Recommender systems
Part 1
The formulation of the problem

Victor Kantor

First view

- There is a list of users and a list of items (products, movies, songs)
- We have a feedback from users (ratings of items, clicks, purchases, likes or dislikes)
- We need to recommend every user items he would like

Movies recommendations: possible problem formalization

- There are ratings that have been chosen by user for movies that he have already watched
- We need to:
  - Predict ratings, that could be chosen by user for other movies
  - Recommend movies that user will like more (according to our predictions)

Movies recommendations

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	
Vladimir			3	5
Nikolay	3		4	5
Peter				4
Ivan		5	3	3

Movies recommendations

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	?
Vladimir			3	5
Nikolay	3		4	5
Peter				4
Ivan		5	3	3

User-based kNN

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	
Vladimir			3	5
Nickolay	3		4	5
Peter				4
Ivan		5	3	3

User-based kNN

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	?
Vladimir			3	5
Nikolay	3		4	5
Peter				4
Ivan		5	3	3

Item-based kNN

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	?
Vladimir			3	5
Nikolay	3		4	5
Peter				4
Ivan		5	3	3

Item-based kNN

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia		5	2	?
Vladimir			3	5
Nikolay	3		4	5
Peter				4
Ivan		5	3	3

Matrix factorizations

j

	Saw	Nightmare on Elm Street	Vanilla Sky	The Intouchables
Maria	5	4	1	2
Julia	5	5	2	
Vladimir			3	5
Nikolay	3	?	4	5
Peter				4
Ivan		5	3	3

i

u_i - "user interests"
v_j - "movies parameters"

$$x_{ij} \approx \langle u_i, v_j \rangle = \sum_{k=1}^{K} u_{ik} v_{jk}$$

Matrix factorizations: fitting model

$$x_{ij} \approx \langle u_i, v_j \rangle$$

$$\sum_{i,j} \left(\langle u_i, v_j \rangle - x_{ij} \right)^2 \to min$$

Measuring the quality

Is the recommendations quality == rating predicting quality?

Variants:

- Root mean square error (RMSE) of ratings predictions
- Mean absolute error (MAE) of ratings predictions

Do we measure quality in a right way?

- We are measuring: the quality of ratings predictions
- What should be measured: the quality of recommendations

Recall@k

Recommended items
Blue T-shirt
Red T-shirt
Boots
Cap
Green T-shirt

Purchased items
Red T-shirt
Boots
Cap

k – count of the recommended items

$$\text{Recall@k} = \frac{purchased\ from\ recommended}{purchased\ items}$$

AverageRecall@k - average in user sessions Recall@k

Precision@k

Recommended items
Blue T-shirt
Red T-shirt
Boots
Cap
Green T-shirt

Purchased items
Red T-shirt
Boots
Cap

k – count of the recommended items

$$\text{Precision@k} = \frac{purchased\ from\ recommended}{k}$$

AveragePrecision@k - average in user sessions Precision@k

Recommendations in retail

j

i

	Dress	Boots	Jeans	T-shirt
Maria	1		1	
Julia	1	1		1
Vladimir		1	1	
Nikolay	1	?	1	
Peter		1	1	
Ivan			1	1

Difference from movie recommendations

- No negative examples
- Simple connection with revenue

Possible solution

- Predict which items will be purchased by user
- Maximize the revenue

Maximizing income

Item 1	**Item 2**	**Item 3**	**Item 4**

Maximizing income

	Item 1	**Item 2**	**Item 3**	**Item 4**
Probability:	p_1	p_2	p_3	p_4
Price:	c_1	c_2	c_3	c_4

Maximizing income

Puma
Ветровка
3 490 руб.

Crocs
Сланцы
1 990 руб.

Tony-p
Слипоны
~~1 999 руб.~~ 1 590 руб.

Champion
Брюки спортивные
~~3 599 руб.~~ 1 970 руб.

Probability:	0.05	0.02	0.015	0.009
Price:	3490	1990	1590	1970

Maximizing revenue

Puma
Ветровка
3 490 руб.

Crocs
Сланцы
1 990 руб.

Tony-p
Слипоны
~~1 999 руб.~~ 1 590 руб.

Champion
Брюки спортивные
~~3 599 руб.~~ 1 970 руб.

Probability:	0.05	0.02	0.015	0.009
Price:	3490	1990	1590	1970
Marginality:	0.1	0.4	0.4	0.2

Predicting probabilities

- Examples: tuples (user, item, timestamp)
- Classes: 1 – item will be purchased by user in this session, 0 – item won't be purchased
- Features: user parameters, item parameters, timestamp parameters and interactions of these parameters

Negative samples

- Add all other items from catalogue as negative examples for every positive example (unreal)
- Random with uniform distribution
- Random with probabilities proportional to items popularities
- Most popular and not purchased
- Recommendations from other algorithm (not purchased)

Candidates selection

- Popular items
- Popular items from already seen categories
- Items that have high PMI (Pointwise Mutual Information) with already seen items
- Items from custom candidates list

PMI

$$\mathrm{pmi}(x;y) \equiv \log \frac{p(x,y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}$$

Example of items with high PMI in user sessions:
Laptop + Mouse

Examples with medium PMI:
T-shirt + Boots
Dress + Handbag
Dress + Shoes

Example with low PMI:
Shoes + Mouse

Online metrics

The quality on the historical data is high

What can we say about the quality of recommendations in production?

Online metrics

Допустим, на исторических данных качество алгоритма высокое, а будет ли оно высоким в реальности?

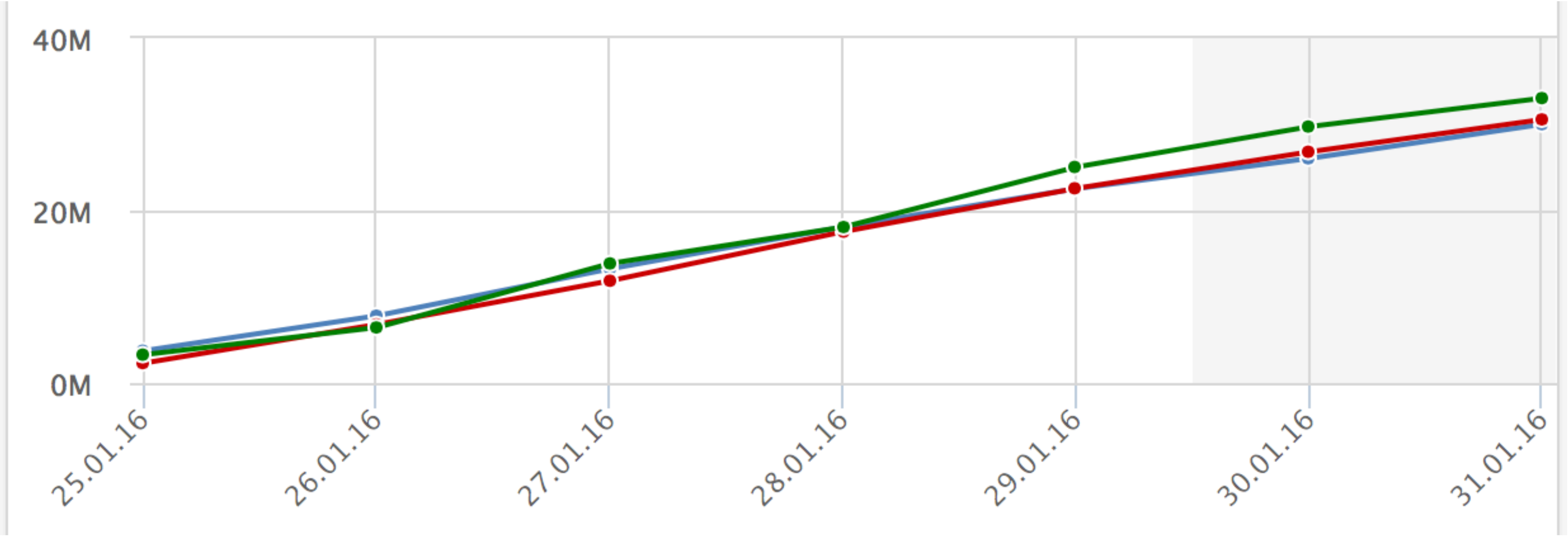
Ideas:

1. A/B test
2. Statistical significance test

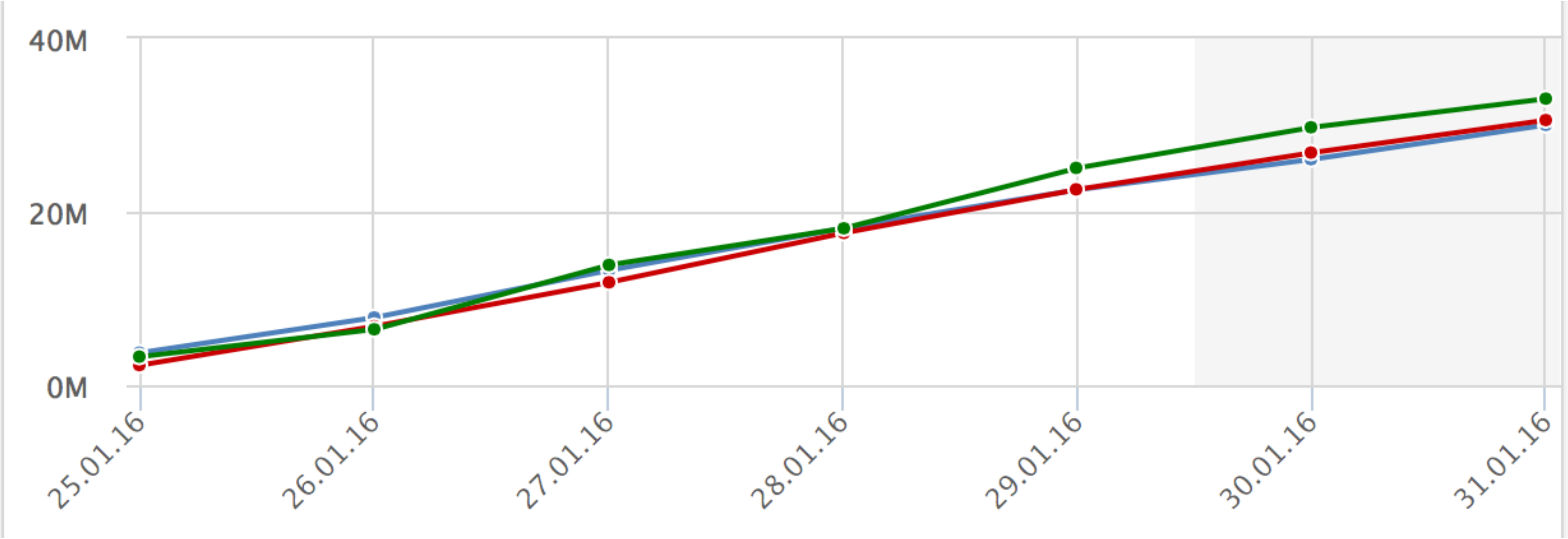
A/B test

1. Split users randomly on two groups
2. Measure online metrics (for example, purchases or revenue) in both groups for rather long period of time
3. Result is two numbers (f. ex. 2.3 M $ in group A and 2.2 M $ in group B)
4. What decision can we make?

Statistical significance: example



Statistical significance: example

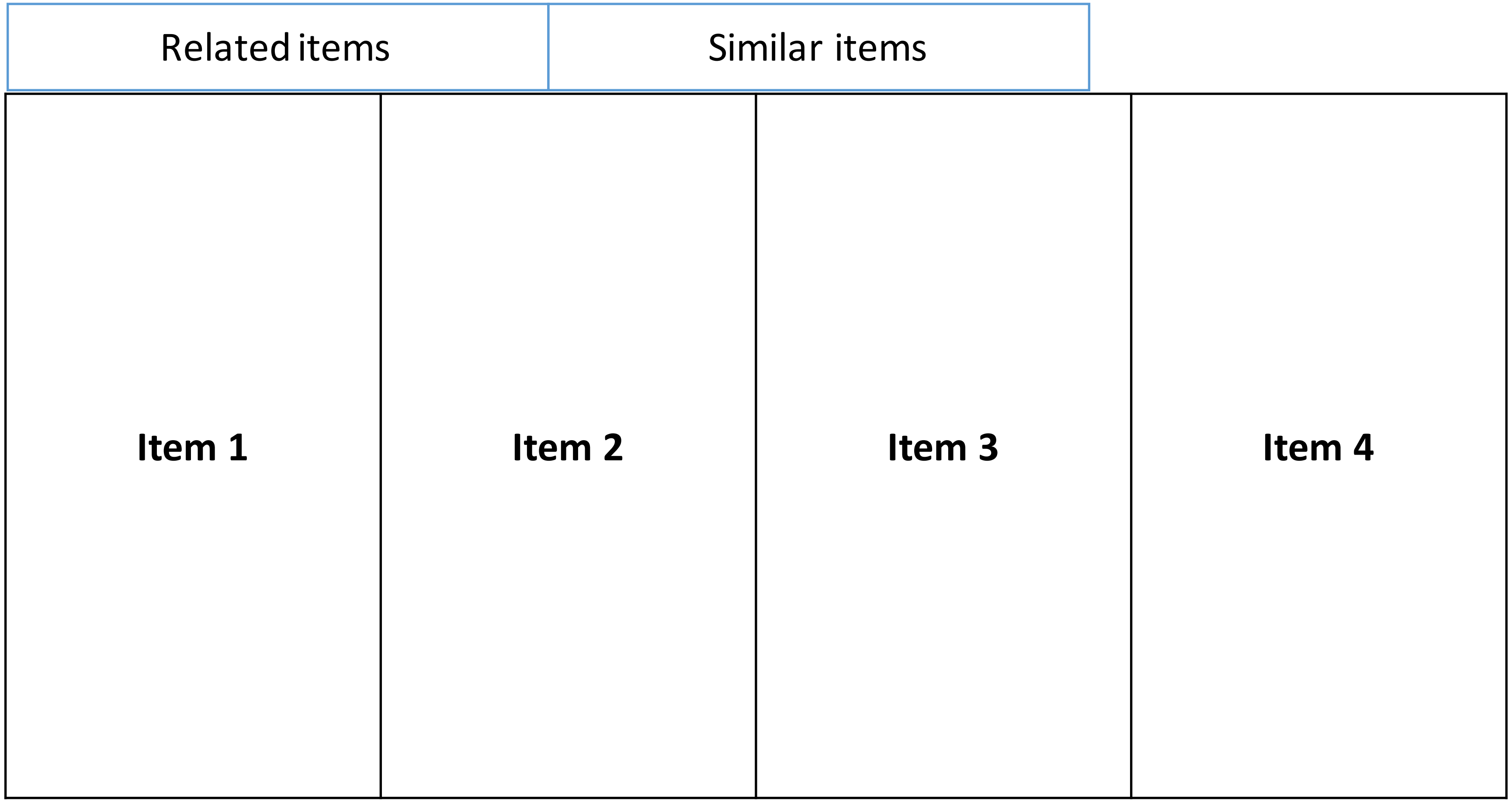


This plot is for random split on three groups

Story 1: bad split in A/B test

- Proposed variant:
  - Group = hash(user_id) % 2
- Released:
  - Group = hash(user_id + user_email) % 2

Story 2: design



Story 3: comparing two solutions

- Compared their solution with recommendations developed in another company
- Offline quality was exactly the same
- Decided not to use this recommender engine
- Some months later discovered the reason of such result

Recap

1. Don’t solve problem until the formulation of the problem is clear
2. Problem formalization should be connected with economics
3. Problem formalization includes the procedure of offline and online quality measuring
4. Good formalization and good model is less valuable than not doing nonsense