Sistemas de Recommendación usando Deep Neural Networks

Denis Parra

IIC3633 PUC Chile

CuratorNet: Visually-aware Recommendation of Art Images

Manuel Cartagena*, Master student

Pablo Messina *, PhD student

Felipe del Río *, PhD student

Patricio Cerda *, PhD student

Denis Parra *, Associate Professor

*CS Department, IA Lab, Pontificia Universidad Católica de Chile

dparras@uc.cl

denisparra

ComplexRec workshop 2020
co-located at ACM RecSys

Thanks to all the people involved in this reseach

Pablo Messina

Manuel Cartagena

Felipe del Río

Patricio Cerda

Denis Parra

Agenda

- **Introduction & motivation**
- Research Problem
- Related Work
- Materials
- Inspiration methods (VBPR, YouTube)
- CuratorNet model
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

The Online Artwork Market

- Online artwork market: Growing since 2008, despite global crises!

- In 2011, art received $11.57 billion in total global annual revenue, over $2 billion versus 2010 (*forbes). In 2018, it grew by 9.8% compared to 2017.

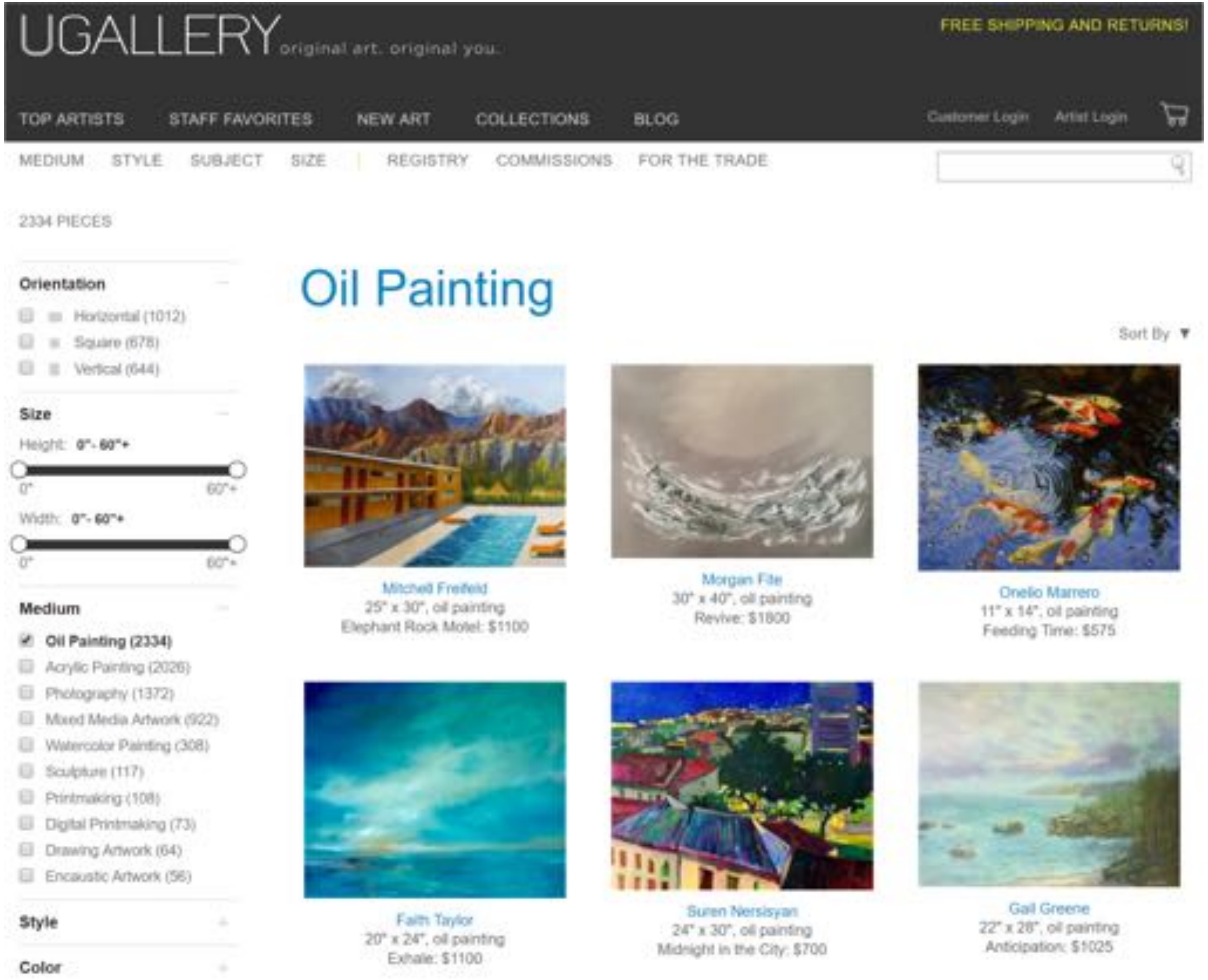
- Recommender systems could play an important role, just like in other e-commerce industries.

https://www.forbes.com/sites/abigailesman/2012/02/29/ the- worlds- strongest- economy- the- global- art- market/ (2012)

https://www.artsy.net/article/artsy-editorial-online-art-market-grew-98-2018-key-takeaways-2019-hiscox-report

UGallery

- Online art store
- Focuses specially in emergent artists: to help them sell their paintings
- Our partners in this project



Agenda

- Introduction & motivation
- **Research Problem**
- Related Work
- Materials
- Inspiration methods (VBPR, YouTube)
- CuratorNet model
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

Research Problem

INPUT

- Anonymized history of art transactions per user
- Images of the art items
- Metadata: artist + some labels by curators (noisy and missing data)

OUTPUT

- Personalized recommendation of paintings for each user

Context: Recomendation of Visual Art

- Few Works in the area (compared to movies or music)

Context: Recomendation of Visual Art

- Few Works in the area (compared to movies or music)
- Most previous work focused on museums and cultural collections

Context: Recomendation of Visual Art

- Few Works in the area (compared to movies or music)
- Most previous work focused on museums and cultural collections
- Few studies about recommending paintings in a commercial setting.

One-of-a-kind Items

- In our dataset we are mostly dealing with a single user feedback: Once purchased, they are removed from the inventory.

One-of-a-kind Items

- In our dataset we are mostly dealing with a single user feedback: Once purchased, they are removed from the inventory.
- There are also photographs, but they account for 10-20% of inventory and purchases
- We have no information of clicks, session length, etc.

Agenda

- Introduction & motivation
- Research Problem
- **Related Work**
- Materials
- Inspiration methods (VBPR, YouTube)
- CuratorNet model
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

Related Work: Art Recommendation

- Previous art recommendation projects date for as long as 2007, such as the CHIP project to recommend paintings from Rijksmuseum. (Aroyo et al., 2007)

Related Work: Art Recommendation

- Previous art recommendation projects date for as long as 2007, such as the CHIP project to recommend paintings from Rijksmuseum. (Aroyo et al., 2007)

- Some approaches have used:
  - Ratings (Aroyo et al., 2007),
  - Tags, Keywords or other textual features (Semeraro et al. 2012)
  - Traditional visual features (van den Broek et al., 2006)

Related Work: Art Recommendation

- Previous art recommendation projects date for as long as 2007, such as the CHIP project to recommend paintings from Rijksmuseum. (Aroyo et al., 2007)

- Some approaches have used:
  - Ratings (Aroyo et al., 2007),
  - Tags, Keywords or other textual features (Semeraro et al. 2012)
  - Traditional visual features (van den Broek et al., 2006)

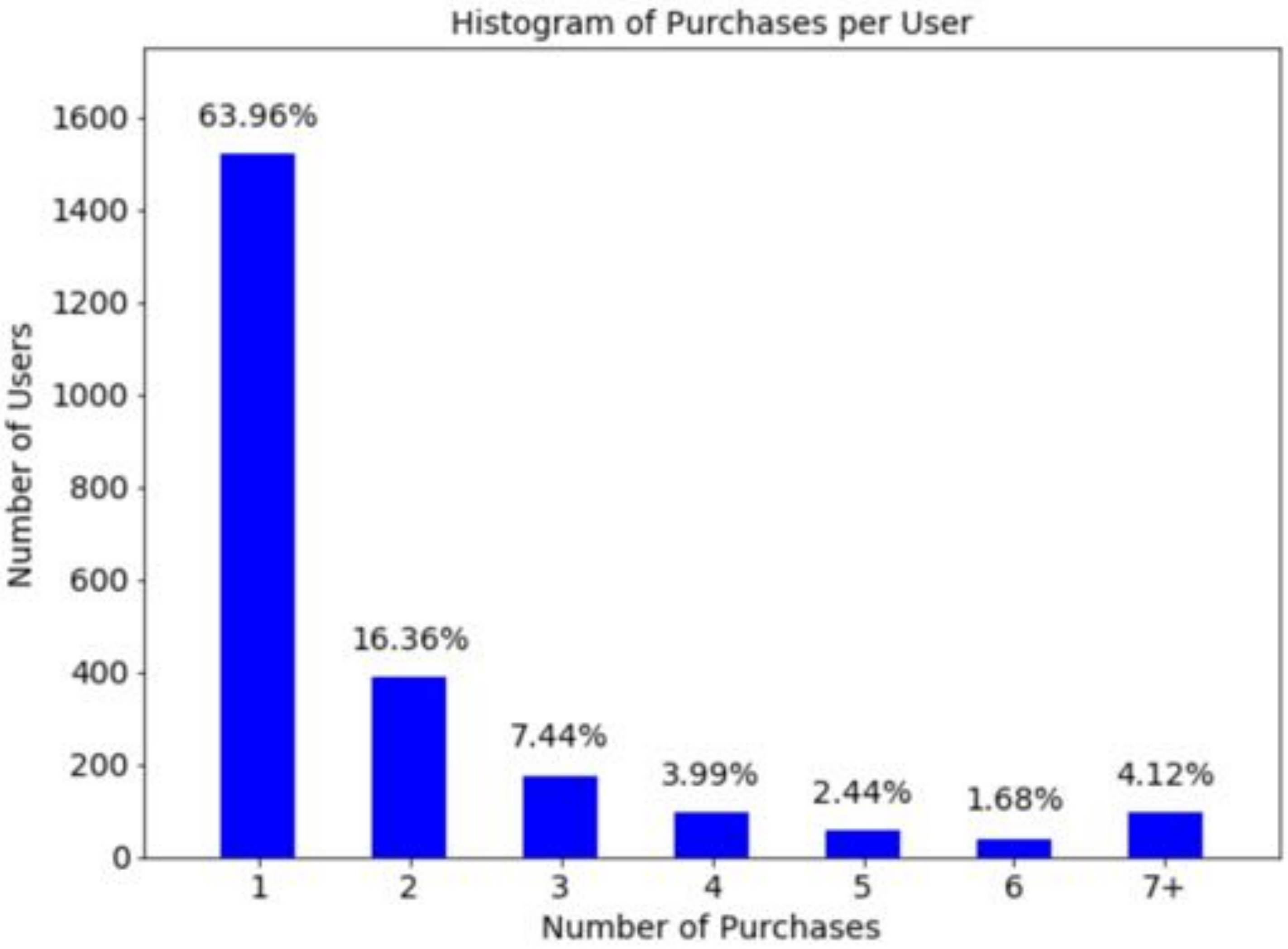
- Deep Learning (DL) have revolutionized computer vision, but few works use DL methods for recommending art (VISTA - He et al. 2016, Messina et al. 2018)

Agenda

- Introduction & motivation
- Research Problem
- Related Work
- **Materials**
- Inspiration methods (VBPR, YouTube)
- CuratorNet model
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

Ugallery Dataset

- 5,336 transactions (purchases)
- 2,378 users
- 6,040 paintings



Agenda

- Introduction & motivation
- Research Problem
- Related Work
- Materials
- **Inspiration methods & their limitations**
- CuratorNet model
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

Inspiration: 1) VBPR (He and McAuley, 2015)

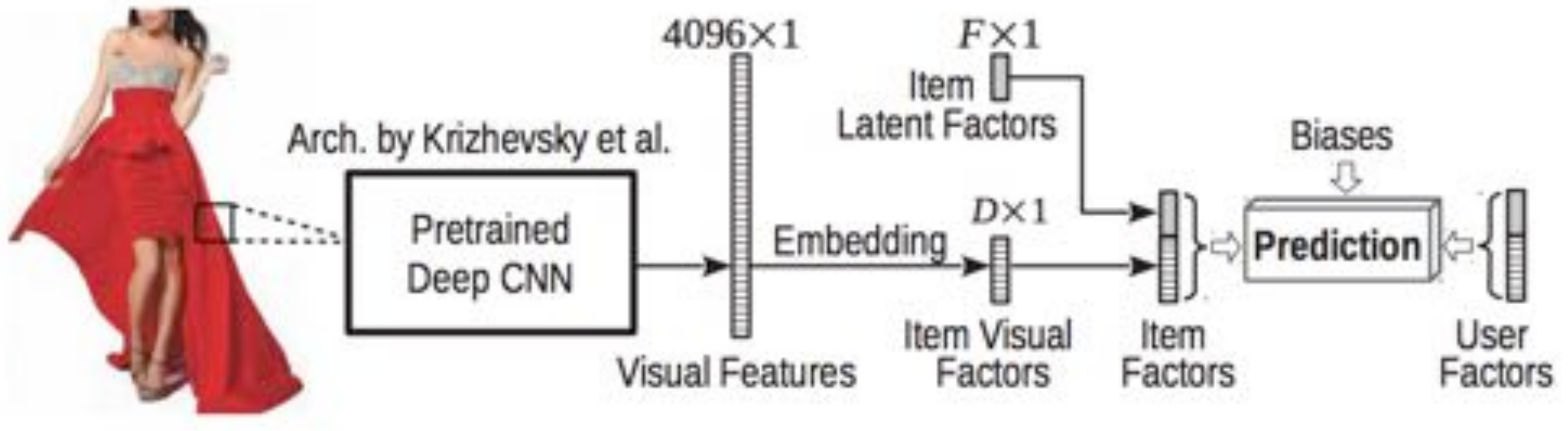


Figure 1: Diagram of our preference predictor. Rating dimensions consist of visual factors and latent (non-visual) factors. Inner products between users and item factors model the compatibility between users and items.

Inspiration: 1) VBPR

VBPR = Visual Bayesian Personalized Ranking (R. He & McAuley, 2016)

$$\hat{x}_{u,i} = \beta_i + \gamma_u^T \gamma_i + \theta_u^T (E f_i) + \beta'^T f_i$$

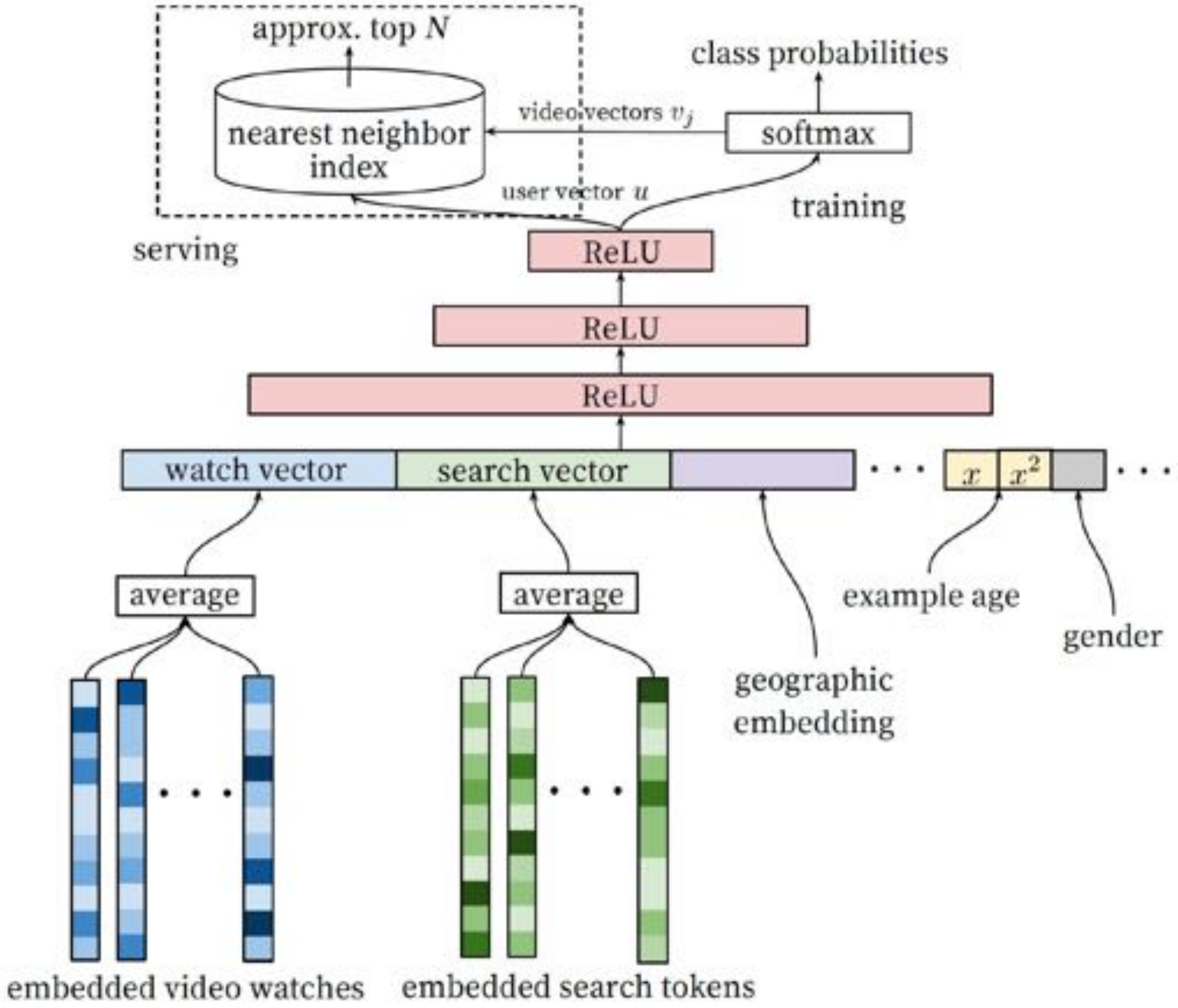
Parameters learned via BPR-OPT (Rendle et al., 2009)

$$D_S = \{(u,i,j) | u \in U \wedge i \in I_u^+ \wedge j \in I \setminus I_u^+\}$$

$$\sum_{(u,i,j) \in D_S} ln(\sigma(\hat{x}_{uij}(\Theta))) - \lambda_\Theta ||\Theta||^2 \qquad \hat{x}_{uij}(\Theta) = \hat{x}_{u,i} - \hat{x}_{u,j}$$

Inspiration:
2) YouTube

Deep Neural Networks for YouTube Recommendations (Covington et.al, 2016)



Conditions for our problem

- *One-of-a-kind items:* We have no explicit co-occurance as in VBPR/VISTA (He & MacAuley, 2016)

- Small dataset, compared to YouTube as well as compared to VBPR/VISTA

- Ideas to addressing our problem: YouTube does not learn a explicit user latent vector and VBPR performs learning by sampling negative feedback.

IDEA: ¿What about combining ***visual content*** *with* ***collaborative information without the need of explicit user latent factors****?*

Agenda

- Introduction & motivation
- Research Problem
- Related Work
- Materials
- Inspiration methods & their limitations
- **CuratorNet model**
- Sampling guidelines
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

CuratorNet

Prediction score:

$$x_{u,i} = \Phi(P_u)^T \Phi(\mathbf{f}_i)$$

Training: Sigmoid Cross-Entropy Loss

$$\mathcal{L} = -\sum_{\mathcal{D}_S} c \ln(\sigma(x_{u,i,j})) + (1-c)\ln(1-\sigma(x_{u,i,j})) + \lambda_\Theta ||\Theta||^2$$

$$P(i >_u j|\Theta) = \sigma(x_{u,i,j}) = \frac{1}{1+e^{-(x_{u,i}-x_{u,j})}}$$

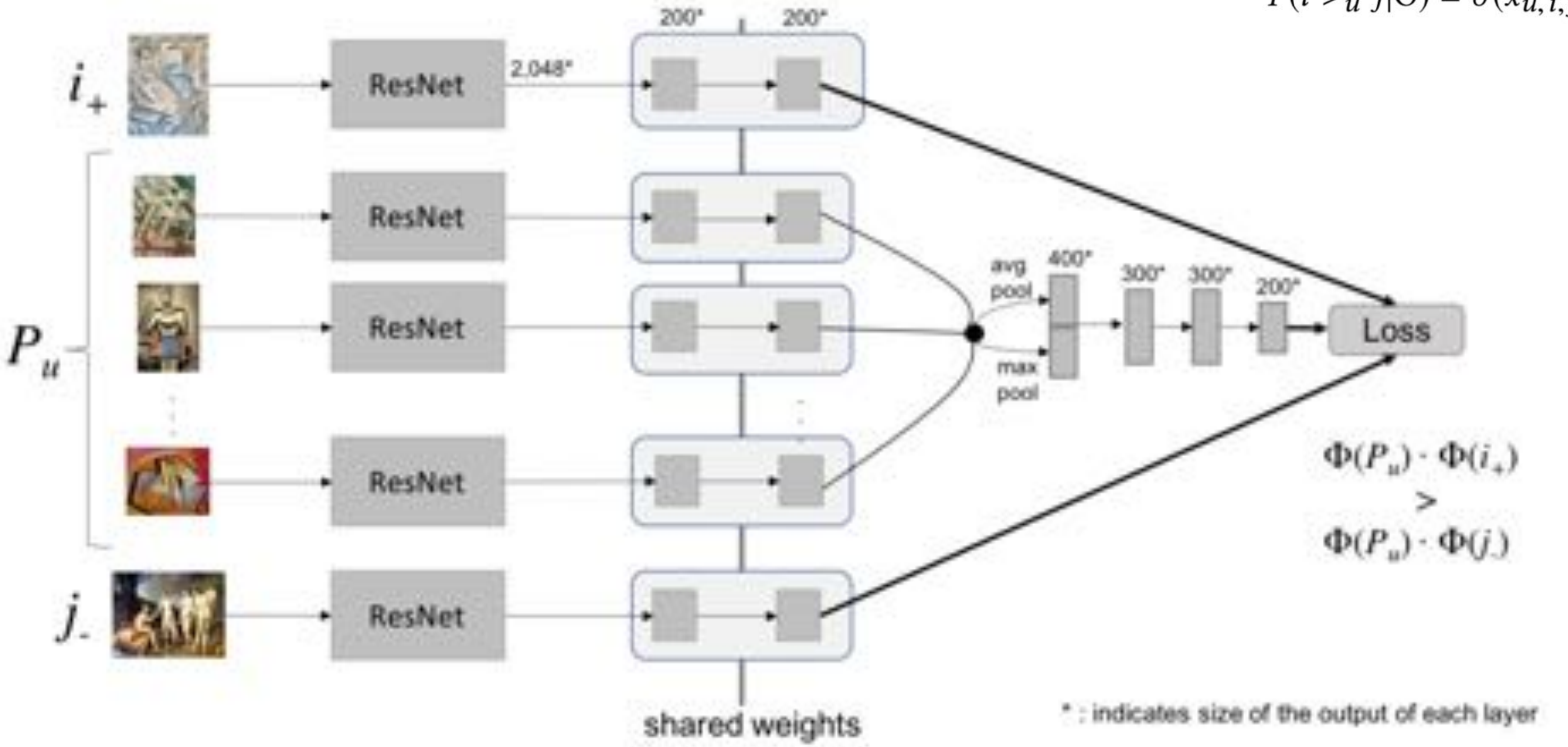


Figure 2: Architecture of CuratorNet showing in detail the layers with shared weights for training.

Training CuratorNet

- Similar to BPR: Given a training set D_S f triples (p, i, j) we aim that our model score like:

$$\vec{u_p} \cdot \vec{i} > \vec{u_p} \cdot \vec{j}$$

- Unlike BPR, we do not randomly sample negative examples for the training set D_S

Agenda

- Introduction & motivation
- Research problem
- Related work
- Materials
- Inspiration methods & their limitations
- CuratorNet model
- **Sampling guidelines**
- Recommendation
- Results
- Conclusion
- Discussion & Future Work

Sampling guidelines for triplets

- Based on findings of our previous work (favorite artist)
- Using notion of visual clusters

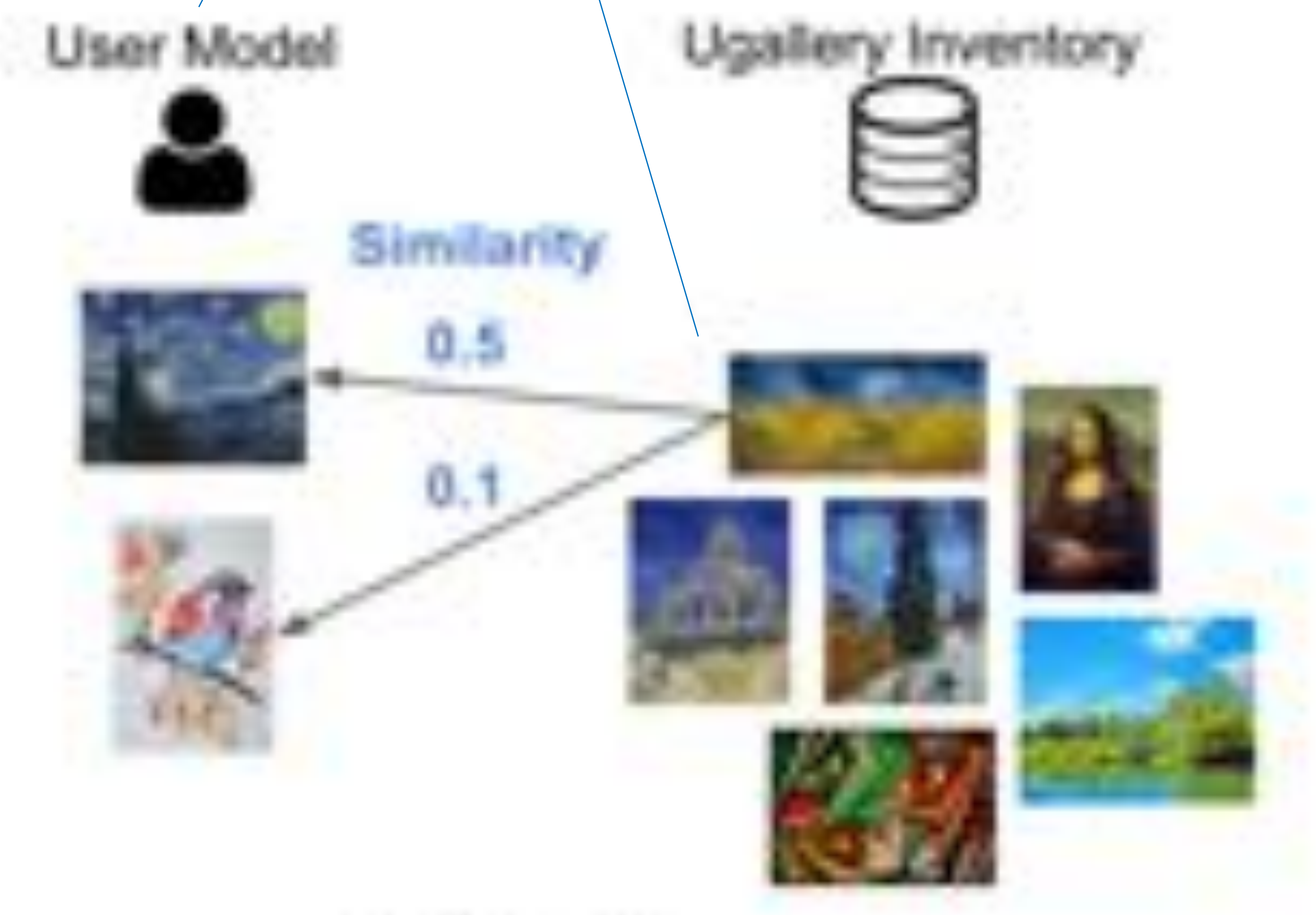
Figure 3: Examples of visual clusters automatically generated to sample triples for the training set.

Agenda

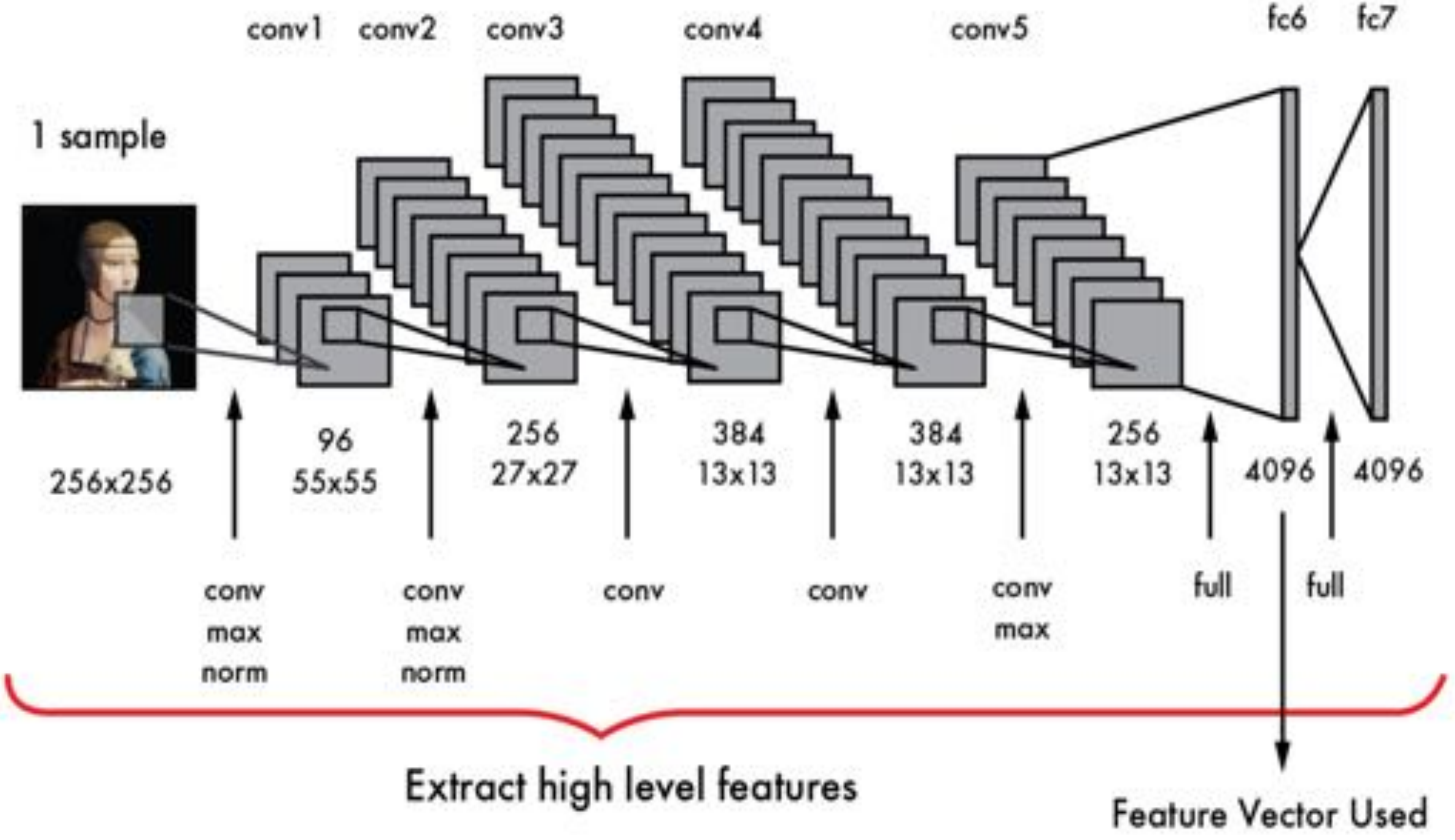
- Introduction & motivation
- Research problem
- Related work
- Materials
- Inspiration methods & their limitations
- CuratorNet model
- Sampling guidelines
- **Recommendation step**
- Results
- Conclusion
- Discussion & Future Work

Content-based recommendation: VisRank (baseline)

s(u,i) = score(ContentBasedProfile(u), Content(i))

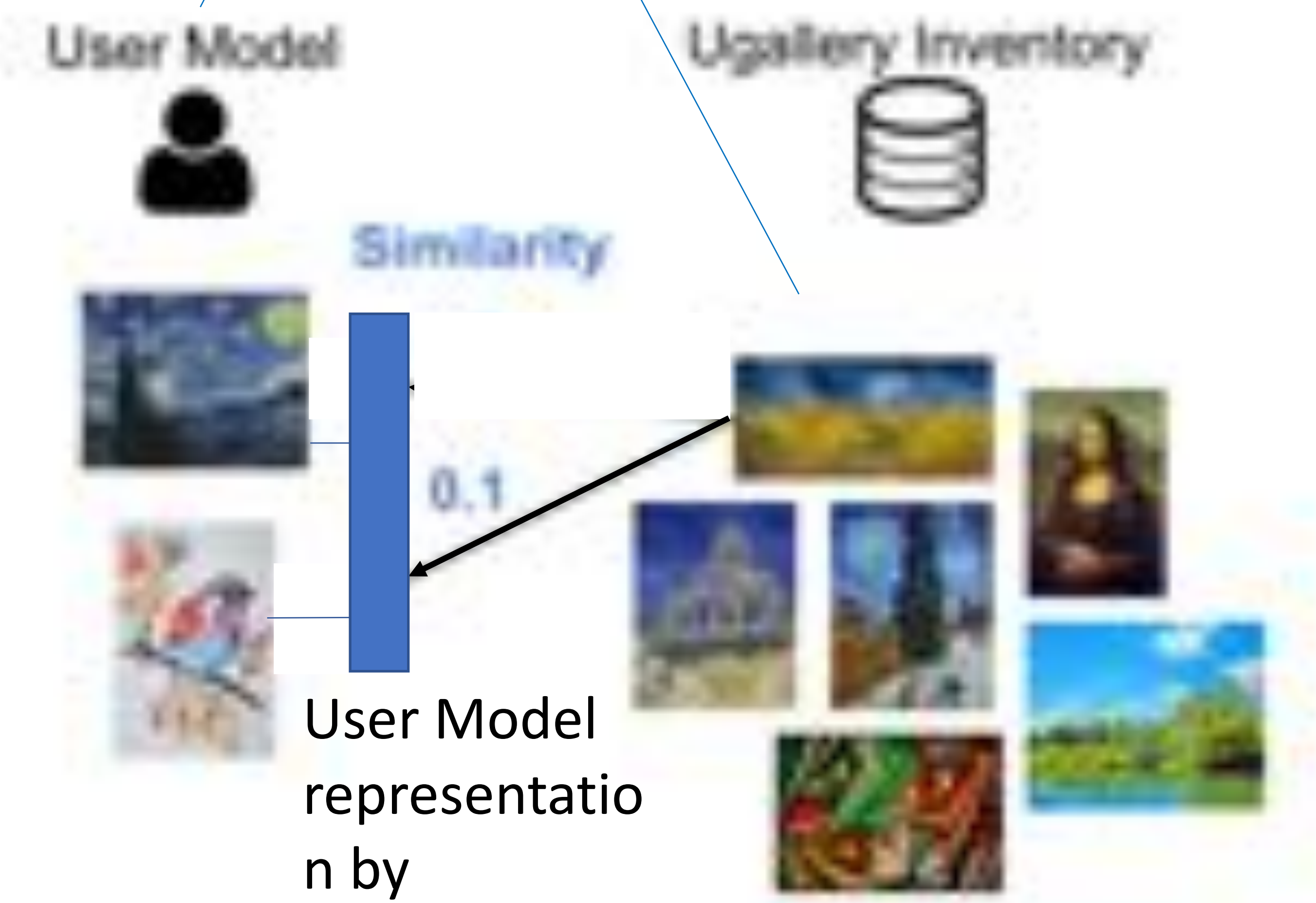


VisRank: use Deep Convolutional Neural Networks (**CNN**)



Content-based recommendation: CuratorNet

$s(u,i) \;= score(\, ContentBasedProfile(u), Content(i)\,)$

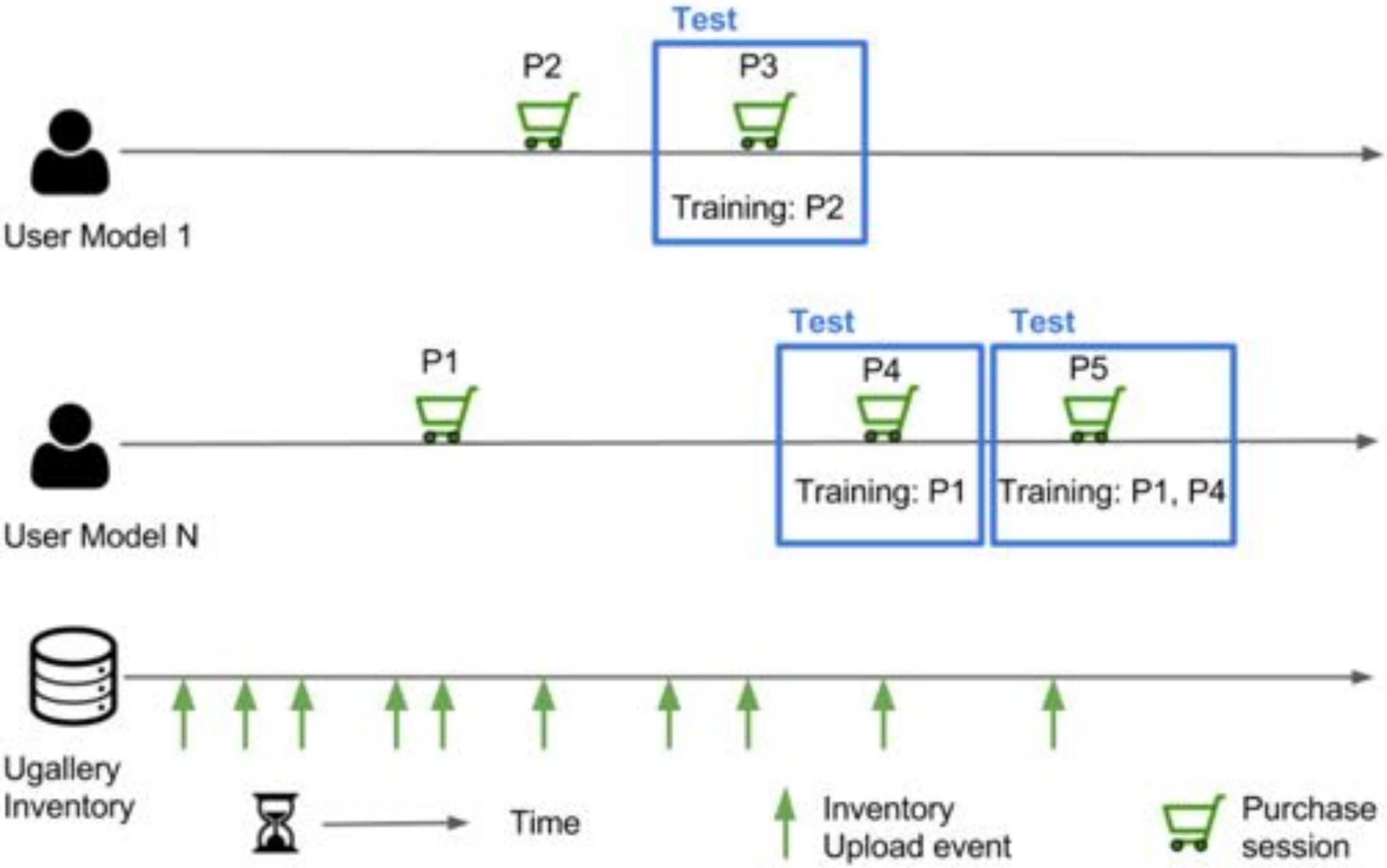


User Model representation by CuratorNet

Agenda

- Introduction & motivation
- Research problem
- Related work
- Materials
- Inspiration methods & their limitations
- CuratorNet model
- Sampling guidelines
- Recommendation step
- **Evaluation and results**
- Conclusion
- Discussion & Future Work

Offline Evaluation (Purchase Records)



Results I

Method	λ (L2 Reg.)	AUC	R@20	P@20	nDCG@20	R@100	P@100	nDCG@100
Oracle	–	**1.0000**	**1.0000**	**.0655**	**1.0000**	**1.0000**	**.0131**	**1.0000**
CuratorNet	.0001	**.7204**	**.1683**	**.0106**	**.0966**	**.3200**	**.0040**	**.1246**
CuratorNet	.001	**.7177**	**.1566**	**.0094**	**.0895**	**.2937**	**.0037**	**.1160**
VisRank	–	**.7151**	**.1521**	**.0093**	**.0956**	**.2765**	**.0034**	**.1195**
CuratorNet	0	**.7131**	**.1689**	**.0100**	**.0977**	**.3048**	**.0038**	**.1239**
CuratorNet	.01	.7125	.1235	.0075	.0635	.2548	.0032	.0904
VBPR	.0001	.6641	.1368	.0081	.0728	.2399	.0030	.0923
VBPR	0	.6543	.1287	.0078	.0670	.2077	.0026	.0829
VBPR	.001	.6410	.0830	.0047	.0387	.1948	.0024	.0620
VBPR	.01	.5489	.0101	.0005	.0039	.0506	.0006	.0118
Random	–	.4973	.0103	.0006	.0041	.0322	.0005	.0098

Results I

Method	λ (L2 Reg.)	AUC	R@20	P@20	nDCG@20	R@100	P@100	nDCG@100
Oracle	–	**1.0000**	**1.0000**	**.0655**	**1.0000**	**1.0000**	**.0131**	**1.0000**
CuratorNet	.0001	**.7204**	**.1683**	**.0106**	**.0966**	**.3200**	**.0040**	**.1246**
CuratorNet	.001	**.7177**	**.1566**	**.0094**	**.0895**	**.2937**	**.0037**	**.1160**
VisRank	–	**.7151**	**.1521**	**.0093**	**.0956**	**.2765**	**.0034**	**.1195**
CuratorNet	0	**.7131**	**.1689**	**.0100**	**.0977**	**.3048**	**.0038**	**.1239**
CuratorNet	.01	.7125	.1235	.0075	.0635	.2548	.0032	.0904
VBPR	.0001	.6641	.1368	.0081	.0728	.2399	.0030	.0923
VBPR	0	.6543	.1287	.0078	.0670	.2077	.0026	.0829
VBPR	.001	.6410	.0830	.0047	.0387	.1948	.0024	.0620
VBPR	.01	.5489	.0101	.0005	.0039	.0506	.0006	.0118
Random	–	.4973	.0103	.0006	.0041	.0322	.0005	.0098

Results II: Effect of sampling guidelines

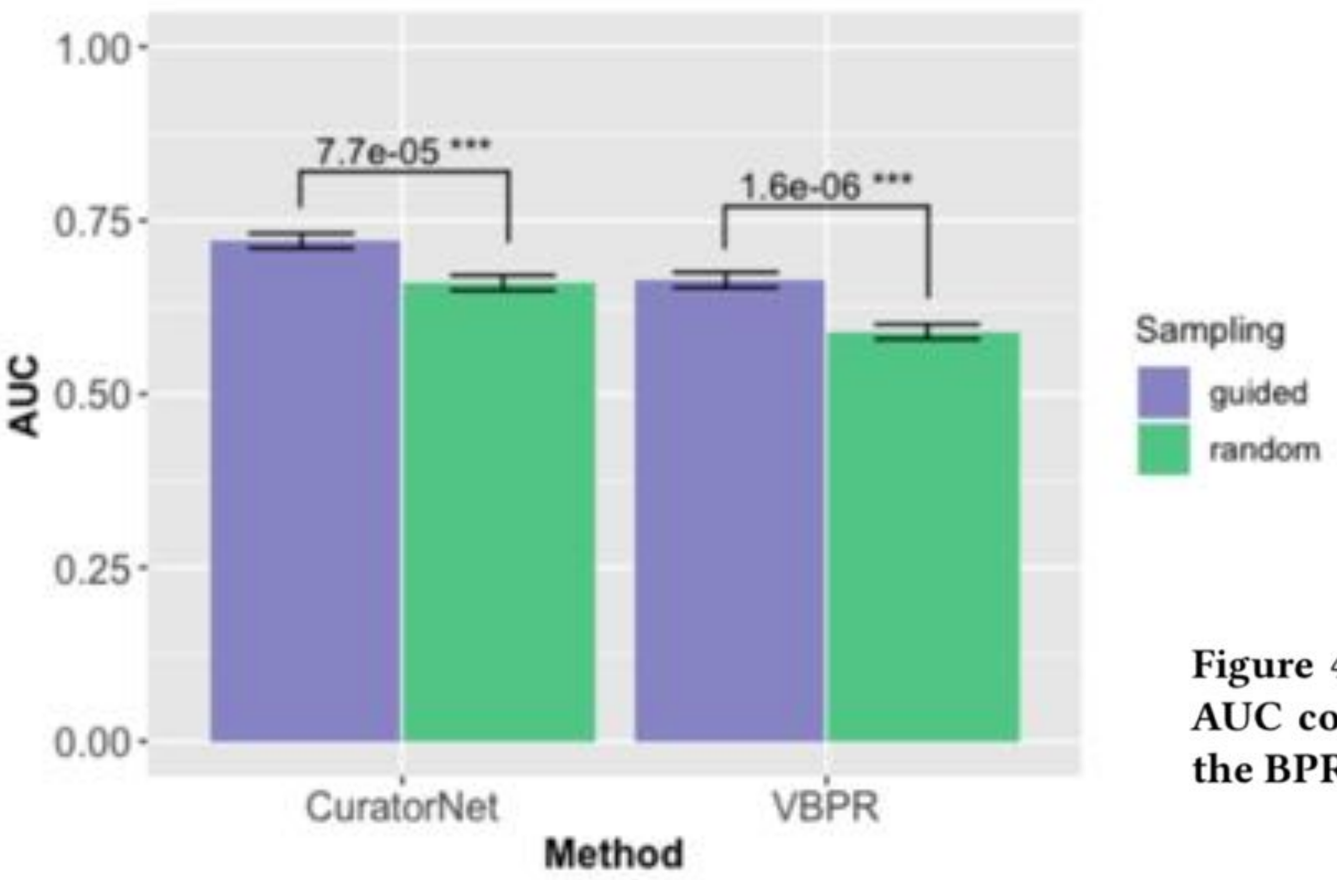
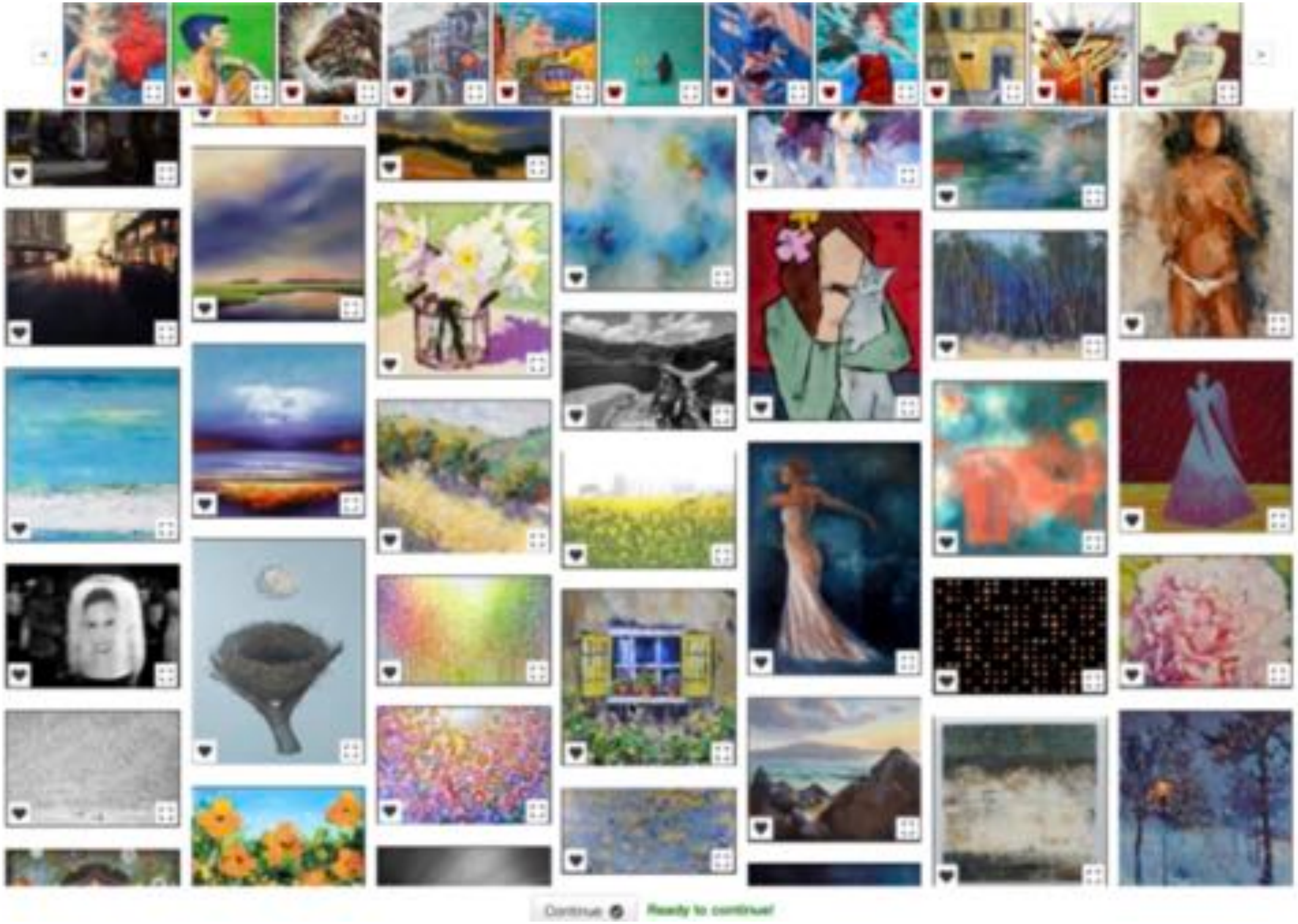


Figure 4: The sampling guidelines had a positive effect on AUC compared to random negative sampling for building the BPR training set.

Demo



Agenda

- Introduction & motivation
- Research problem
- Related work
- Materials
- Inspiration methods & their limitations
- CuratorNet model
- Sampling guidelines
- Recommendation step
- Evaluation and results
- **Conclusion**
- Discussion & Future Work

Conclusion

- CuratorNet improves upon VBPR and VisRank

- Unline VBPR, CuratorNet does not need re-training to recommend to new users since it does not explicitly train user factors

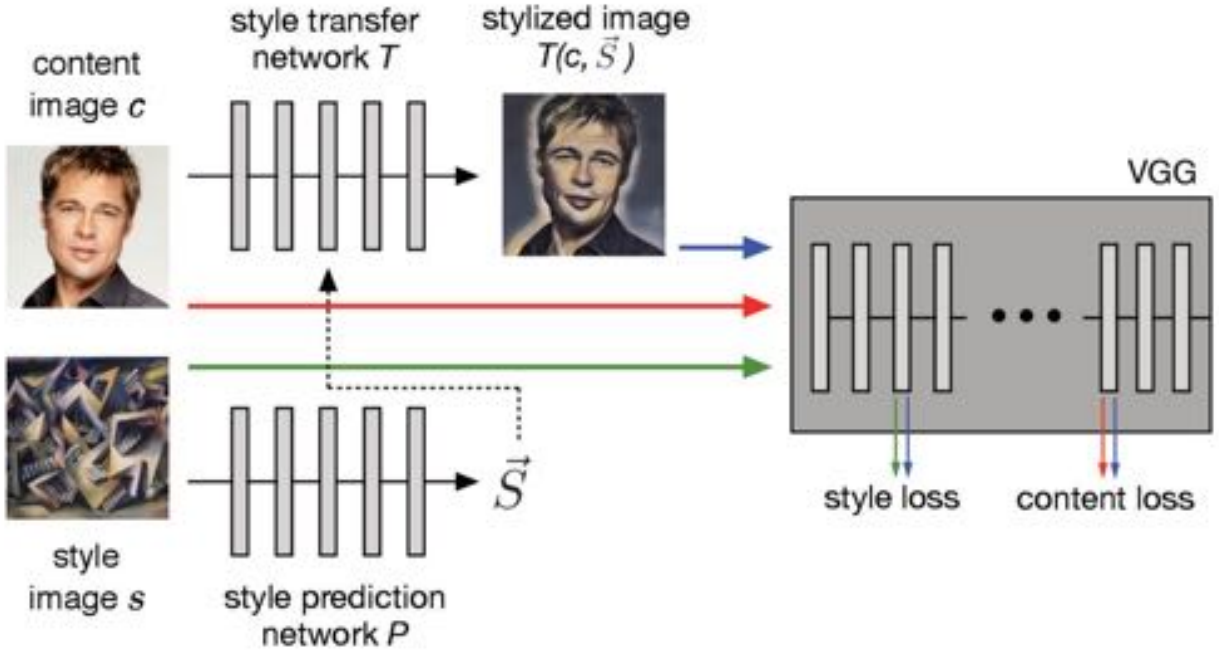
- The proposed sampling guidelines benefit both CuratorNet and VBPR

Agenda

- Introduction & motivation
- Research problem
- Related work
- Materials
- Inspiration methods & their limitations
- CuratorNet model
- Sampling guidelines
- Recommendation step
- Evaluation and results
- Conclusion
- **Discussion & Future Work**

Discussion and Future Work

- End to end learning visual features, other losses
- Separate style from content Ghiasi et al. (2017): *Exploring the structure of a real-time, arbitrary neural artistic stylization network*



Millennium Institute Foundational Research on Data

SoCVis PUC SOCIAL COMPUTING & VISUALIZATION GROUP

IA Lab

Interpretable Contextual Team-aware Item Recommendation (TTIR)

Andrés Villa*, PhD student

Vladimir Araujo *, PhD student

Francisca Cattan *, PhD student

Denis Parra *, Associate Professor

*CS Department, IA Lab, Pontificia Universidad Católica de Chile

dparras@uc.cl

denisparra

ACM RecSys 2020

Motivation

- Global e-sports revenues and its audience will grow to $1.1 billion and 495 million people in 2020

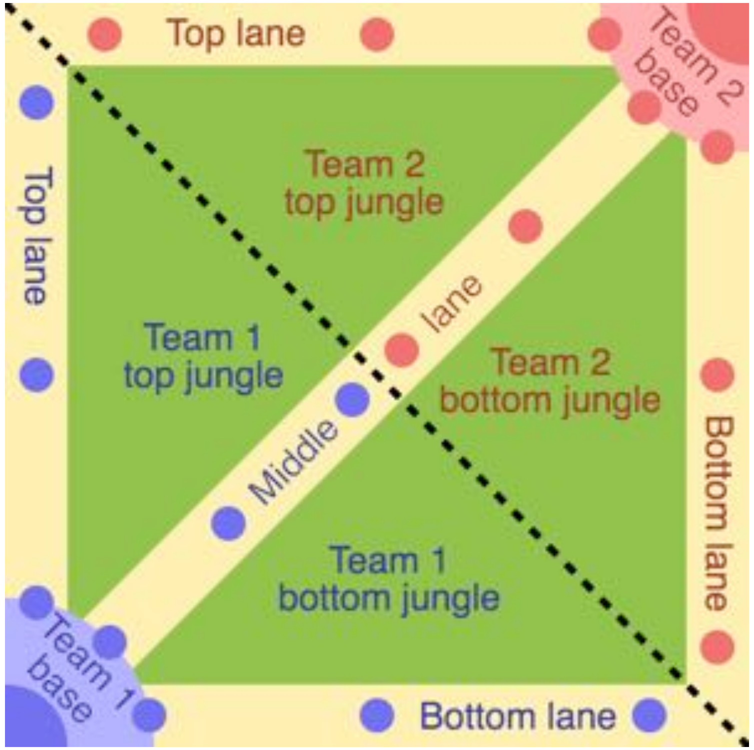
- MOBA (Multiplayer Online Battle Arena) is one of the most significant social gaming genres contributing to that growth.

- The League of Legends World Championship, which was the biggest tournament of 2019 with more than 105 million hours live on Twitch and YouTube.

https://strivesponsorship.com/wp-content/uploads/2020/03/Global-Esports-Market-Report-2020.pdf

LoL and a typical MOBA map

- League of Legends (LoL) has dominated the market since 2012.
- The game consists of two teams (red and blue) of five players each, that compete to be the first to destroy the enemy base.
- Each player controls one character (champion) that interacts with the rest through combats, which are carried out in a particular arena.



Typical MOBA map. Circles represent turrets. The green area is the jungle and the yellow area are the lanes. The corner areas are the team's base.

The Recommendation Problem

- This game presents at least two recommendation problems: champion (Chen et al., 2018), there are more than 140 characters, and
- Item recommendation (Looi et al., 2018), there are around 240 items.

Example of two teams matchup and item recommendations for the Blue team.

*Zhengxing **Chen**, Truong-Huy D Nguyen, Yuyu Xu, Christopher Amato, Seth Cooper, Yizhou Sun, and Magy Seif El-Nasr. **2018**. The art of drafting: a team-oriented hero recommendation system for multiplayer online battle arena games.*

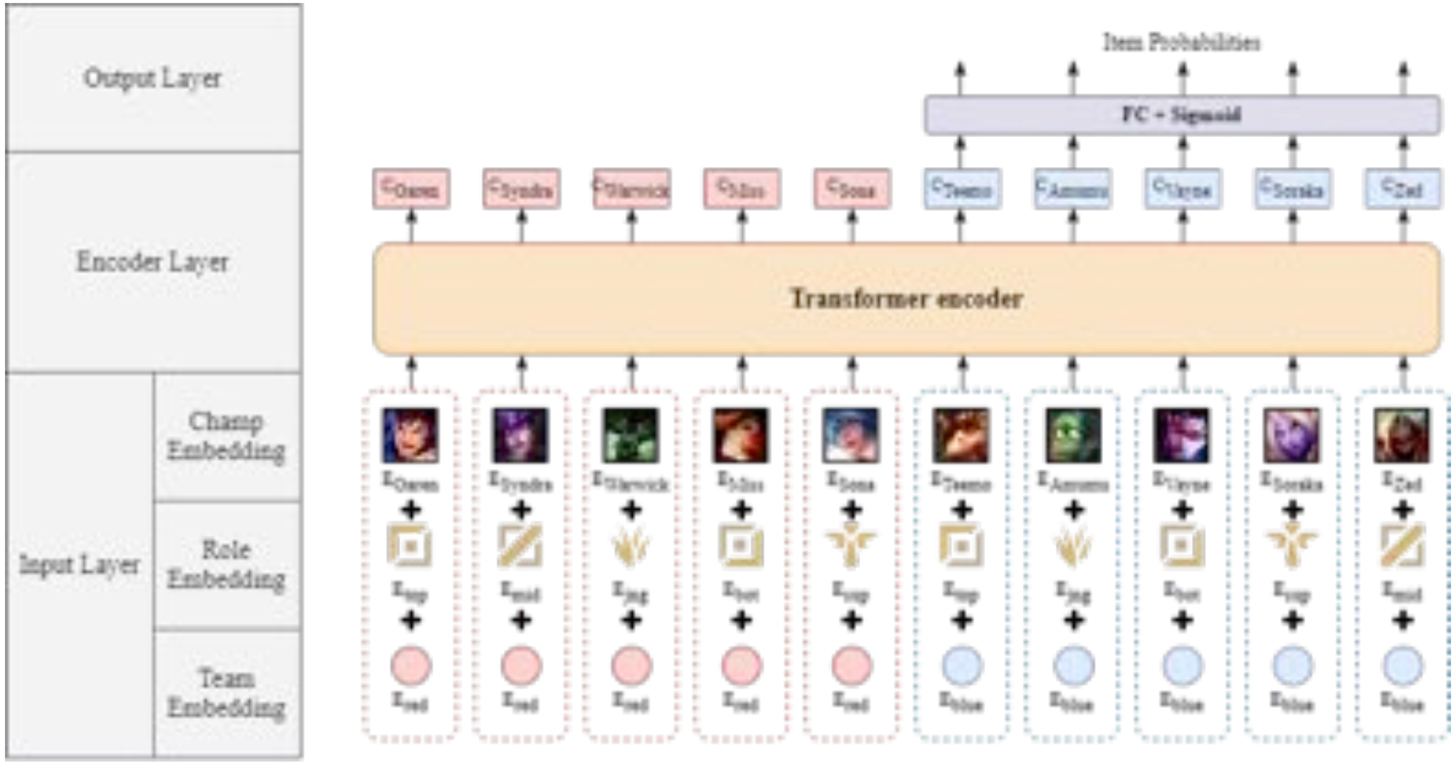
*W. **Looi**, M. Dhaliwal, R. Alhajj and J. Rokne. **2019**. "Recommender System for Items in Dota 2," in IEEE Transactions on Games,*

Our Approach: The transformer

- Figure shows the Transformer for Team-aware Item Recommendation architecture (TTIR).

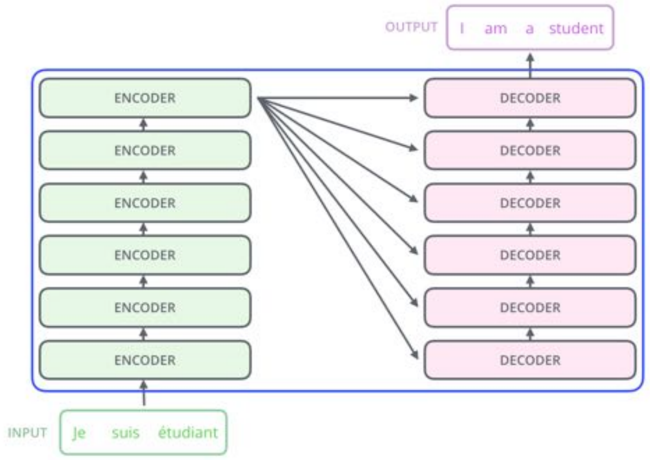
- This model is made up of three major parts: the input representation layer, the encoder layer, and the output layer for recommendation.

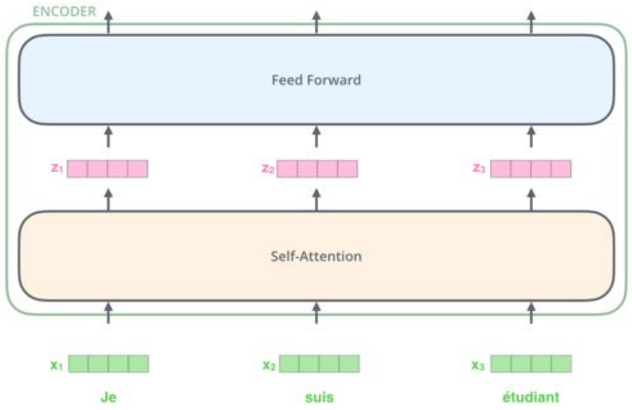
$$E_{input} = E_{champ} + E_{role} + E_{team}$$

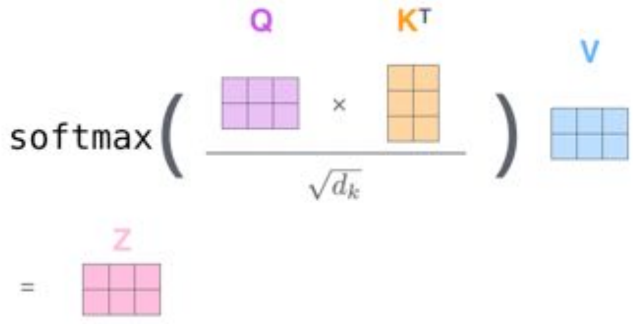


What is the Transformer ? How does it work?

- It is a neural architecture proposed in the article "Attention is all you need" by Vaswani et al. (2017)
- I suggest you to read the article, but for a visual detailed description I suggest you read this https://jalammar.github.io/illustrated-transformer/ by Jay Alammar







Evaluation

- Kaggle Dataset*. 184,070 game sessions in the ranked category, a competitive alternative to the normal match.

Table 1. Overview of the dataset

	LoL Ranked Matches 7th Season
# Items	89
# Champions	136
# Matches	157,584
Roles	Top, Mid, Jungle, Support, Bot
Train / Test	1,261,280 / 314,560

* http://www.kaggle.com/paololol/league-of-legends-ranked-matches

Results

(a) Results for top @k recommendation. TTIR is significantly better than the second best method, CNN.

	Method					T-test(df = 314557)
	D-Tree	Logit	ANN	CNN	TTIR	p-Value (t-Stat)
Precision@1	0.516	0.672	0.771	0.790	**0.803**	5.30e-20 (9.158)
Recall@1	0.135	0.178	0.205	0.209	**0.214**	9.54e-18 (8.580)
F1@1	0.210	0.277	0.318	0.331	**0.338**	7.23e-20 (9.124)
MAP@1	0.516	0.672	0.771	0.790	**0.803**	5.30e-20 (9.158)
Precision@6	0.319	0.393	0.476	0.484	**0.492**	2.41e-22 (9.723)
Recall@6	0.491	0.607	0.732	0.744	**0.756**	1.93e-27 (10.854)
F1@6	0.379	0.468	0.566	0.586	**0.596**	2.57e-27 (10.828)
MAP@6	0.648	0.714	0.785	0.795	**0.805**	3.77e-30 (11.410)
Precision@10	0.204	0.285	0.341	0.348	**0.351**	2.49e-11 (6.674)
Recall@10	0.520	0.726	0.864	0.882	**0.889**	1.43e-24 (10.232)
F1@10	0.289	0.403	0.481	0.499	**0.503**	3.34e-15 (7.878)
MAP@10	0.636	0.672	0.743	0.754	**0.764**	1.32e-34 (12.270)

Results – Ablation analysis

(b) Ablation study of TTIR

	P@6	R@6	MAP@6
Default (h=2, l=1)	0.492	0.756	0.805
Multiheads (h=1)	0.462	0.726	0.778
Multiheads (h=4)	0.492	0.756	0.806
Layers (l=2)	0.493	0.757	0.806
Layers (l=3)	0.493	0.758	0.807
w/o $E_{enemies}$	0.487	0.749	0.798
w/o E_{role}	0.484	0.742	0.794
w/o $E_{enemies}, E_{role}$	0.479	0.736	0.787
CNN w/ $E_{enemies}, E_{role}$	0.484	0.744	0.795

User Survey

Fig. 3. Visualization of the attention weights for each member of the Blue team on each member of both teams (bottom row).

User Survey

Table 3. Results of the preliminary user survey (N=16), ratings in range [1-10]

Question	Global M±SD (N=16)	Subjects by year of first play: 2009-11 (N=5)	2012-14 (N=5)	2015-2017 (N=6)
Q1. How good were the recommendations for the *Blue team* ?	7.98±1.22	7.7±1.24	7.7±1.16	8.46±1.3
Q2. Is it understandable the influence of every team member upon each champion being recommended ?	7.44±1.72	7.4±1.55	7.1±0.8	7.75±2.49
Q3. Is it useful the information provided by the visualization in order to understand the item recommendations made ?	6.9±2.15	6.7±1.98	6.6±1.65	7.33±2.87

User Survey – Samples of comments

- Positive Comments
  - "you can see exactly the focus of each champion with respect to the main enemies on the facing team..."
  - "useful build to prevent enemy ganking...",
  - "...with this build, Vlad hinders the enemy, making Lucian suffer. Then Ezreal with that build can damage both Lux and Fizz"

- Critical Comments
  - "this explanation missed armor penetration and grievous wounds..."
  - "it doesn't show with which item I need to start and ***the sequence to progress..***."
  - "recommendations do not show magic resistance..."

Code and data

https://github.com/ojedaf/IC-TIR-Lol

Denis Parra
Associate Professor
Pontificia Universidad Católica de Chile
dparra@ing.puc.cl

Thanks!

User Survey – Samples of comments

- Positive Comments
  - “you can see exactly the focus of each champion with respect to the main enemies on the facing team...”
  - “useful build to prevent enemy ganking...”,
  - “...with this build, Vlad hinders the enemy, making Lucian suffer. Then Ezreal with that build can damage both Lux and Fizz”

- Critical Comments
  - “this explanation missed armor penetration and grievous wounds...”
  - “it doesn’t show with which item I need to start and ***the sequence to progress..***.”
  - “recommendations do not show magic resistance...”

Scalable Recommendation of Wikipedia Articles to Editors (WikiRecNet)

Oleksii Moskalenko (UCU),
Denis Parra (PUC), and Diego Sáez-Trumper* (WMF)

ComplexRec Workshop @ RecSys - Online Event
Sept. 2020

Wikipedia is edited by volunteer editors around the world.

Considering the large amount of existing content, *over 5M articles just in English Wikipedia*, deciding what to edit next can be difficult.

Average: 390K

The count of editors with one or more edits, including on redirect pages. More info about this metric.
You can see or add annotations for this data on this wiki article.

Challenges

- Minimal user data collection
- Diversity of edit types
- Different topical interests
- Different editors expertises

Opportunities

- 18 years of open data
- Full article content and history available
- Complete editors history is also open (logged users)
- Articles written in WikiText provides semi-structured data

WIKIPEDIA
The Free Encyclopedia

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute

Help
Community portal
Recent changes
Upload file

Tools

What links here
Related changes

Technology in *Star Trek*

From Wikipedia, the free encyclopedia

"Physics and Star Trek" redirects here. For the nonfiction book, see The Physics of Star Trek.

This article has multiple issues. Please help **improve it** or [hide] discuss these issues on the **talk page**. *(Learn how and when to remove these template messages)*

- This article **needs additional citations for verification**. *(August 2008)*
- This Star Trek-related article **describes a work or element of fiction in a primarily in-universe style**. *(October 2007)*

The **technology in *Star Trek*** has borrowed many ideas from the scientific world. Episodes often contain technologies named after real-world scientific phenomena, such as tachyon beams, baryon sweeps, quantum slipstream drives, and photon torpedoes. Some of the technologies created for the *Star Trek* universe were done so out of financial necessity. For instance, the transporter was created because the limited budget of *Star Trek: The Original Series* (*TOS*) in the 1960s did not allow expensive shots of spaceships landing on planets.[1][*page needed*]

Harmonica

From Wikipedia, the free encyclopedia

This article has multiple issues. Please help **improve it** or [hide] discuss these issues on the **talk page**. *(Learn how and when to remove these template messages)*

- This article **needs additional citations for verification**. *(February 2017)*
- This article **possibly contains original research**. *(February 2017)*

The **harmonica**, also known as a **French harp** or **mouth organ**, is a free reed wind instrument used worldwide in many musical genres, notably in blues, American folk music, classical music, jazz, country, and rock. The many types of harmonica include diatonic, chromatic, tremolo, octave, orchestral, and bass versions. A harmonica is played by using the mouth (lips and tongue) to direct air into or out of one (or more) holes along a mouthpiece. Behind each hole is a chamber containing at least one reed. A harmonica reed is a flat, elongated spring typically made of brass, stainless steel, or bronze, which is secured at one end over a slot that serves as an airway. When the free end is made to vibrate by the player's air, it alternately blocks and unblocks the airway to produce sound.

Harmonica

A 16-hole chromatic (top) and 10-hole diatonic harmonica

Other instrument	
Other names	French harp ·

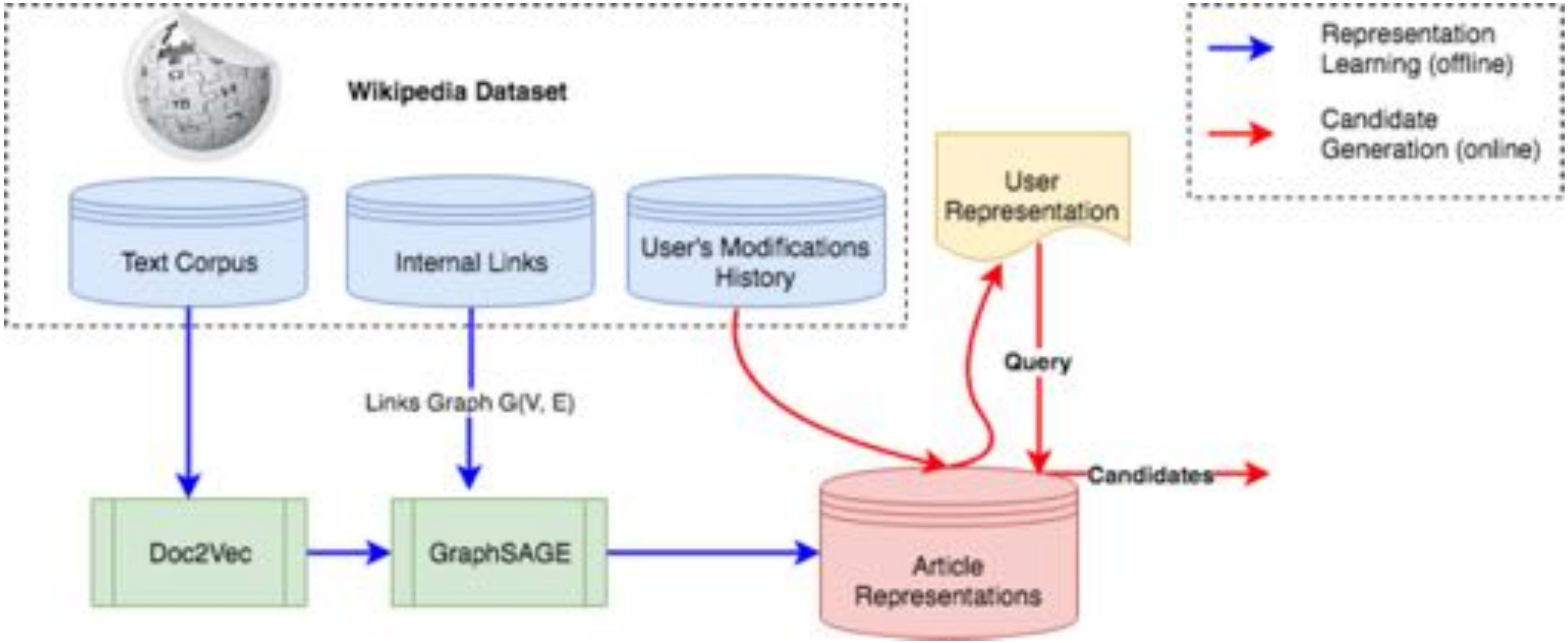
Research Resources

- Static Dumps
- MediaWiki Utilities
- Wikimedia API
- Page Views
- SQL Replicas / Quarry
- Clicks
- Event Stream
- Wikidata
- Commons
- ORES
- PAWS

On this paper

Recommend relevant articles to editors using GCN and Doc2Vec, learning how to represent articles

Model summary



GraphSAGE

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In Advances in neural information processing systems (pp. 1024-1034).

1. Sample neighborhood

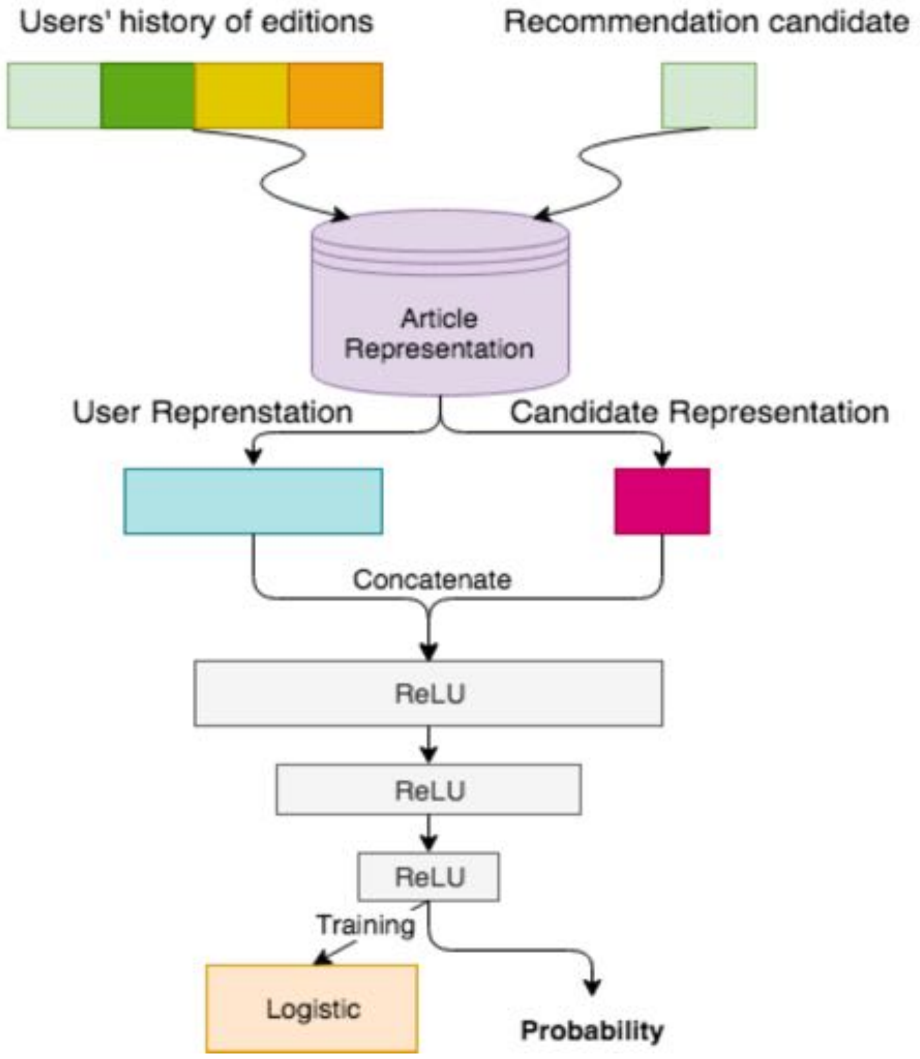
2. Aggregate feature information from neighbors

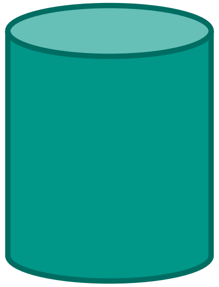
3. Predict graph context and label using aggregated information

To run GraphSAGE, it needs to train on an example graph or set of graphs. After training, GraphSAGE can be used to generate node embeddings for previously unseen nodes or entirely new input graphs, as long as these graphs have the same attribute schema as the training data.

Candidate Ranking

User history along with candidate are passed through articles' representation database (EmbeddingLayer) and then through several fully-connected layers to train inthe log-regression setup





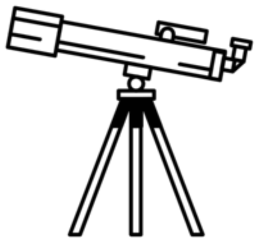
Dataset

Table 1: Performance of different algorithms for K-NN search. All tests were conducted with English Wikipedia articles ($|V| = 5,251,875$). Setup is measured in seconds. Secs. req. means seconds per request.

Algorithm	Setup	Secs./req.	Recall	MRR
Exact search	3.91	0.81	0.224	0.0220
IVF	207.02	0.07	0.206	0.0212
HNSW	232.68	0.04	0.224	0.0220
LSH	472.31	0.15	0.215	0.0219

Table 2: Specifications of built Wikipedia Graph

Specification	English Wikipedia
Amount of vertices ($\|V\|$)	5,251,875
Amount of Edges ($\|E\|$)	458,867,626
Average Degree ($\overline{d_{all}}$)	174
Median Degree ($\widetilde{d_{all}}$)	60
Approx. Diameter (D)	23
Amount of labeled nodes	4,652,604



Evaluation

- Consider users with ten edits or more
- Sample a window of 10 edits
- Use the first 5 articles edited to predict the next 5
- Compare against Doc2Vec, ALS and BM25

Experiment

1. Take first 5 articles per user. Calculate avg of their embeddings
2. Generate candidates by NN search
3. Sort candidates according to ranking algorithm
4. Compare Top-K recommendations with the next 5 articles

Results

Table 3: Offline evaluation of generated recommendations on the task of predicting next 5 articles edited by user with percentage improvement over content-based model Doc2Vec (mean-pool) with cosine similarity.

			K=50			K=100		
Model	Aggregate	Rank	MAP	nDCG	Recall	MAP	nDCG	Recall
WikiRecNet	mean	cosine	0.0221	0.1361	0.0846	0.0238 (+78%)	0.1468 (+66%)	0.1179 (+99%)
	mean	deep-rank	0.0228	0.1363	0.0841	0.0243 (+82%)	0.1493 (+70%)	0.1134 (+92%)
	max	cosine	0.0192	0.1196	0.0672	0.0206 (+54 %)	0.1299 (+47%)	0.0923 (+56%)
	merge	cosine	0.0208	0.1412	0.0825	0.0227 (+70%)	0.1538 (+75%)	0.1175 (+99%)
	merge	deep-rank	**0.0262**	**0.1625**	**0.0935**	**0.0282 (+111%)**	**0.1760 (+100%)**	**0.1302 (+120%)**
Doc2Vec	merge	cosine	0.0085	0.0805	0.0438	0.0092	0.0883	0.0600
	mean	cosine	0.0126	0.0821	0.0436	0.0133	0.0880	0.0590
BM25			0.0251	0.1602	0.0921	0.0273	0.1710	0.1290
ALS MF			0.0027	0.0163	0.044	0.0063	0.0204	0.0609

Final remarks

- WikiRecNet architecture is composed of two networks, improving existing implementations that were not capable to work in such scenarios.
- Our approach does not need to be retrained when new items are added.
- Code and links to data available here:
  - https://github.com/digitalTranshumant/WikiRecNet-ComplexRec2020

Future Work

- Go beyond text, and use the image information on articles to improve pages and users representation
- Enrich the graph neural network by adding entities information coming from a knowledge graph (Wikidata)
- Run a user experiment, to test our recommendation with real Wikipedia editors
- Add fairness metrics (eg. gender of the main subject of the article) to deal with known biases on Wikipedia data.

Thank you!

diego@wikimedia.org