Cold Start Thread Recommendation as Extreme Multi-label Classification

April 23, 2018

Kishaloy Halder, Lahari Poddar, Min-Yen Kan

XMLC for Social Media, Lyon, France

NUS National University of Singapore

School of Computing

WING

Web Information Retrieval
Natural Language Processing Group

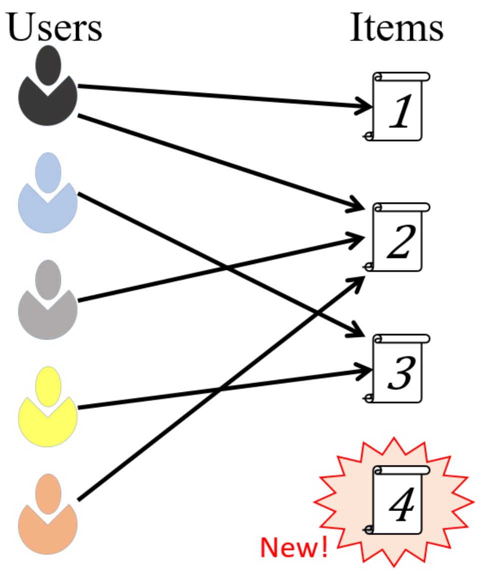
THE WEB CONFERENCE

Cold Start Thread Recommendation

- New threads/contents are created continuously in Web2.0 applications
  - Threads in discussion forums, questions in community question answering platforms, Social Media posts and so on
- To increase visibility of a new thread, the platforms need to ensure that the members find questions relevant to their interests
- **Task**: Recommend *newly created threads* to potentially interested users in order to get them answered
- In recommendation literature, this is known as **cold-start** problem

Cold Start Item Recommendation

- Typically user and item are represented as vectors in latent factor models
  - i^{th} User → $\mathbf{u_i}$
  - j^{th} Item → $\mathbf{v_j}$
- Predicted recommendation is obtained by,
  - $r_{ij} = \mathbf{u_i} . \mathbf{v_j}^T$



Interaction Graph

	1	2	3	4
	1	1	0	?
	0	0	1	?
	0	1	0	?
	0	0	1	?
	0	1	0	?

Interaction Matrix

For New Item j = 4:

- $\mathbf{v}_{j=4}$ is randomly initialized
- Rating for it can not be predicted for any user

Revisiting Cold Start as XMLC

- In absence of interaction history for a newly created thread, traditional recommendation systems suffer

- Need to use the textual content of a thread in order to find potentially interested users.

- Can be viewed as an Extreme Multi-Label Text Classification problem
  - Existing users → Class labels
  - Out-of-matrix thread recommendation → multi-label classification

Extreme Multi-label Text Classification

- Number of labels are “extremely” high i.e.,
  - Thousands, or even more
  - Typically used for tag prediction – wiki pages, amazon products

- Multi-Label Classification Models
  - Embedding based Method: SLEEC (NIPS ‘15)
  - Tree based Method: FastXML (KDD ‘14)
  - Deep Learning based Method: XML-CNN (SIGIR ‘17)
    - State-of-the-art for XMLC!

Our Approach

- Propose a neural network to predict the subset of users interested in a new thread from the extremely large set of users in the forum community

- Textual content is encoded to a lower dimensional space
  - Word embedding: maps words to vectors
  - Bi-directional GRUs: encodes sequence of words

- A universal encoding of a post text might not be enough
  - Different users have different interests

Challenges

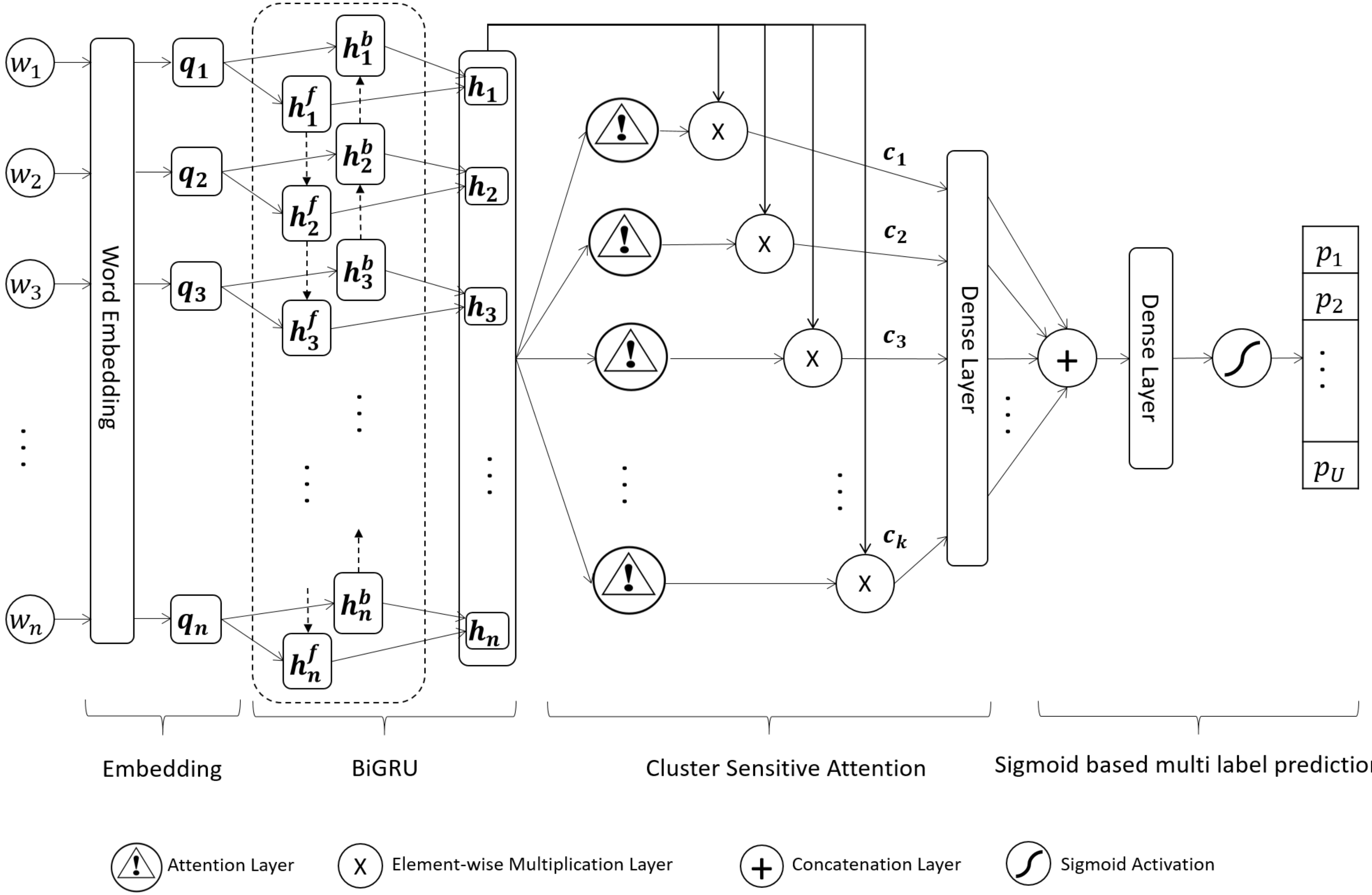
"I have been recommended to undergo tracheotomy and put in a PEG. I am wondering how many days I'll have to stay in the hospital? Will I have a hard time adjusting afterwards? Does the hose need to be connected while transferring? Will the equipments take up a lot of room? How do you call for help?.."

- The post contains diverse questions – different parts of it could potentially be answered by different users

Cluster Sensitive Attention

- Attention mechanism: Effective in capturing important parts of the text
  - Gives weights to words of post
  - Post encoding: weighted sum of word encodings
- Separate attention for every user: not scalable due to huge number of parameters
- Hypothesis: Clusters of users exist who are interested in similar items
- Cluster sensitive attention on textual content
  - N users, K clusters where K << N
  - K attention layers
- Each attention layer captures cluster-specific preferences

Overall Architecture



Experiments - Datasets

- Have experimented with 4 forum datasets from multiple domains
  - Online Health Forum: Epilepsy, ALS, MS
  - Stackoverflow

Dataset	#users	#threads		Avg #word in thread		Avg #user per thread		Sparsity
		train	test	train	test	train	test	
1. Epilepsy	1506	1644	412	147	168	7.39	9.29	99.49%
2. ALS	3182	6466	1617	148	135	9.85	9.75	99.69%
3. Fibromyalgia	5669	8576	2144	203	233	9.02	9.14	99.84%
4. Stackoverflow	69,631	20,137	5035	93	99	6.81	7.29	99.99%

- Metrics: Recall@M, nDCG@M, MRR

Experiments - Baselines

- CVAE: Collaborative Variational Auto Encoder (KDD’17)
- CTR : Out-of-matrix setting (KDD ‘11)
- CNN-KIM: CNN based Text classifier (EMNLP ‘14)
- XML-CNN (SIGIR ‘17)
- Bi-GRU2: Our Model without cluster sensitive attention

Experiments – Results (MRR)

Dataset	Methods					
	CVAE	CTR	CNN-Kim	XML-CNN	BiGRU-2	Our Model
1.Epilepsy	0.159	0.443	0.536	0.551	0.631	**0.671**
2.ALS	0.201	0.275	0.270	0.293	0.297	**0.306**
3.Fibromyalgia	0.304	0.435	0.669	0.668	0.740	**0.773**
4.Stackoverflow	0.003	0.032	0.025	0.029	0.047	**0.050**

Our model outperforms the baselines in all cases

Experiments – Results (Recall@M)

Dataset	Metric	Method					
		CVAE	CTR	CNN-Kim	XML-CNN	BiGRU-2	Our Model
1. Epilepsy	recall@5	3.69	17.46	17.23	**22.76**	22.64	22.65
	recall@10	7.22	27.67	22.93	**34.67**	29.22	29.26
	recall@30	21.14	43.83	44.63	49.08	50.99	**51.21**
	recall@50	29.62	50.86	52.45	53.69	59.47	**59.80**
	recall@100	42.44	59.93	65.77	63.67	68.23	**69.37**
2. ALS	recall@5	4.17	7.05	6.19	6.51	7.63	**9.23**
	recall@10	7.07	12.08	10.09	11.44	**14.65**	13.89
	recall@30	17.04	25.00	22.15	23.56	30.18	**31.84**
	recall@50	24.07	32.46	31.27	30.61	36.32	**36.55**
	recall@100	35.77	44.14	43.82	43.14	48.14	**49.78**
3. Fibromyalgia	recall@5	8.24	14.58	23.01	22.11	25.63	**25.97**
	recall@10	14.93	27.18	34.77	33.88	35.18	**37.38**
	recall@30	32.83	54.39	58.04	61.83	62.39	**63.06**
	recall@50	42.43	63.91	67.83	68.92	69.17	**72.04**
	recall@100	55.02	72.31	76.37	75.74	77.98	**78.19**
4. Stackoverflow	recall@5	0.02	0.59	0.46	0.51	0.66	**0.86**
	recall@10	0.06	1.14	0.73	0.97	1.15	**1.30**
	recall@30	0.16	2.73	1.84	2.42	**2.94**	2.80
	recall@50	0.31	4.02	2.74	3.43	4.03	**4.11**
	recall@100	0.69	**6.36**	4.43	5.35	6.09	6.33

- Our Model outperforms baselines in most cases
- Scores at smaller M are not important
  - A new content is targeted to a much larger audience by common practice
- The cluster sensitive attention boosts performance

Conclusion

- The age-old cold start problem can be seen and solved as an Extreme Multi-label Classification problem
- A cluster sensitive attention mechanism can capture user groups with similar preferences, and it helps with addressing scalability as well
- Our method outperforms traditional state-of-the-art recommendation, and other XMLC approaches for this task

Thanks for listening!

kishaloy@comp.nus.edu.sg
lahari@comp.nus.edu.sg