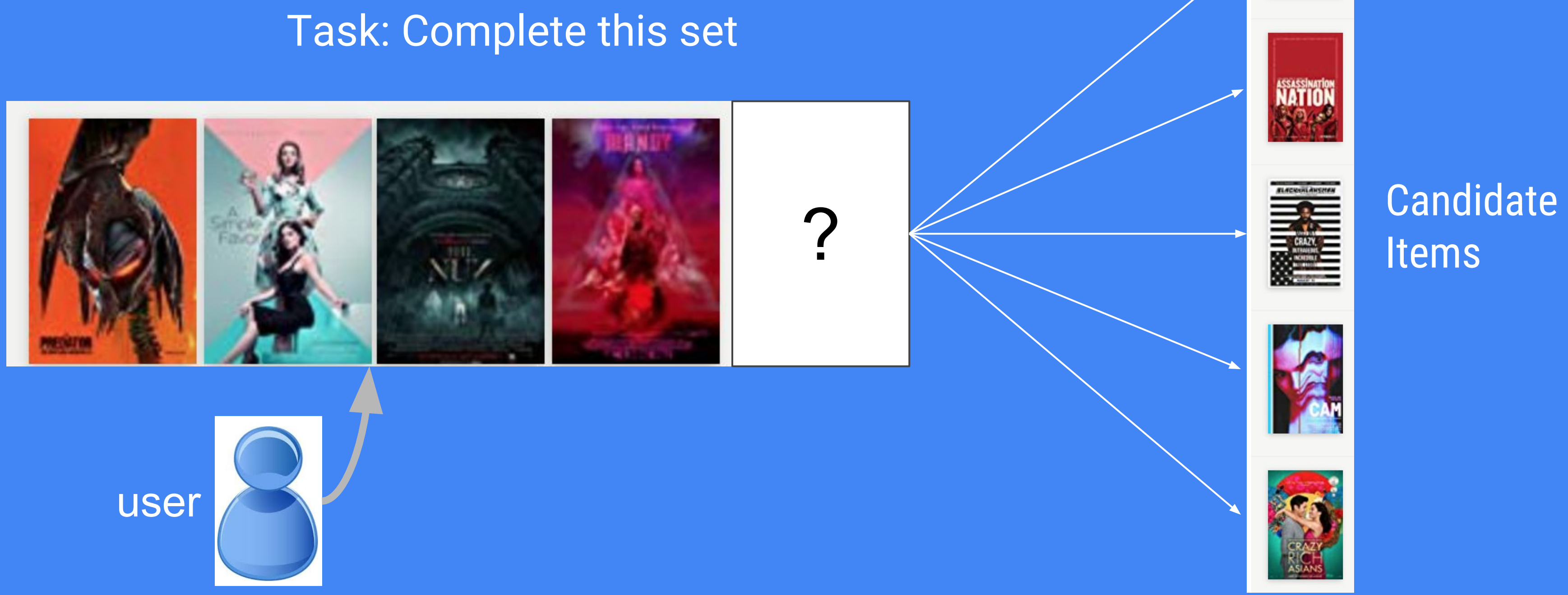
Categorical-Attributes-Based Item Classification for Recommender Systems

Qian Zhao (Bloomberg L.P. *)
Jilin Chen, Minmin Chen, Sagar Jain, Alex Beutel, Francois Belletti, Ed H. Chi (Google Inc.)

* Work done while interning at Google Inc.

Multi-Class Classification for Recommendation



Problem

- The possible number of classes or items is huge.
- Expensive softmax normalization across the whole catalog of items
    - Normalization goal is to compute the negative log likelihood loss.

$$\widehat{p_a}(a_i = 1) = \frac{exp(f(s, o_i))}{\sum_{k=1}^{\alpha+1} exp(f(s, o_k))}$$

In Practice

- To speed up: use negative sampling.
  - Normalize across *sampled smaller set* of classes or items (with correction)

$$\widehat{p_a}(a_i = 1) = \frac{exp(f(s, o_i))}{\sum_{k=1}^{\alpha+1} exp(f(s, o_k))}$$

Bengio, Y. and Senécal, J.S., *AISTATS 2003*

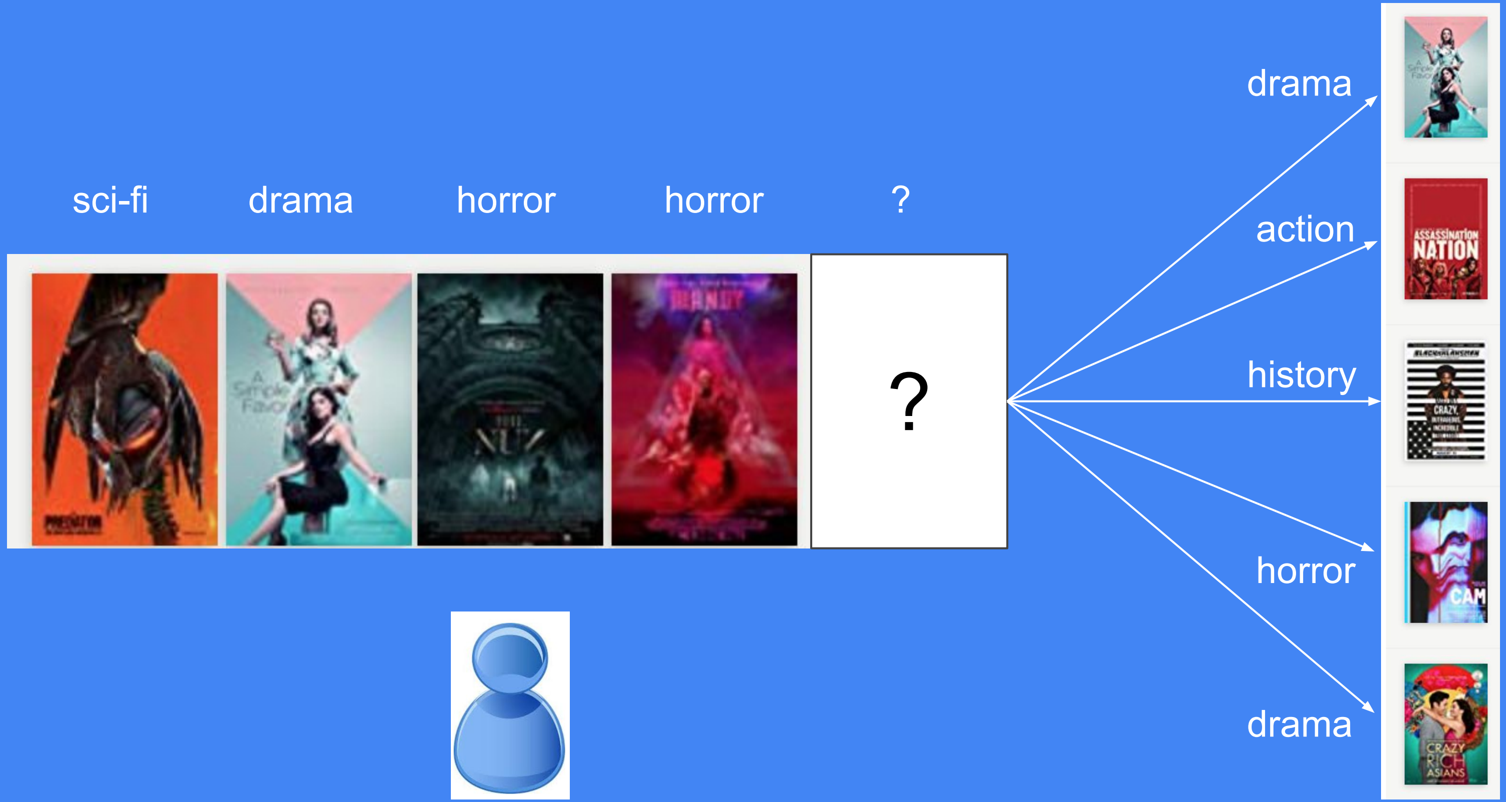
In Practice

- To speed up: use negative sampling.
  - Normalize across *sampled smaller set* of classes or items (with correction)

$$\widehat{p_a}(a_i = 1) = \frac{exp(f(s, o_i))}{\sum_{k=1}^{\alpha+1} exp(f(s, o_k))}$$

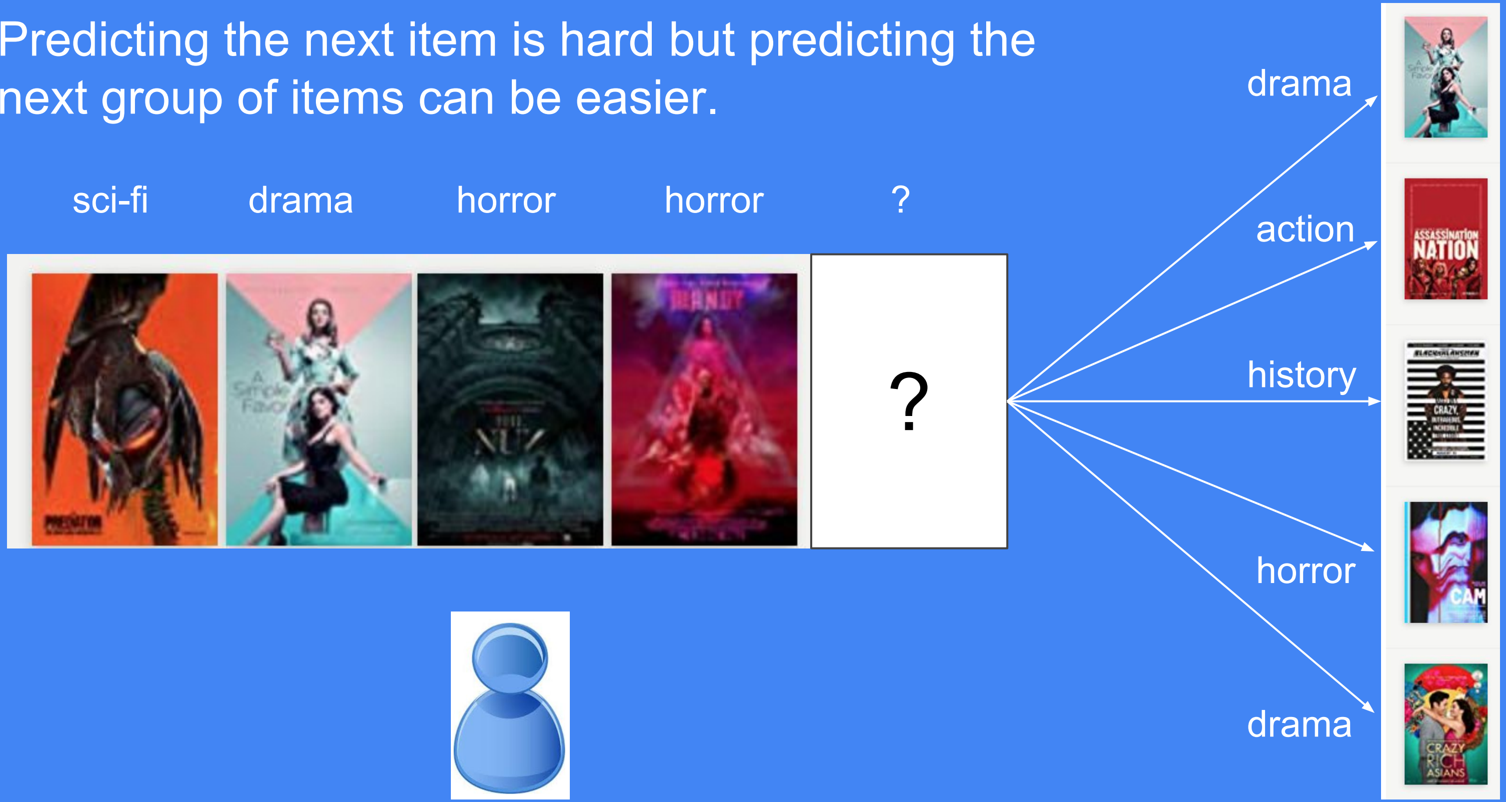
Can we improve on top of this?

Motivation: Utilizing Item Grouping Structure



Motivation: Utilizing Item Grouping Structure

Predicting the next item is hard but predicting the next group of items can be easier.



How to utilize this grouping structure?

Two Approaches:

- Multi-task learning (MTL)
- Hierarchical classification or softmax (HSM)

MTL


drama
action
history ...
horror
sci-fi
drama
horror
horror
topic softmax
item softmax
NATION
CAM
CRAZY RICH ASIANS

MTL

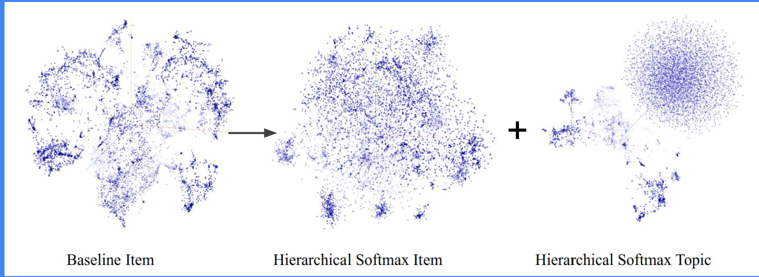
Improving user embedding model?

HSM

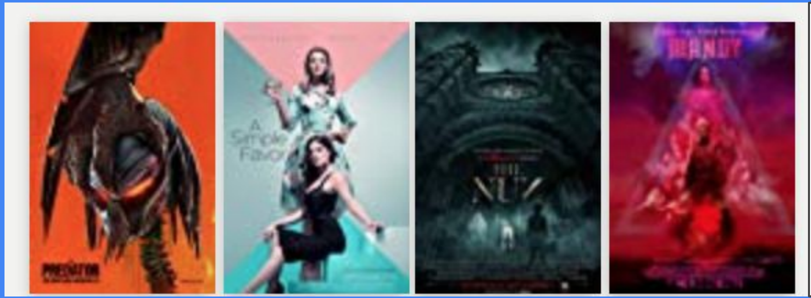
sci-fi
drama
horror
horror
topic softmax
drama
action
history
horror
...
within-topic item softmax

HSM

Reducing one harder task to two simpler tasks?



sci-fi drama horror horror



topic softmax

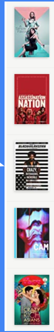
drama

action

history

horror

within-topic item softmax



...

Experiments

Data Set 1: Public Behance Data Set (He et al. RecSys'16)

Statistic	Size	Summary	Size
#users	63,497	min	1
#items	178,788	25%	1
#owners	51487	median	2
#appreciations	1M	75%	4
		max	153

Attribute

Data Set 2: Proprietary Large-Scale Data Set

Statistic	Size	Summary	Topic Size	Publisher Size
#users	Hundreds of Millions	**min**	1	1
#items	2M	**25%**	1	1
#topics	600K	**median**	1	1
#publishers	800K	**75%**	3	2
#consumptions	Hundreds of Billions	**max**	38K	3.5K

Results

Model	Item MAP@5
SVDFeature	0.0035
SVDFeature+MTL	0.0044 (+25.7%)
SVDFeature+HSM	**0.0046 (+31.4%)**
RNN	0.0099
RNN+MTL	0.0104 (+5%)
RNN+HSM	**0.0129 (+30.3%)**

Model	Attribute	Item MAP@5
RNN	N.A.	0.151
RNN+MTL	Topic	0.163 (+7.9%)
	Publisher	0.156 (+3.3%)
RNN+HSM	Topic	**0.184 (+21.8%)**
	Publisher	0.182 (+20.5%)

Summary:
MTL and HSM better than both SVDFeature and RNN alone.
But HSM has larger gain.

Results

Model	Item MAP@5	Attribute MAP@5
SVDFeature	0.0035	N.A.
SVDFeature+MTL	0.0044 (+25.7%)	0.020
SVDFeature+HSM	**0.0046 (+31.4%)**	0.025
RNN	0.0099	N.A.
RNN+MTL	0.0104 (+5%)	0.027
RNN+HSM	**0.0129 (+30.3%)**	0.024

Summary:
Predicting item groups (attribute) is indeed easier than predicting items.

Results

Model	Attribute	Item MAP@5	Attribute MAP@5
RNN	N.A.	0.151	N.A.
RNN+MTL	Topic	0.163 (+7.9%)	0.336
	Publisher	0.156 (+3.3%)	0.293
RNN+HSM	Topic	**0.184 (+21.8%)**	0.337
	Publisher	0.182 (+20.5%)	0.295

Summary:
Predicting item groups (attribute) is indeed easier than predicting items.

What if the grouping structure is noisy?

- Does the benefit of HSM still hold?

Results

Model	Noise	Item MAP@5
SVDFeature	N.A.	0.0035
SVDFeature+HSM	0.0	0.0046 (+31.4%)
	0.1	0.0045 (+28.5%)
	0.2	0.0047 (+34.2%)
	0.6	0.0038 (+8.5%)
	1.0	0.0029 (-17.1%)

Model	Randomization	Item MAP@5
RNN	N.A.	0.151
RNN+HSM	0.0	0.182 (+20.5%)
	0.1	0.168 (+11.2%)
	0.2	0.166 (+9.9%)
	0.6	0.152 (+0.6%)
	1.0	0.150 (-0.6%)

Summary:
HSM is robust to noisy grouping structure, but purely random grouping doesn't improve.

Does this help Cold-start items?

H1: Does HSM have more improvement for the long-tail items' prediction?

Does this help Cold-start items?

H1: Does HSM have more improvement for the long-tail items' prediction?

Yes!

Different user generative models?

H2: Does the advantage of HSM generalize across different types of generative models of users?

Two Types of User Generative Models

- Single-Level
- Two-Level

#generating the item according to user attribute interest.

if *type is Single-Level* **then**

$a \in Z_n \sim_n Softmax(Wu)$

$c \in Z_n$, where $c_j = C_{a_j}, j = 1 : n$

end

else if *type is Two-Level* **then**

$c \in Z_n \sim_n Softmax(Vu)$

$a \in Z_n$, where $a_j \sim Softmax(W_{c_j} u)$, W_{c_j} is the embeddings of the items involved by $c_j, j = 1 : n$.

end

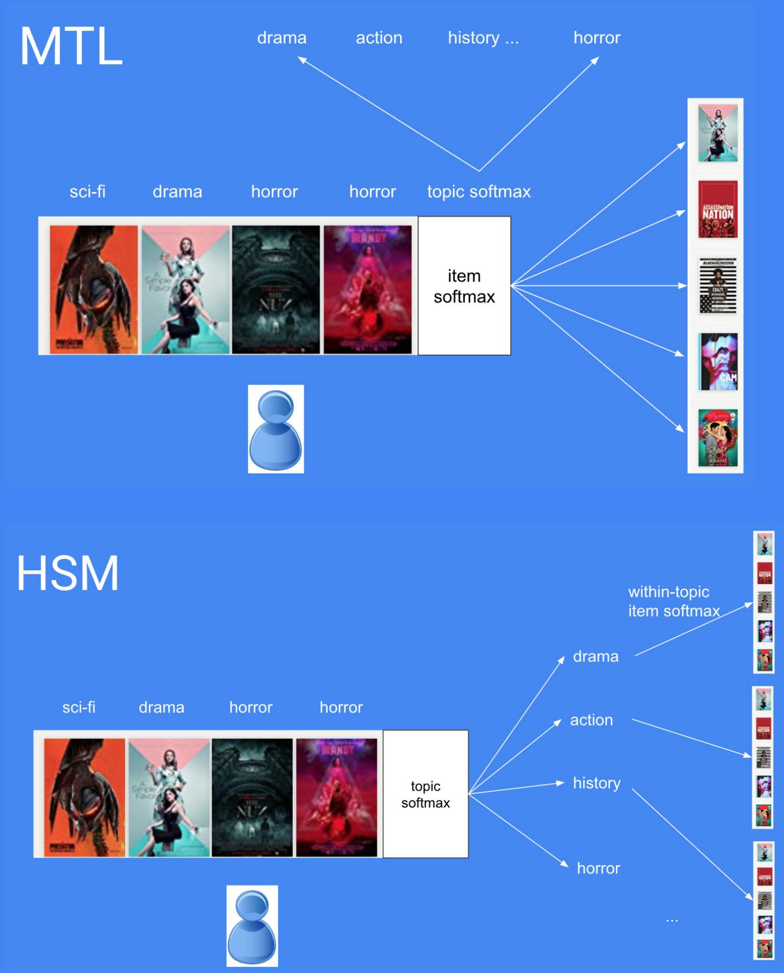
Results

Data Set	Model	AUC
Single-Level	SVDFeature	0.592
	SVDFeature+HSM	0.668 (+12.8%)
Two-Level	SVDFeature	0.652
	SVDFeature+HSM	0.705 (+8.1%)

Summary:
HSM improves in both cases!

Summary

The grouping structure of item categorical attributes can be used to improve the multi-class classification model accuracy in recommender systems through MTL and especially HSM.



Implications

- For user-centric researchers
  - Multi-level human preference and decision making process?
- For practitioners
  - Dynamic list/slate recommendation
  - Interactive preference elicitation or conversational recommendation

Acknowledgments

Thanks to

- Google Inc. for supporting this work with an internship
- Bloomberg L.P. for supporting my traveling here

Thanks! Questions?

“Categorical-Attributes-Based Item Classification for Recommender Systems”

Qian Zhao (now at Bloomberg L.P.), Jilin Chen, Minmin Chen, Sagar Jain, Alex Beutel, Francois Belletti, Ed H. Chi (Google Inc.)

Contact:

- jilinc,minminc@google.com
- qzhao101@bloomberg.net (http://qianzhao.me)