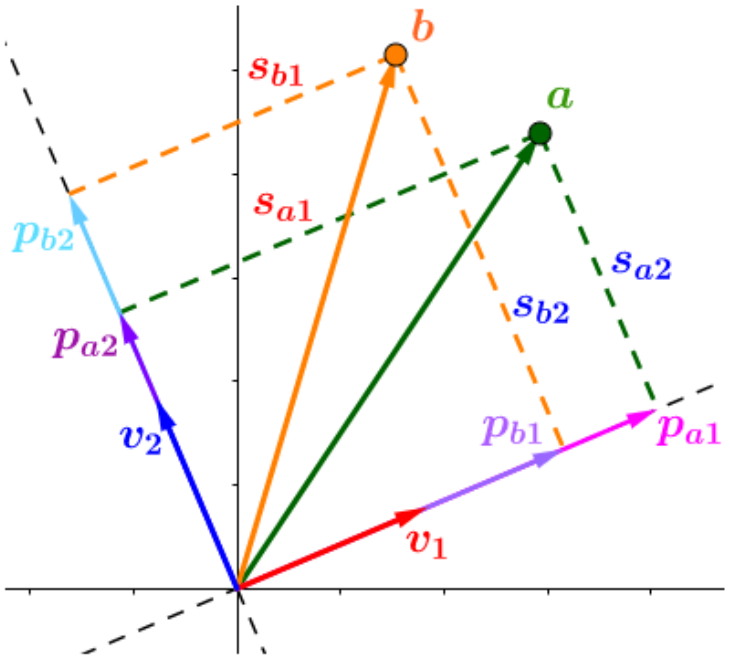
$$A = U \Sigma V^T$$

SVDpp =GV+A

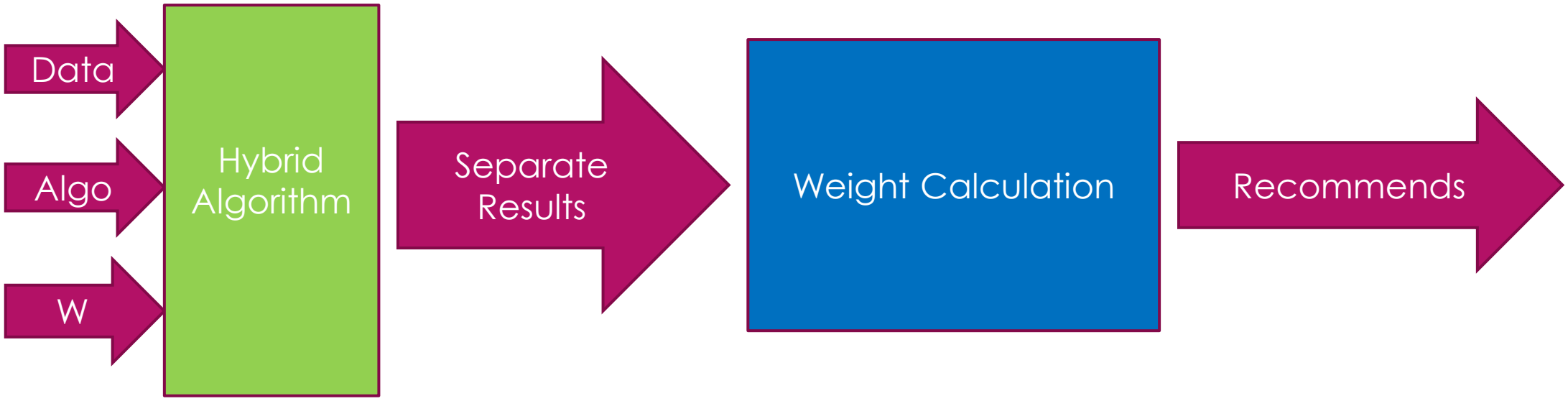
	Bob	Ted	Ann
Indiana Jones	4	3	4
Star Wars	5	3	5
Empire Strikes Back	5	3	5
Incredibles	4	5	5
Casablanca	4	4	2

PCA

	"Action"	"Sci-Fi"	"Classic"
Indiana Jones	0.6	0.3	0.1
Star Wars	0.4	0.6	0
Empire Strikes Back	0.4	0.6	0
Incredibles	0.8	0.2	0
Casablanca	0.2	0	0.8



Methodology



Results

```
Evaluating  Random ...
Evaluating accuracy...
Analysis complete.


Algorithm  RMSE       MAE
ContentKNN 1.0460     0.8198
Random     1.4385     1.1478

Legend:

RMSE:      Root Mean Squared Error. Lower values mean better
accuracy.
MAE:       Mean Absolute Error. Lower values mean better
accuracy.
```

```
...done.
Analysis complete.


Algorithm  RMSE       MAE
SVDpp      0.9002     0.6958
ContentKNN 0.9375     0.7263
Hybrid     0.8998     0.6961

Legend:
```

Results

```
Computing recommendations...

We recommend:
Inception (2010) 3.9710086765133914
Touch of Evil (1958) 3.9349401871607217
Fight Club (1999) 3.885915398141114
French Connection, The (1971) 3.8558946140756425
City of God (Cidade de Deus) (2002) 3.8515050078395734
Star Trek: First Contact (1996) 3.850014218802074
Independence Day (a.k.a. ID4) (1996) 3.833685147457443
Indiana Jones and the Last Crusade (1989) 3.824705292208889
Heat (1995) 3.8064215431021307
True Grit (2010) 3.796431973892063
```

Hybrid Algorithm

```
...done.
Computing recommendations...

We recommend:
Presidio, The (1988) 3.841314676872932
Femme Nikita, La (Nikita) (1990) 3.839613347087336
Wyatt Earp (1994) 3.8125061475551796
Shooter, The (1997) 3.8125061475551796
Bad Girls (1994) 3.8125061475551796
The Hateful Eight (2015) 3.812506147555179
True Grit (2010) 3.812506147555179
Open Range (2003) 3.812506147555179
Big Easy, The (1987) 3.7835412549266985
Point Break (1991) 3.764158410102279
```

KNN Algorithm

Results

Algorithm	MAE	RMSE	Bench(RMSE)
KNNBasic Cont	0.8198	1.046	1.154
KNN tuned Cont	0.7263	0.9375	0.952
SVDpp	0.6958	0.9002	0.926
Random	1.1478	1.4385	1.501
Hybrid	0.6961	0.8998	-

Challenges

- Scaling Up the System

100k to 20 million

- Predict Rating problem
- Success is based on real test
- Resources
- SVDpp Need to modify

Future Works

- Test Hybrid Method with different Algorithms
- Increase the number of algorithms for hybrid system

Q & A