arXiv.org > cs > arXiv:1901.07829

Computer Science > Computation and Language

AspeRa: Aspect-based Rating Prediction Model

Sergey I. Nikolenko, Elena Tutubalina, Valentin Malykh, Ilya Shenbin, Anton Alekseev

(our paper accepted to ECIR2019 Cologne)

St. Petersburg Department of Steklov Mathematical Institute of Russian Academy of Sciences

NLP seminar @ NRU ITMO, St Petersburg
Feb 16, 2019

(a total of 33 slides)

NLP gang

Sergey Nikolenko
Laboratory Head

Elena Tutubalina
Researcher

Valentin Malykh
Researcher

Anton Alekseev
Researcher

Ilya Shenbin
Researcher

Andrey Savchenko
Senior Researcher

Alexandr Rassadin
Researcher

Ekaterina Arkhangelskaya
Researcher

Ivan Grechikhin
Researcher

Who are we?

Joint Samsung-PDMI AI Center
(Artificial Intelligence Laboratory,
Steklov Mathematical Institute
St Petersburg)

A distributed team of researchers working on a pool of different tasks requiring the usage of neural methods

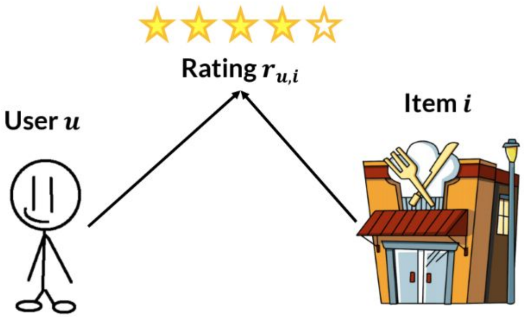
for questions regarding possible collaboration please contact Sergey
snikolenko@gmail.com

Plan

1. Recommender systems
2. Aspects extraction
3. Aspect-based recommendations: AspeRa
4. Discussion and future work directions

Recommender systems: task

(may I please skip the explanation what recommender systems are for?)



Goal: based on users' feedback (of any kind!) on items, suggesting users new items

Usually are roughly divided into

- **content-based** ones (determining what kind of content the user likes)
- those utilizing **collaborative filtering** (predicting preferences based on similarity to other users)

Img src: https://drive.google.com/file/d/1ILrOhHO8K9V-euVg0Ur2xtsKYhRPNyfb/view

Recommender systems: task

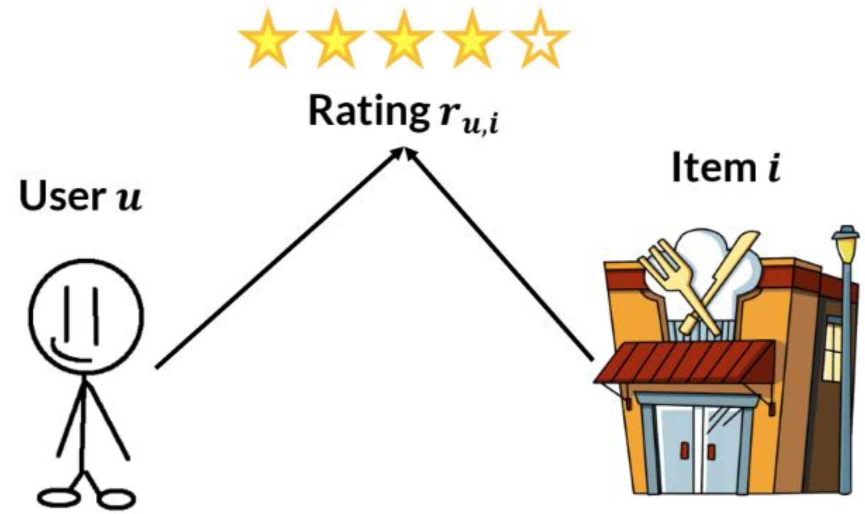
Explicit feedback matrix of ratings

$R_{NxM} = \{ r_{u,i} \text{ in } 0..5 \}$

value 1..5 if the user **u**
has interacted with the item **i**

0 -- if one has’t

Task: predict new items the user
Is supposed to rate higher

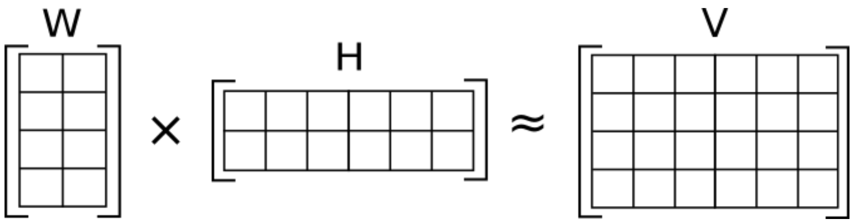


Img src:: https://drive.google.com/file/d/1lLrOhHO8K9V-euVg0Ur2xtsKYhRPNyfb/view

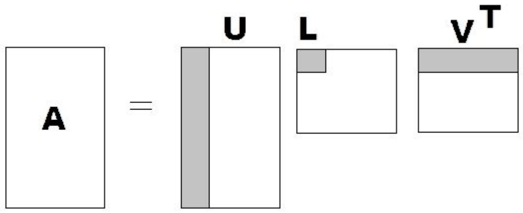
Recommender systems: traditional CF approach

Methods that are well-known and widely used in industry

NMF
nonnegative matrix factorization



SVD
singular value decomposition



...And all other kinds and shapes of matrix approximation ways

Text-aware recommender systems: why?

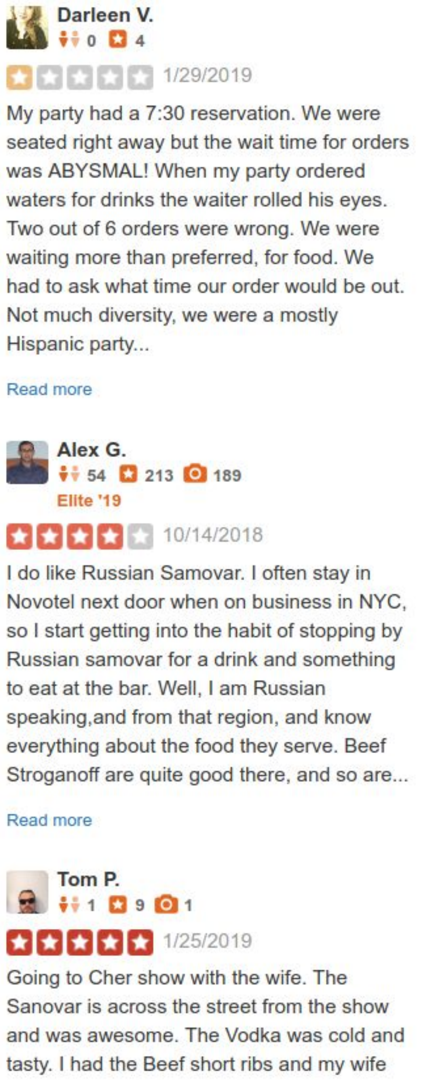
...more useful data.

A common situation: users are asked to write reviews as their response becomes more trustworthy and informative for other users

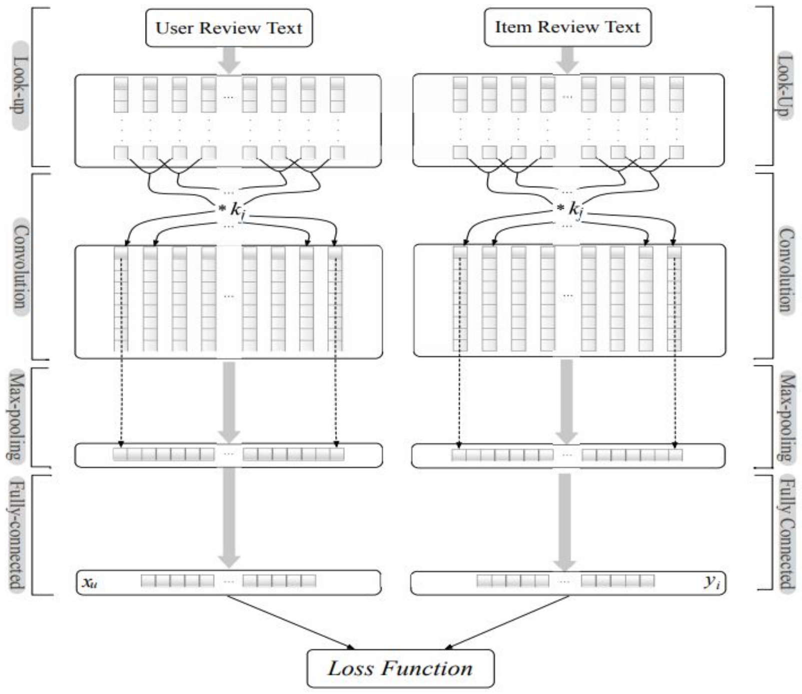
User-item interaction becomes:

$$(u, i, r_{u,i}, d_{u,i})$$

Img src:: https://drive.google.com/file/d/1ILrOhHO8K9V-euVg0Ur2xtsKYhRPNyfb/view



Baseline example: DeepCoNN



Lei Zheng, Vahid Noroozi, and Philip S Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In WSDM. ACM, 425-434.

Concatenation of the word vectors for all the concatenated reviews

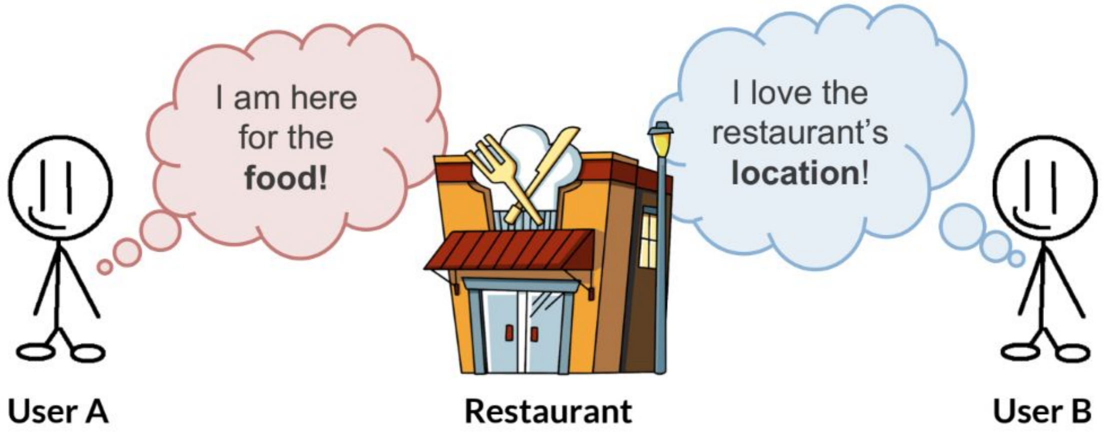
- written by the user in concern
- devoted to the item in concern

and a standard two-tower architecture

Aspect-based recommender systems

Users value **different features (aspects) in items**, and have different opinion each of aspects of one item, so taking this into account may

- improve predictions
- help analyze items' most/least valued features or users' preferences



Img src:: https://drive.google.com/file/d/1ILrOhHO8K9V-euVg0Ur2xtsKYhRPNyfb/view

Room for improvement

- text-based systems are limited in terms of interpretability of predictions
- aspect-based systems often rely on external aspects extraction instruments, which is the cause of the quality limits

Plan

1. ~~Recommender systems~~
2. Aspects extraction
3. Aspect-based recommendations: AspeRa
4. Discussion and future work directions

Aspects Extraction

the goal = to extract **entity aspects on which opinions have been expressed**

In practice may be is **rule-based/supervised**
BUT: suffers from domain adaptation

Hence, **unsupervised** approaches are to be worked on; usually -- all kinds and flavours of **LDA, BTM**, etc.
BUT: reviews are short, ‘extra efforts’ on estimating the document distribution, could try something else

#	sent.	sentiment words
1	neu	соус [sauce], салат [salad], кусочек [slice], сыр [cheese], тарелка [plate], овощ [vegetable], масло [oil], лук [onions], перец [pepper]
	pos	приятный [pleasant], атмосфера [atmosphere], уютный [cozy], вечер [evening], музыка [music], ужин [dinner], романтический [romantic]
	neg	ресторан [restaurant], официант [waiter], внимание [attention], сервис [service], обращать [to notice], обслуживать [to serve], уровень [level]
2	neu	столик [table], заказывать [to order], вечер [evening], стол [table], приходить [to come], место [place], заранее [in advance], встречать [to meet]
	pos	место [place], хороший [good], вкус [taste], самый [most], приятный [pleasant], вполне [quite], отличный [excellent], интересный [interesting]
	neg	еда [food], вообще [in general], никакой [none], заказывать [to order], оказываться [appear], вкус [taste], ужасный [awful], ничто [nothing]
3	neu	девушка [girl], спрашивать [to ask], вопрос [question], подходить [to come], официантка [waitress], официант [waiter], говорить [to speak]
	pos	большой [big], место [place], выбор [choice], хороший [good], блюдо [dish], цена [price], порция [portion], небольшой [small], плюс [plus]
	neg	цена [price], обслуживание [service], качество [quality], уровень [level], кухня [kitten], средний [average], ценник [price tag], высоко [high]

img src: https://www.slideshare.net/aistconf/sergey-nikolenko-and-elena-tutubalina-constructing-aspectbased-sentiment-lexicons-with-topic-modeling

An Unsupervised Neural Attention Model for Aspect Extraction

ACL 2017

Ruidan He[†‡], Wee Sun Lee[†], Hwee Tou Ng[†], and Daniel Dahlmeier[‡]

ABAE, a simple yet effective neural method, works as an autoencoder:

1. each word is assigned **a word embedding** (e.g. skip-grams)
2. words in a sentence are **summed weighted with attention** (hopefully down-weighting the non-aspect words)
3. the resulting sentence embedding is transformed with **one FF layer**
4. ...and then that sentence embedding is **recovered with aspects matrix** T

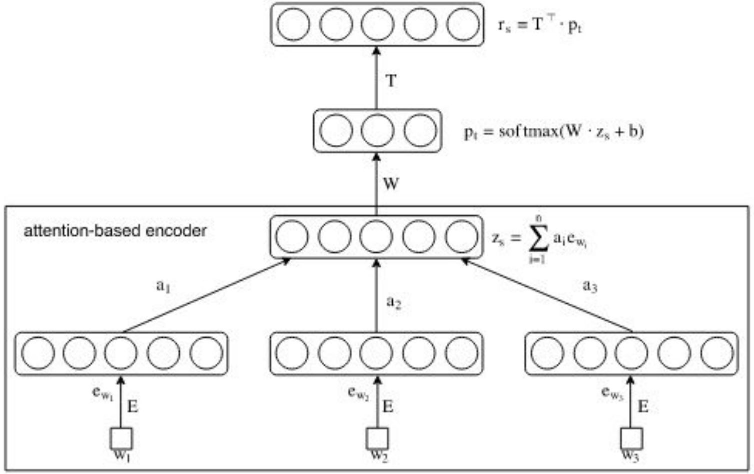


Figure 1: An example of the ABAE structure.

An Unsupervised Neural Attention Model for Aspect Extraction

DETAILS

ACL 2017

Ruidan He[†‡], Wee Sun Lee[†], Hwee Tou Ng[†], and Daniel Dahlmeier[‡]

Attention over words e_w is 'paid' to the mean word embedding y_s

$$a_i = \frac{\exp(d_i)}{\sum_{j=1}^{n} \exp(d_j)}$$

$$d_i = \mathbf{e}_{w_i}^{\top} \cdot \mathbf{M} \cdot \mathbf{y}_s$$

$$\mathbf{y}_s = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_{w_i}$$

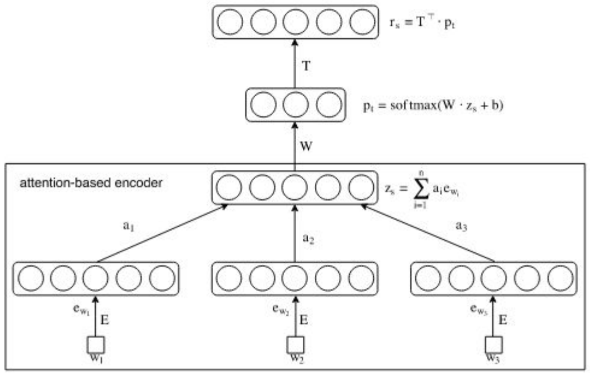


Figure 1: An example of the ABAE structure.

$T_n T_n^T$ is forced to be **close to orthogonal** to penalize redundancy

$$U(\theta) = \|\mathbf{T}_n \cdot \mathbf{T}_n^{\top} - \mathbf{I}\|$$

Main recovery loss (max-margin)

$$J(\theta) = \sum_{s \in D} \sum_{i=1}^{m} \max(0, 1 - \mathbf{r}_s \mathbf{z}_s + \mathbf{r}_s \mathbf{n}_i)$$

negative sample

recovered sentence embedding

original sentence embedding

EXAMPLES

An Unsupervised Neural Attention Model for Aspect Extraction

ACL 2017

Ruidan He[†‡], **Wee Sun Lee**[†], **Hwee Tou Ng**[†], and **Daniel Dahlmeier**[‡]

Inferred Aspects	**Representative Words**	**Gold Aspects**
Main Dishes	beef, duck, pork, mahi, filet, veal	
Dessert	gelato, banana, caramel, cheesecake, pudding, vanilla	
Drink	bottle, selection, cocktail, beverage, pinot, sangria	Food
Ingredient	cucumber, scallion, smothered, stewed, chilli, cheddar	
General	cooking, homestyle, traditional, cuisine, authentic, freshness	
Physical Ambience	wall, lighting, ceiling, wood, lounge, floor	Ambience
Adjectives	intimate, comfy, spacious, modern, relaxing, chic	
Staff	waitstaff, server, staff, waitress, bartender, waiter	Staff
Service	unprofessional, response, condescending, aggressive, behavior, rudeness	
Price	charge, paid, bill, reservation, came, dollar	Price
Anecdotes	celebrate, anniversary, wife, fiance, recently, wedding	Anecdotes
Location	park, street, village, avenue, manhattan, brooklyn	
General	excellent, great, enjoyed, best, wonderful, fantastic	Misc.
Other	aged, reward, white, maison, mediocrity, principle	

An Unsupervised Neural Attention Model for Aspect Extraction

ACL 2017

Ruidan He[†‡], Wee Sun Lee[†], Hwee Tou Ng[†], and Daniel Dahlmeier[‡]

On prediction stage,

1) the **aspects rates** for the text are obtained as outputs of the FF layer with softmax

2) the words describing aspects (please see the prev. slide) are obtained as nearest neighbours to aspects embeddings from the **T** matrix

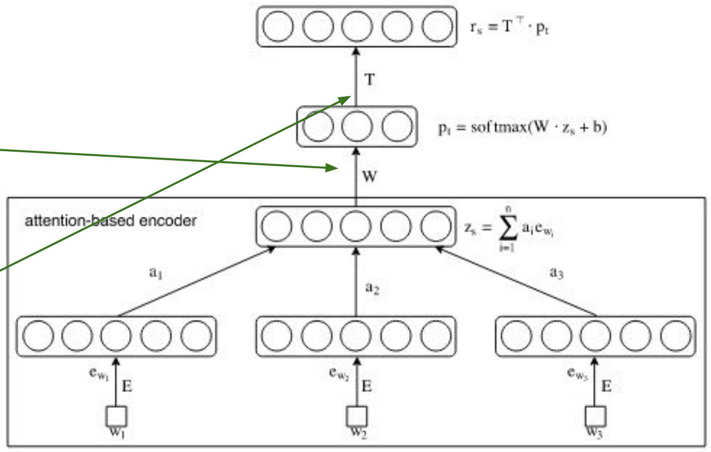


Figure 1: An example of the ABAE structure.

Discussion

Findings from experience:

- stopwords filtering is important
- custom word2vec vectors perform better in terms of interpretability
- couldn’t make work on non-review data (larger texts where aspects may not be present, e.g. news, fashion reviews)
- good aspects matrix initialization **is extremely important**

5	'wordprocessing blogging picasa blackboard sketchup youtubing ev
6	'kardon harman altec sound harmon lansing speaker treble audiop
7	'price clearance officemax mart wal msrp bb reward tax financing'
8	'screen display tint uniform ppi saturated representation colorful viv
9	'lug dorm tote bike carrying travel motorcycle briefcase commuting
10	pant perpendicular bathroom thump nose upwards cringe blanket f
11	'circa xt spectre theasus sb fujitsu weighed hp pavillion thinnest'
12	'redundantly liveupdate inflate deference <unk> exasperated staun
13	'resists grippy faux rubberized rubbery machined velvet alloy foam
14	'headquarters phoned mailed dept authorization kentucky apologiz

Plan

1. ~~Recommender systems~~
2. ~~Aspects extraction~~
3. Aspect-based recommendations: AspeRa
4. Discussion and future work directions

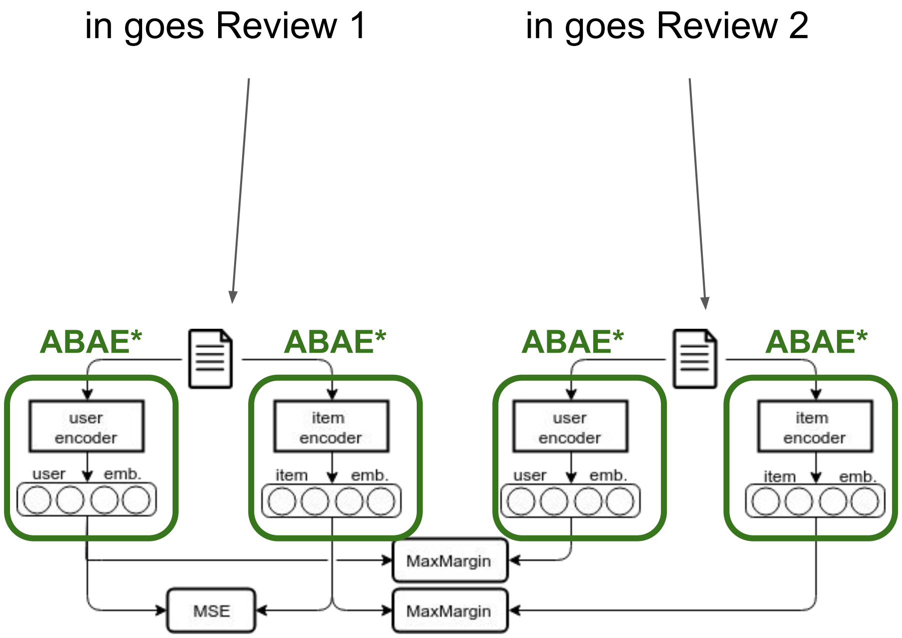
AspeRa: the idea

Two towers, each contains an ABAE-like encoder, which provides reconstruction "**user embeddings**" and "**item embeddings**"

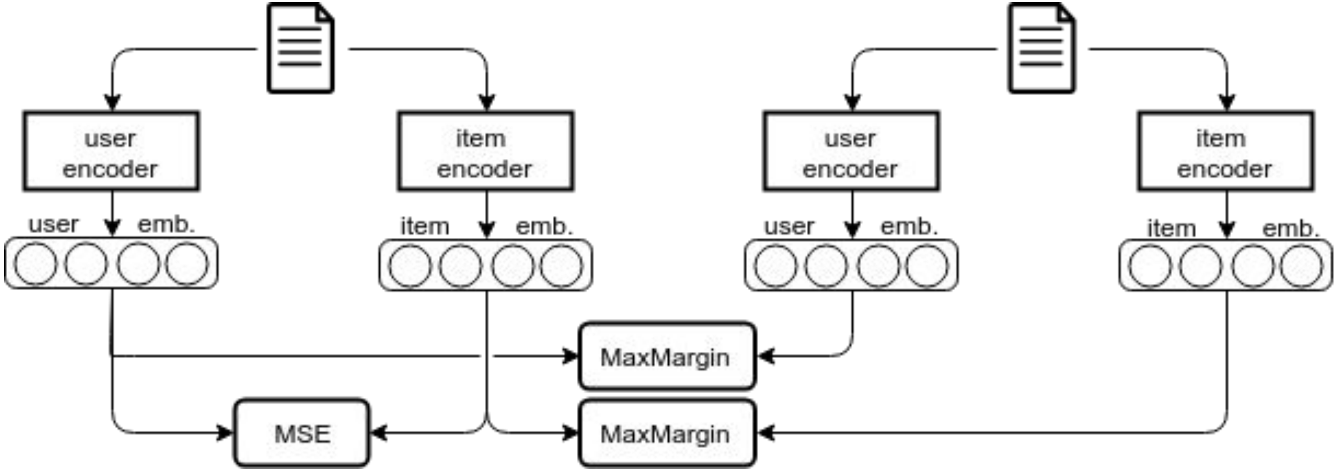
The intuition: there are some features (aspects) important in an item **for a user**, but the item also has **other features**.

The idea was to **separate** item-related and user-related information and to have the possibility to **analyze** users' preferences and items' features

Note: no explicit usage of user ID or item ID



AspeRa: a closer look

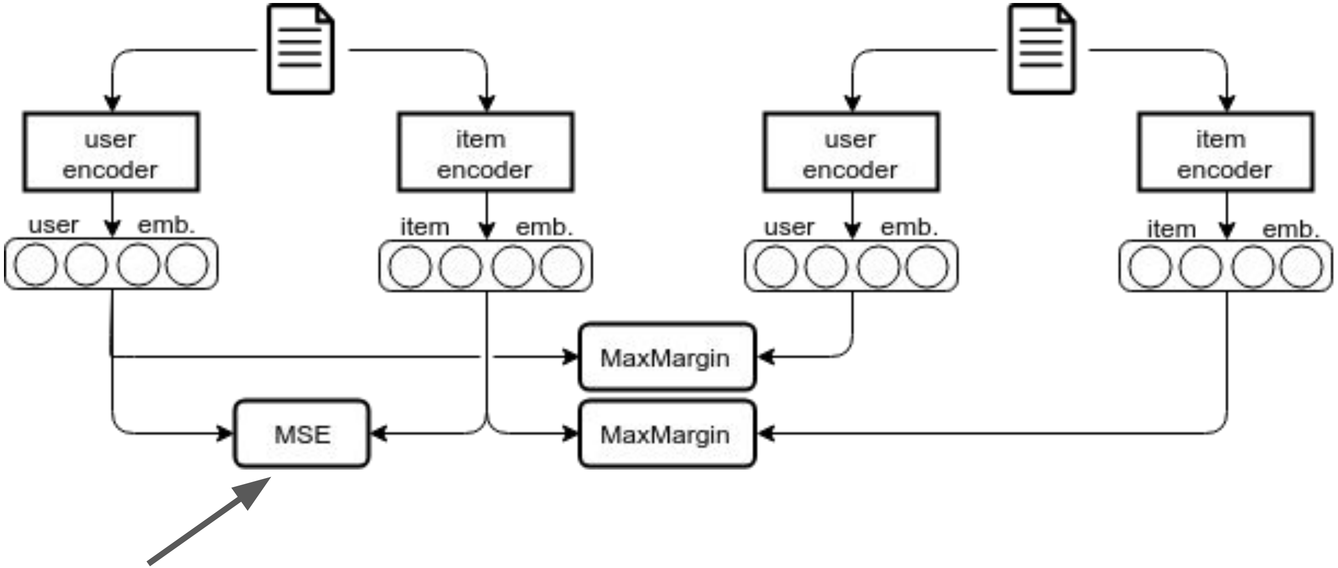


A training sample = two reviews; as in ABAE, each is represented as a set of word embeddings (GloVe/word2vec-SGNS), which are then fed into the ABAE-like encoders

Either those are reviews **from the same user**, **or** both are **about the same item**

But, unlike in ABAE, there is **no explicit negative sampling**

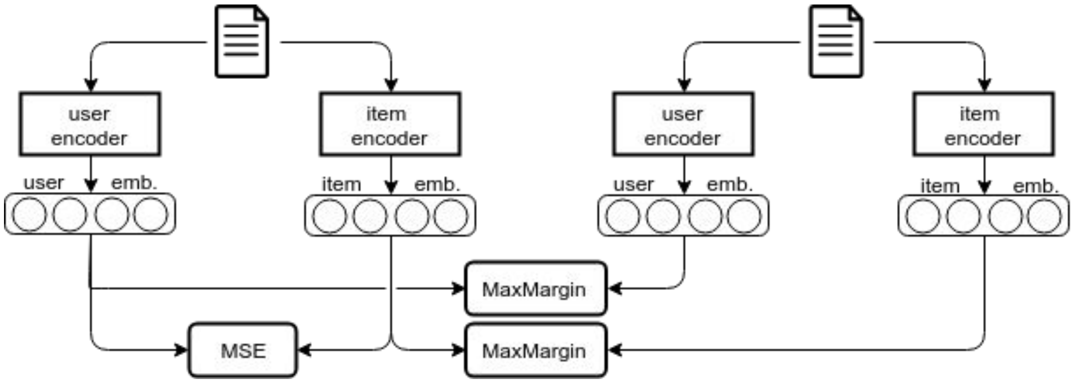
AspeRa: losses (quite a few tbh)



$$MSE = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{z}_j^{u\top}\mathbf{z}_j^{i} - r_j)^2$$

user and item sentence embeddings scalar product for rating prediction, with MSE as loss

AspeRa: losses (quite a few tbh)



$$\mathrm{MaxMargin}(i,j) = \frac{1}{N}\sum_{i,j} \max(0, 1 - {\mathbf{r}_i^u}^\top \mathbf{z}_i^u + {\mathbf{r}_i^u}^\top \mathbf{z}_i^i + {\mathbf{r}_i^u}^\top \mathbf{z}_j^i)$$

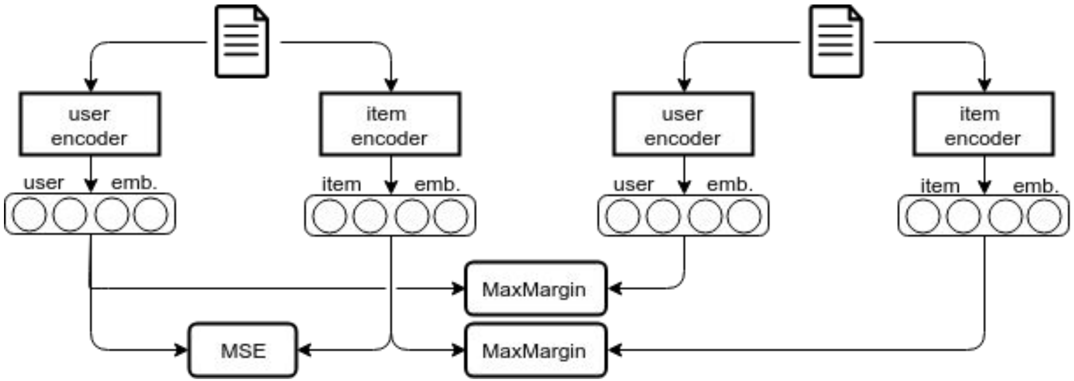
(1) **(2)**

instead of negative sampling as in ABAE

for **every user**: we push

1) reconstructed and original sentence embeddings to be closer for each user ***i***
2) original sentence embeddings for both considered items away from the user's reconstructed embedding

AspeRa: losses (quite a few tbh)



$$\mathrm{MaxMargin}(i,j) = \frac{1}{N}\sum_{i,j}\max(0, 1 - \mathbf{z}_i^{u\top}\mathbf{z}_j^{u} + \mathbf{z}_i^{u\top}\mathbf{z}_i^{i} + \mathbf{z}_i^{u\top}\mathbf{z}_j^{i})$$

(1) (2)

the case of **the same user**: we push

1) original sentence embeddings to be closer for each user ***i***
2) original sentence embeddings for both considered items away from the user

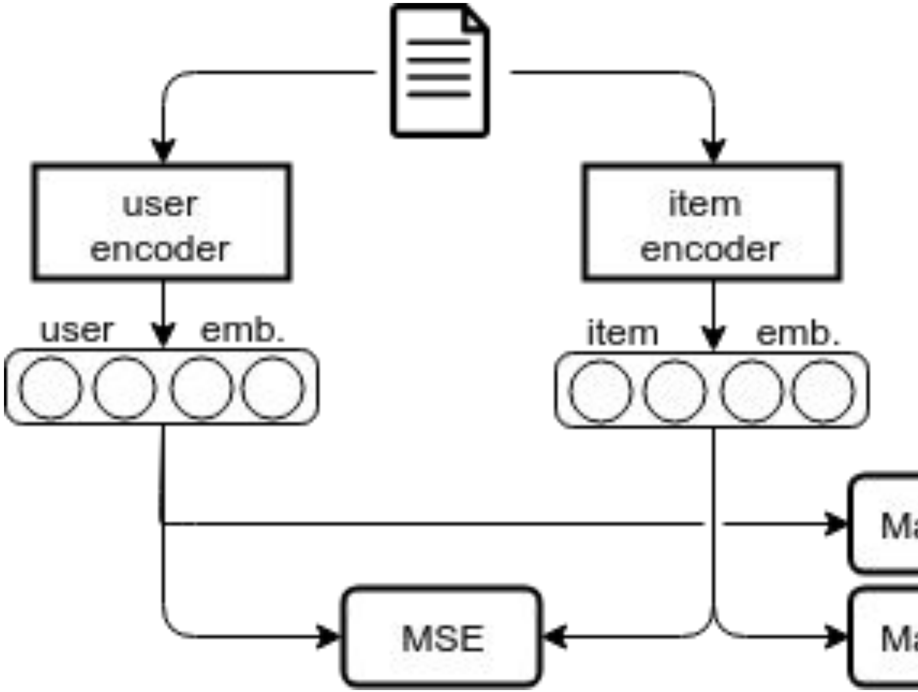
AspeRa: losses (quite a few tbh)

… + almost exactly the same **max-margin losses** of the like for **items**, **not users** in focus

AspeRa: prediction

For prediction just one tower is used, the one with the MSE loss: rating is predicted as a scalar product of a 'user embedding' and an 'item embedding'

Review text IN



Rating value OUT

AspeRa: rating prediction evaluation

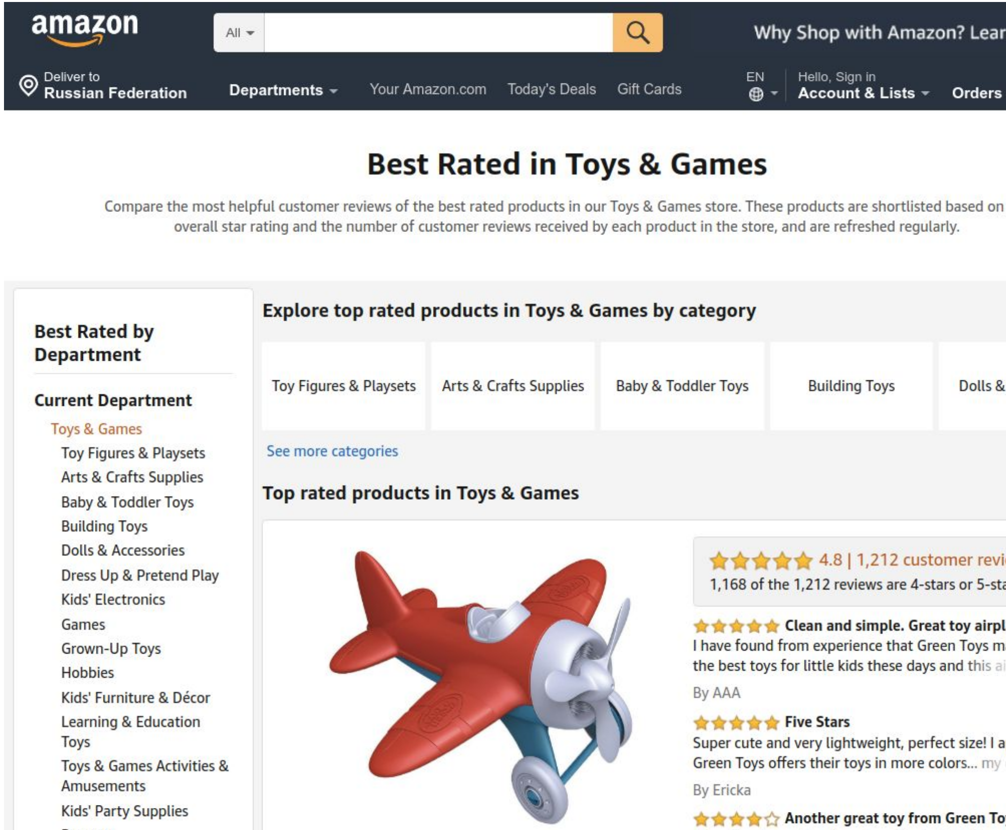
Amazon Instant Videos 5-core reviews

- 37,126 reviews,
- 5,130 users,
- 1,685 items,
- 3,454,453 non-unique tokens.

Amazon Toys and Games 5-core reviews

- 167,597 reviews,
- 19,412 users,
- 11,924 items,
- 17,082,324 non-unique tokens

http://jmcauley.ucsd.edu/data/amazon/



AspeRa: quantitative evaluation

Rating prediction: a document from the test set is fed into the first 'tower', and rating prediction is compared with ground truth

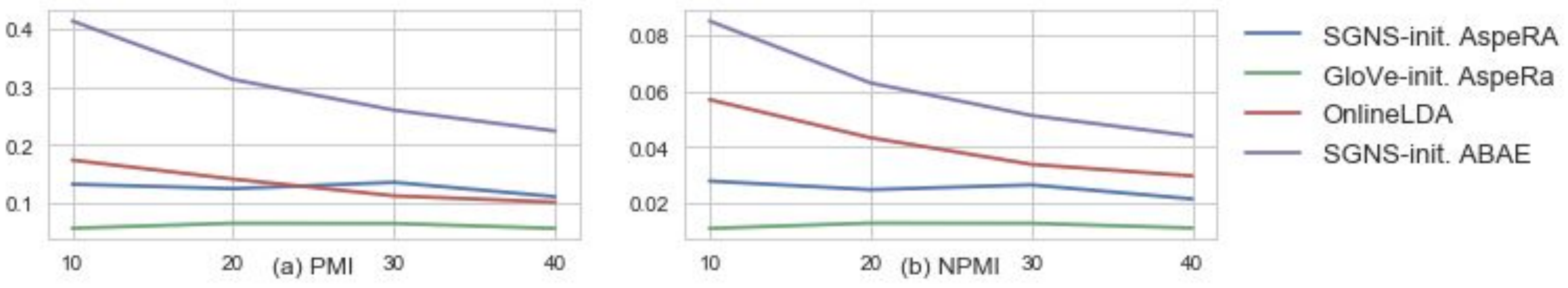
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Model	MSE	
	Instant Videos	Toys & Games
NMF	0.946	0.821
DeepCoNN	0.943	0.851
Attn+CNN	0.936	-
SVD	0.904	0.788
HFT	0.888	0.784
TransRev	0.884	0.784
NARRE	-	0.769
AspeRa (GloVe)	0.870	0.730
AspeRa (SGNS)	**0.660**	**0.571**

Domain-specific custom word vectors have boosted the quality!

img src: https://en.wikipedia.org/wiki/Mean_squared_error

Aspect (topic) coherence quantitative analysis



Standard topic coherence measures on Amazon Instant Videos 5-Core dataset:

- PMI-coherence computed using top N words in topics (see ***x*** axis)

- its normalized modification (NPMI)

$$\mathrm{PMI}(w_i, w_j) = \log \frac{P(w_i, w_j) + \epsilon}{P(w_i)P(w_j)}$$

$$C_{uci} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{PMI}(w_i, w_j)$$

$$\mathrm{NPMI}(w_i, w_j) = \left(\frac{\mathrm{PMI}(w_i, w_j)}{-\log P(w_i, w_j) + \epsilon} \right)^{\gamma}$$

AspeRa: qualitative analysis

#	Aspect words
1	communities governments incidents poverty unity hardships slaves citizens fought
2	coppola guillermo bram kurosawa toro ridley del prolific ti festivals
3	brisk dialouge manipulation snappy plotlines dialogues taunt camerawork muddled
4	sock vegans peanut stifling bats buh ammonium trollstench vegetables pepsi
5	the a and to is of joe's enters that fatal

Based on SGNS

#	Aspect words
1	protein diagnose cell genes brain membrane interacts interact oxygen spinal
2	boost monetary raise introduce measures credit expects increase push demand
3	towel soaked greasy towels cloth dripping tucked crisp coat buckets
4	offbeat comic parody spoof comedic quirky cinematic campy parodies animated
5	sheesh wham whew hurrah oops yikes c'mon shhh oooh och

Based on GloVe

Plan

1. ~~Recommender systems~~
2. ~~Aspects extraction~~
3. ~~Aspect-based recommendations: AspeRa~~
4. Discussion and future work directions

AspeRa: lessons learned so far

- aspect-based recommendations can be effective
- ABAE can serve as an aspects encoder in larger systems
- the choice of word vectors can matter
- interpretability of the aspects in AspeRa is still a challenge

What next?

- comparison with more recent models, such as ANR (in progress!)
- working on aspects quality
- thorough cold-start tolerance analysis
- cross-domain text-based recommendations

etc.

arXiv.org > cs > arXiv:1901.07829

Computer Science > Computation and Language

ECIR2019 Cologne

AspeRa: Aspect-based Rating Prediction Model

Sergey I. Nikolenko, Elena Tutubalina, Valentin Malykh, Ilya Shenbin, Anton Alekseev

Thank you for your attention!

link to the paper preprint

val@maly.hk
Valentin
AspeRa idea author,
main contributor

anton.m.alexeyev@gmail.com
Anton
contributor responsible
for evaluation