Predicting Patients' Conversation Transitions in Online Health Support Groups

Anusha Prakash, Srividya Potharaju

Advisor: **Prof Eduard Hovy w/ Diyi Yang**

Carnegie Mellon University
School of Computer Science

Language Technologies Institute

Agenda

Carnegie Mellon

- Introduction and Motivation
- Problem Definition
- CSN
- Approach Overview
- User Profile Creation
  - Handcrafted Approach
  - Attention Based Approaches
- User Matching Mechanism
- Results
- Analysis and Visualization
- Conclusion and Future Work
- Key Takeaways
- Q & A

Online Health Support Groups

Carnegie Mellon

- Platform where people can connect with others suffering from similar issues
- Hosts Discussion Boards and Chat Rooms
- Hope is that people can find strength and inspiration from each other

Medium of Communication used by Users :

1. Thread Creation and Commenting on public discussion boards
2. Personal Message Exchanges

Significance of Conversation Medium

Medium used is a conscious choice - influenced by various factors - content of discussion, nature of the user in question, social role of the user (caregiver, care-seeker etc.), health background of the user at that point of time etc.

Helps model users’ social relations and concerns

Helps to recommend potential users and helpers to connect

Predicting Patients' Conversation Transitions

“We address the task of predicting patients' conversation transitions, i.e., predicting whether two users will move their conversations to the private chat based on their textual communication in the public discussion board, thereby also making recommendations for people to connect privately”

American Cancer Society

Cancer Survivors Network

Carnegie Mellon

CSN

- Discussion Boards
- Announcements
- Member Resource library
- CSN Chatroom

cancer.org

- Cancer Information
- Community Resources
- Making Strides Against Breast Cancer
- ACS News
- Caregivers
- After Treatment
- In Treatment
- Rides To Treatment
- Lodging
- Hair Loss and Mastectomy Products
- Breast Cancer Support
- Look Good...Feel Better

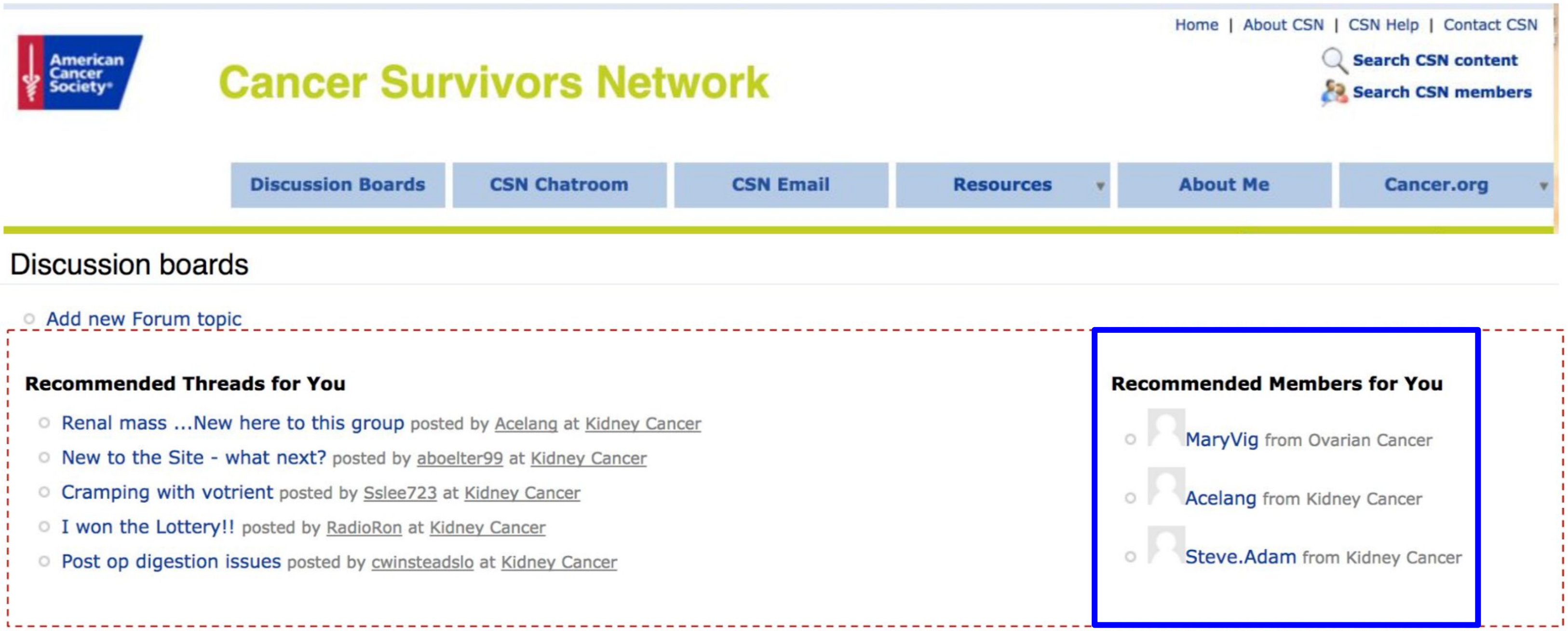
CSN Home

Discussion boards

- Log in to post new content in the forum.

Forum	Topics	Posts	Last post
Cancer specific			
Please remember that these discussion boards are a public forum, which means open to the public (i.e. non-CSN members) and the content can be found via internet search engines. Members are strongly advised not to share personal identifiers such as real names, email address, telephone, street address, etc. can be used to identify you and link you to the content you provide. Other areas of CSN are restricted to members only and cannot be found by search engines.			
			By jacobmom
Brain Cancer	1557	10755	Apr 29, 2018 - 9:32 am By Josey
Breast Cancer	27708	335891	Apr 30, 2018 - 6:51 am By pamelamasterson
Childhood Cancers	365	1913	Apr 25, 2018 - 11:24 am By jnickele
Colorectal Cancer	25991	284686	Apr 30, 2018 - 2:32 pm By ThomasH
Esophageal Cancer	4424	34245	Apr 26, 2018 - 3:09 pm By paul61
Gynecological Cancers (other than ovarian and uterine)	1077	6680	Apr 23, 2018 - 7:27 am By RobLee
Head and Neck Cancer	11653	141710	Apr 30, 2018 - 3:09 pm By GavinP
Kidney Cancer	5308	64702	Apr 30, 2018 - 11:33 am By CRashster
Leukemia	612	2953	Apr 16, 2018 - 3:10 pm By Mattymix

Recommending Threads & Members



CSN Data Statistics

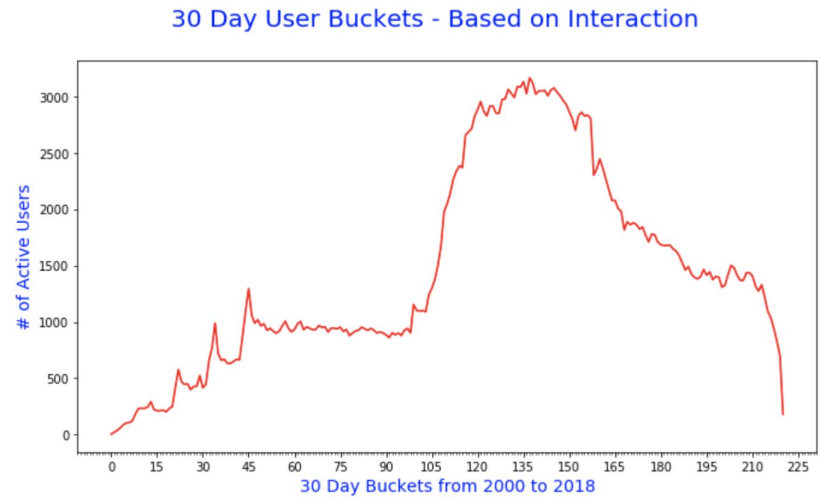
Attribute	Value
# of Active Users	65260
# of Threads	142211
# of Comments	1090683
Avg # of Threads created by each User	3.67
Avg # of Comments by each User	22.2
# of Users who have created New Threads	38675
# Avg Body Length of Comments (words)	87.48
# Avg Body Length of Thread (words)	172.66

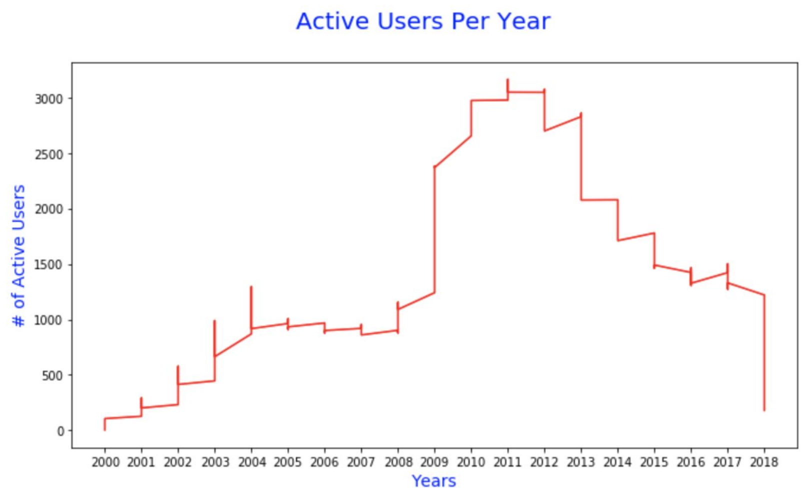
Approach - Overview

Carnegie Mellon

- We frame this task as a binary class classification problem where given two users' public profile, we predict if they are likely to have a private conversation
- We use information from personal chats data as our ground truth
- During training, we create **dynamic, temporal based** user profile based on the users' discussion in public forum
- We later take the difference of the user profiles thus created, pass it through a feedforward neural network to predict the output

Temporal User Bucketing





We calculated

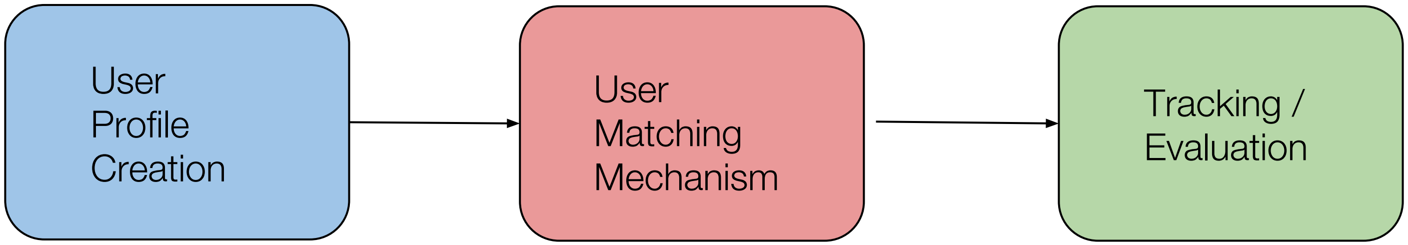
- Users active in each Time Bucket
- The Time Buckets in which a User is active
- Interactions occurring in each Time Bucket

Dynamic and Temporal User Profiling

uid	comment	thread_id	timestamp	
12	c_1	th_1	123476	Train data (80%)
	c_2	th_2	176549	
			: :	
	c_8	**th_i**	**190654**	
	c_9	th_j	201345	Dev data (10%)
	c_{10}	th_m	203569	Test data (10%)

- User profile for Dev data and Test data is sealed at timestamp of 8th comment i.e., (80% th comment)

Prediction System Components



User Profile / Fingerprint Creation

Carnegie Mellon

Created by combining one or more of the following properties :

- Metadata Information
- Content Information

Two approaches for User Profile Creation :

1. Handcrafted Based Approach
2. Attention Based Approaches

Metadata Information

16 Metadata Fields - Gender, Marital Status, Income, Race, Cancer Type, Diagnosis Date, Insurance Status etc.

Metadata available is very sparse :

- 52.7% Users have metadata
- On an average, 3/16 metadata filled per user

Conclusion : Not used due to extreme sparsity and not much value

Only **Content Information** Used - Content of Threads and Comments

Profile Creation - Handcrafted Approach

Carnegie Mellon

Total of 7 Handcrafted features extracted from Content Information :

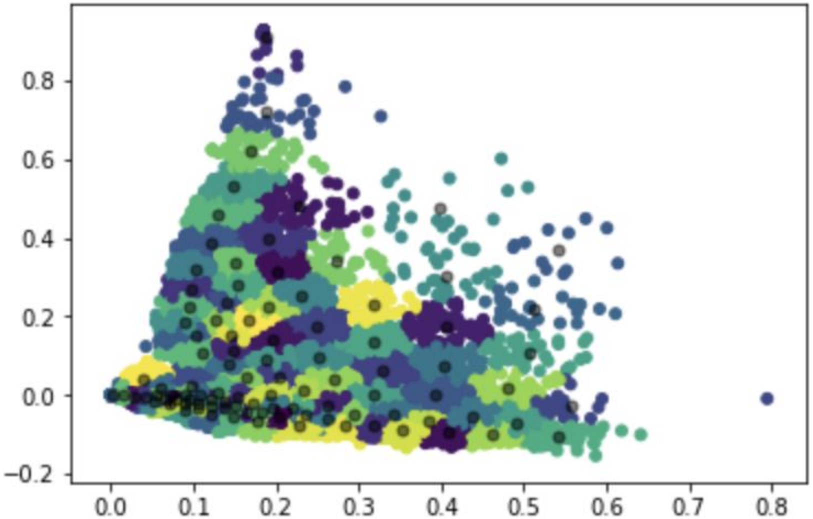
1. **Unigram TF-IDF - Top 3000 TF vocabulary words**

   Words - surgery, thorac, therapy, recovery, radioact, vomit, mastectom etc.

2. **Bigram TF-IDF - Top 3000 TF vocabulary words**

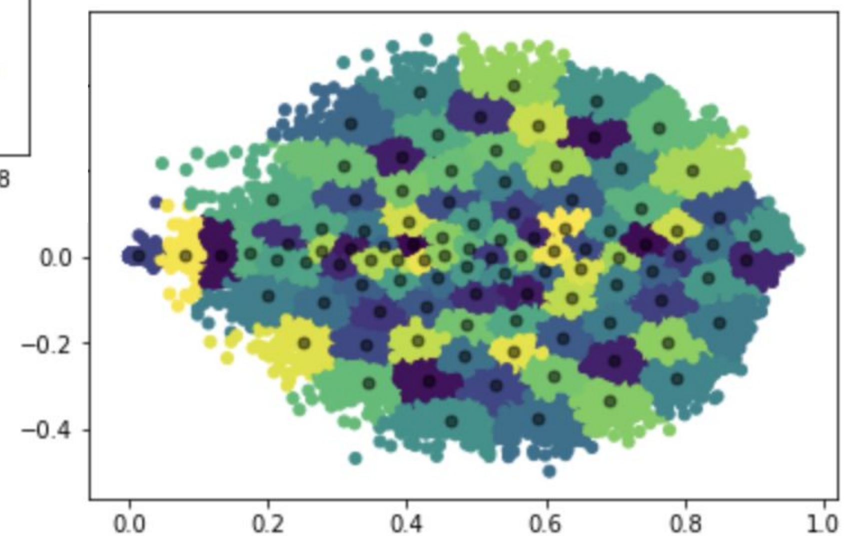
   Words - american cancer, back surgery, bad days, abdominal pain, chemo drugs, enlarged lymph, love prayers etc.

K-Means Cluster Visualization



Unigram User Clusters - K-Means

Bigram User Clusters - K-Means



Profile Creation - Handcrafted Approach

3. Knowledge Promoters

Has_url_feature

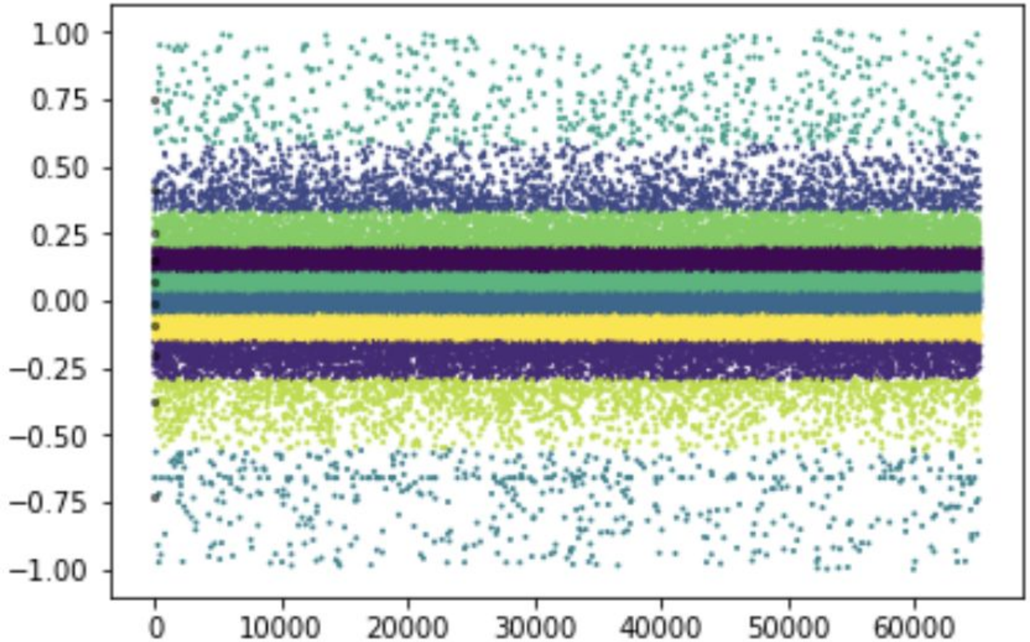
4. Networkers

Has_email
Has_phone

5. User Sentiment

- Vader Sentiment

Sentiment User Clusters - K-Means



Profile Creation - Handcrafted Approach

Carnegie Mellon

6. LDA Topic Modelling - Unigram Topics - 100

(40, '0.324*"iodin" + 0.171*"radioact" + 0.161*"thyroid" + 0.116*"yahoo" + 0.019*"send"')
(89, '0.381*"cyst" + 0.256*"complex" + 0.079*"marilyn" + 0.035*"ston" + 0.000*"pict"')
(74, '0.286*"im" + 0.092*"leukem" + 0.054*"doesnt" + 0.054*"follicul" + 0.046*"didnt"')
(53, '0.308*"pancrea" + 0.283*"gemz" + 0.109*"brand" + 0.101*"pancr" + 0.080*"oil"')

7. LDA Topic Modelling - Bigram Topics - 100

(16, '0.116*"lymph node" + 0.079*"lymph nodes" + 0.067*"pet scan" + 0.049*"scan showed" + 0.045*"high dose"')
20, '0.065*"triple negativ" + 0.041*"breast cancer" + 0.036*"clear cell" + 0.035*"radiation oncologist" + 0.026*"medical oncologist"')
(30, '0.048*"hair loss" + 0.045*"cancer years" + 0.041*"cancer doct" + 0.040*"high school" + 0.032*"go chemo"')
(60, '0.219*"thyroid cancer" + 0.064*"talk someone" + 0.054*"head neck" + 0.047*"neck cancer" + 0.043*"would appreciat"')

Profile Creation - Attention Based Approaches

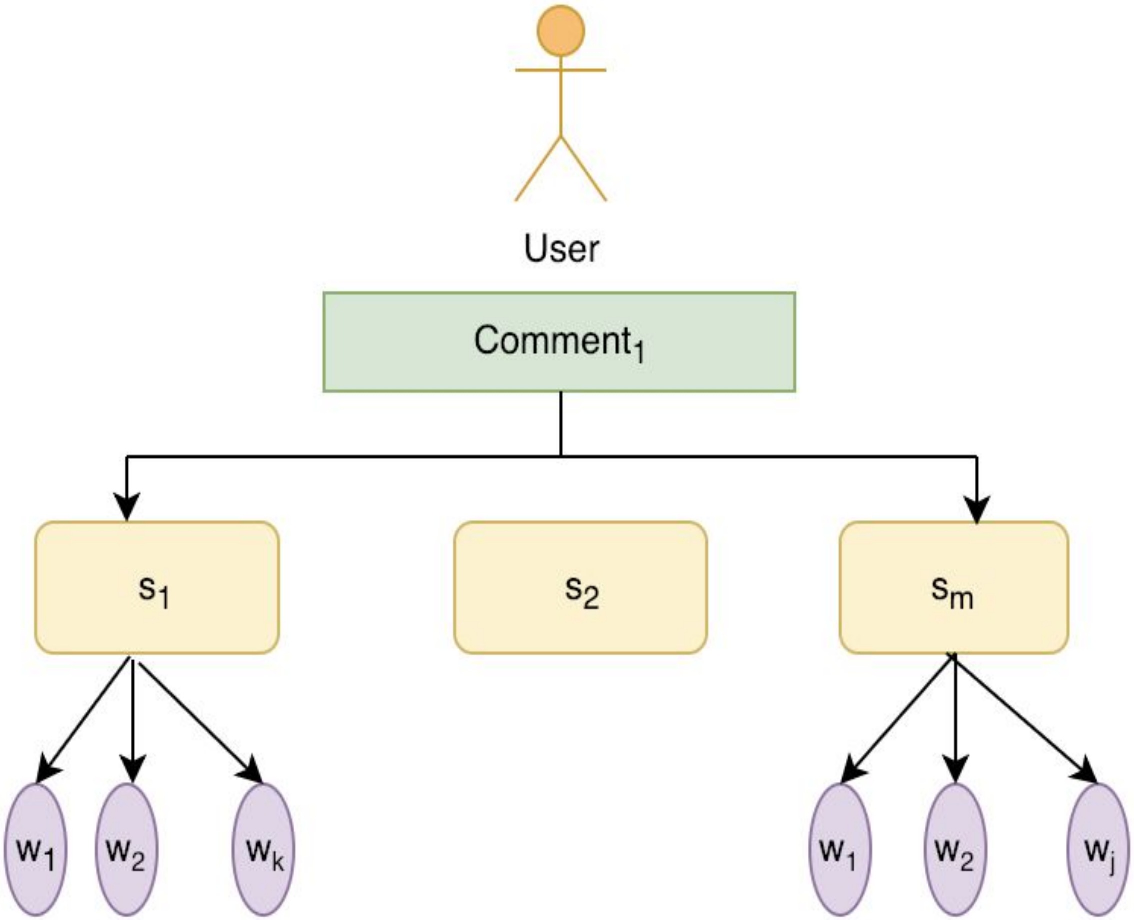
Carnegie Mellon

- All the sentences of the comments - encoded and various architectures are experimented with on these encoded sentences to create user profile :

1. Word level Attention
2. Comment level Attention
3. HAN (Hierarchical Attention Network)

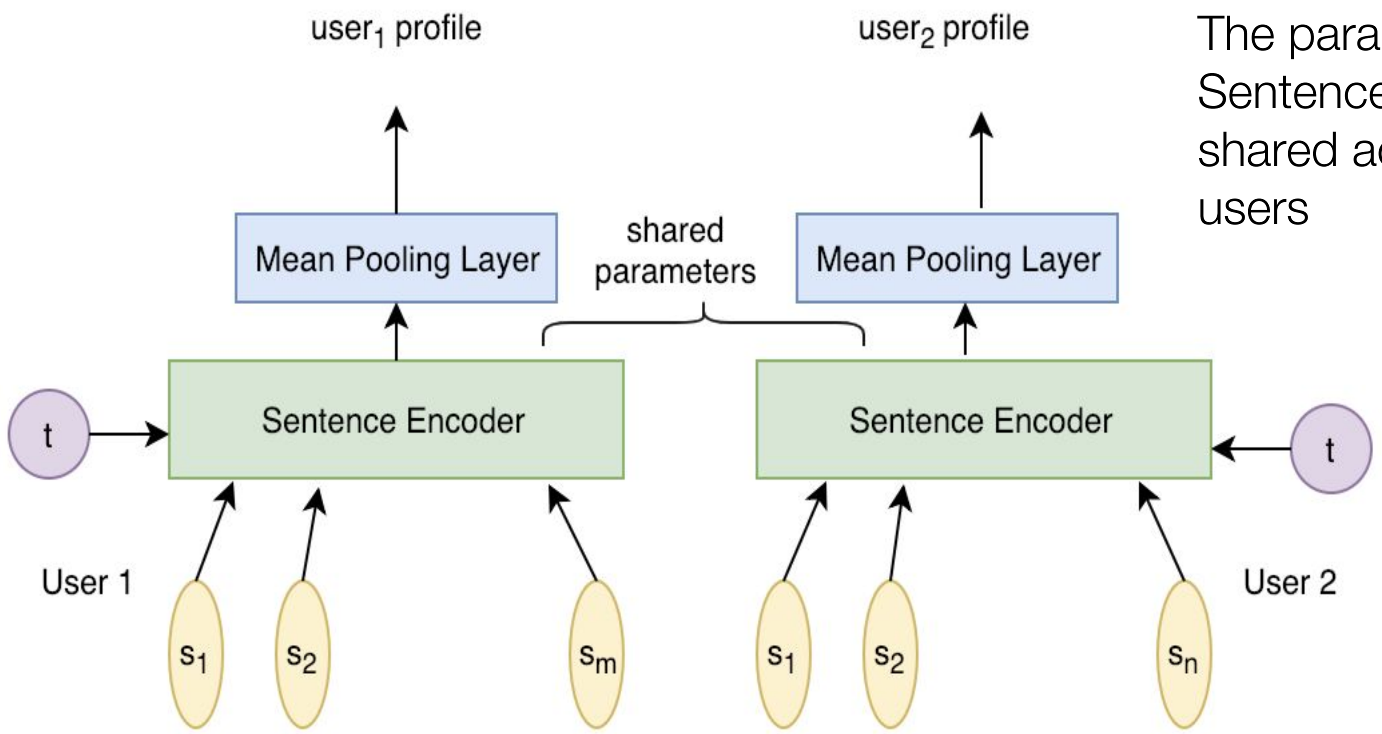
- We re-trained the Facebook's Infersent Model to encode sentences with 512 embedding dimension
- GRU - to obtain the comment and user encoding (all comments)
- Linear function is used for learning attention.

Public Content Structure



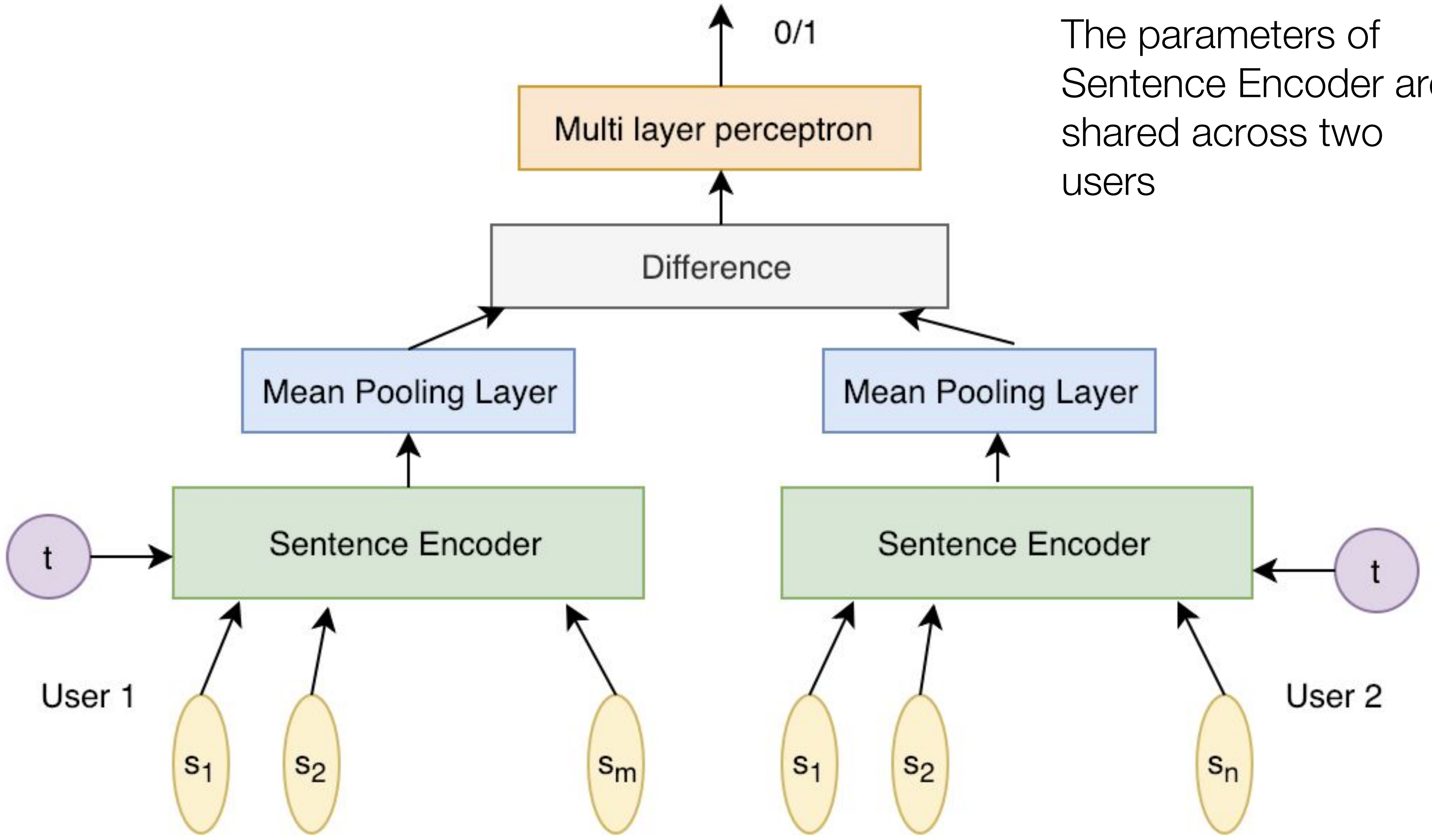
- User profile is built from comments
- s_i is a sentence in a comment c
- w_j is a word present in a sentence s

Word level Attention - Architecture



The parameters of Sentence Encoder are shared across two users

Word level Attention - Architecture

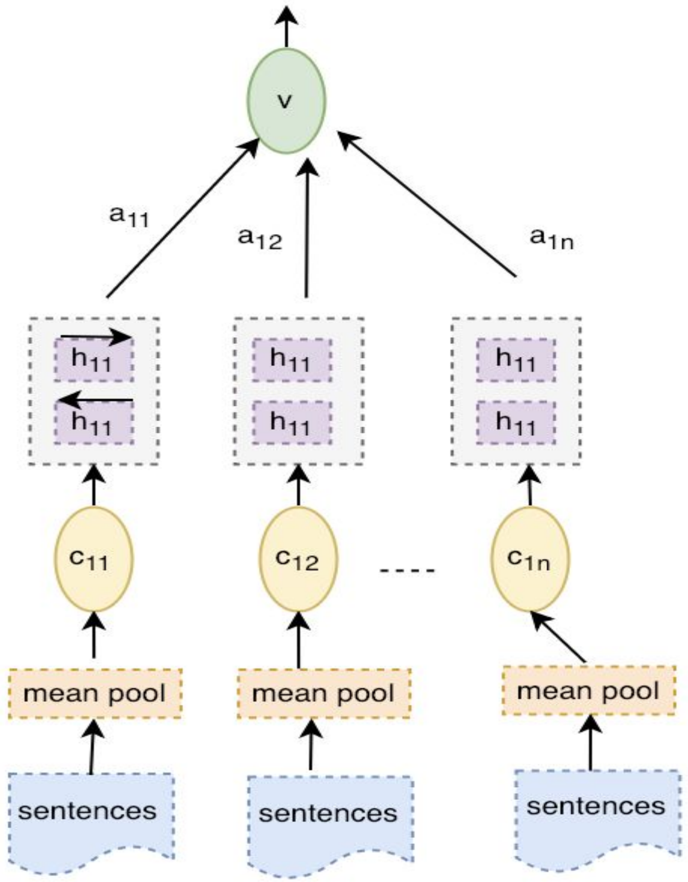


The parameters of Sentence Encoder are shared across two users

Comment level Attention

- Two level Attention

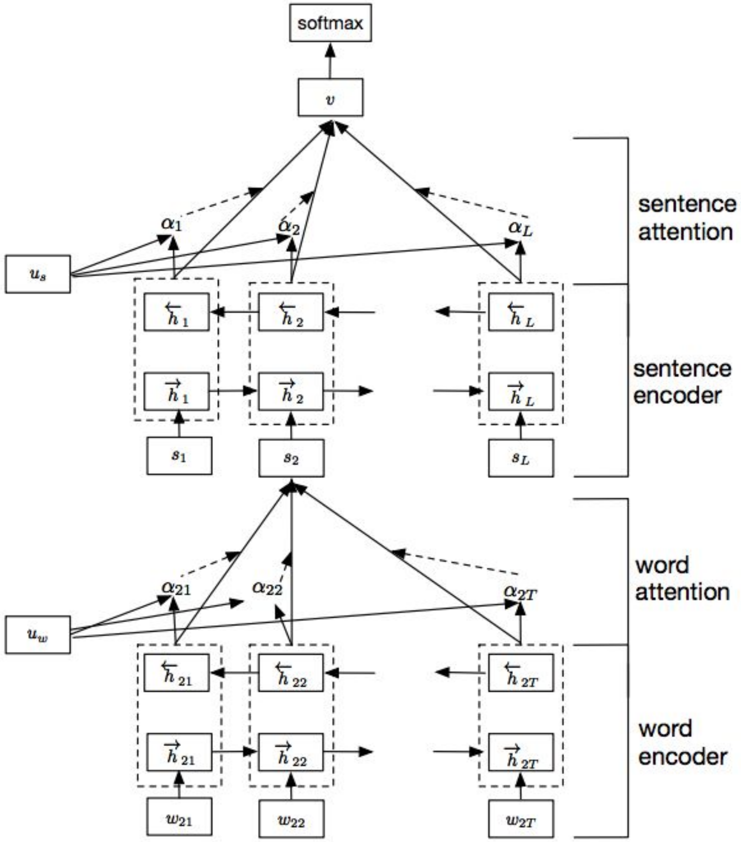
  Attention Over Words
  Attention Over Comments



Carnegie Mellon

- Two level Attention

  Attention Over Sentences
  Attention Over Comments

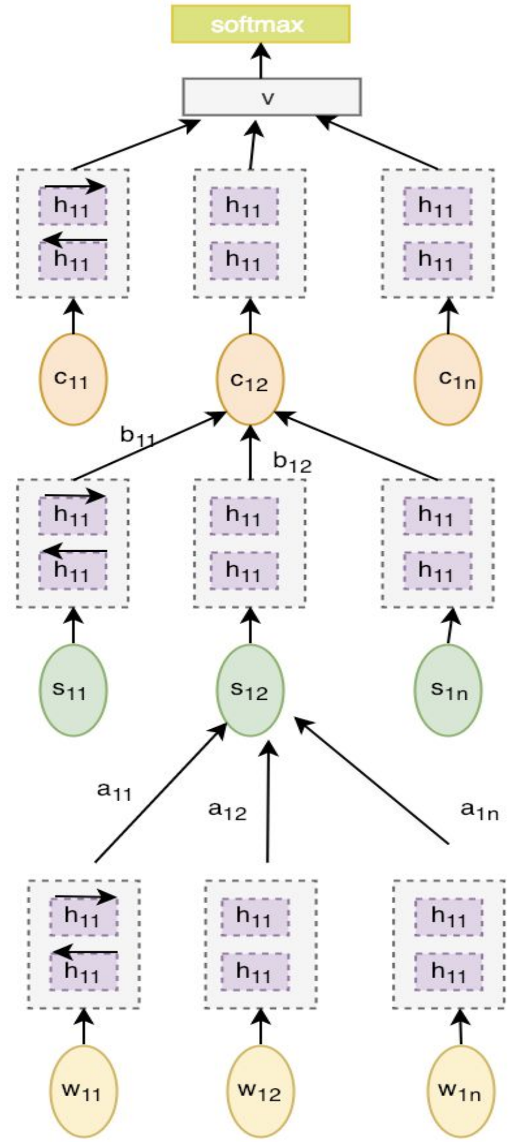


HAN [2]

HAN

- Three level Attention

  Attention Over Words
  Attention Over Sentences
  Attention Over Comments



ARE WE NOT WRITING ALSO ABOUT NARRE / DEEPCONN - I THINK WE SHOULD PUT SOMETHING - WITH THAT ARCHI DIAG - ELSE WILL SEEM LIKE WE DIDNT DO MUCH

- Too many architecture diagrams can confuse

Dataset for Training

Carnegie Mellon

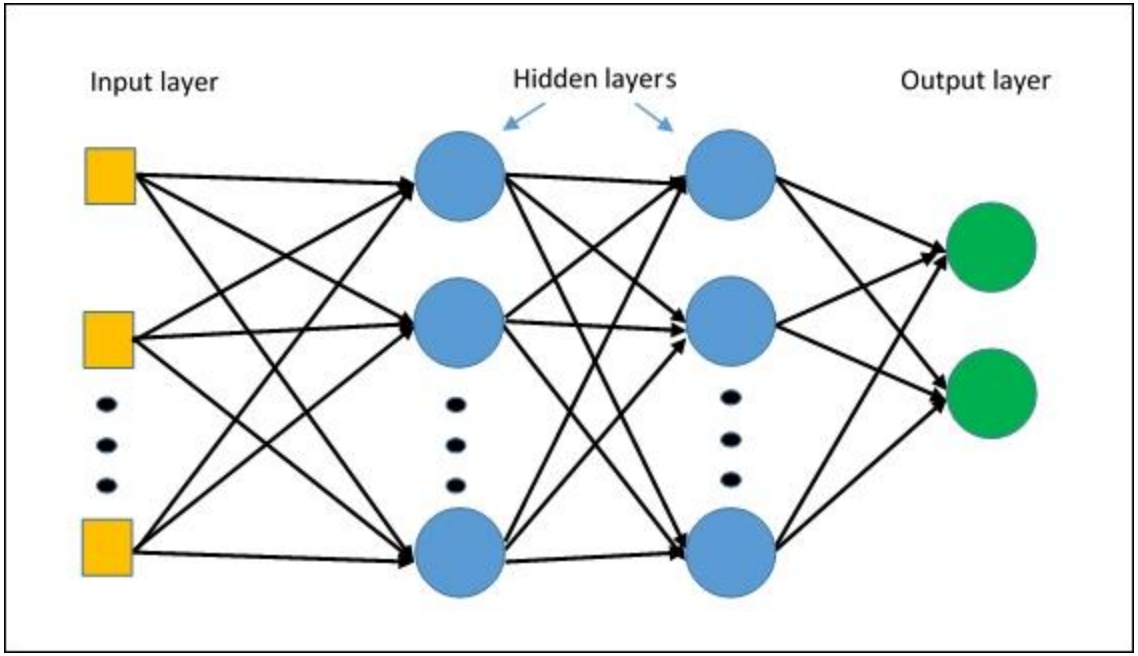
- Training data instance - (u_1, u_2, i), where u_1 and u_2 are the users' profiles and i is either 0 or 1, based on the personal chat ground truth
- Data set split - Train set (29338), Dev set (3668) and Test set (3668) - 80:10:10 ratio
- Only user pairs where each user has made between 5 and 500 comments used
- Subsampled from total of 106794 positive interaction pairs of users

User Matching Mechanism

Carnegie Mellon

Feed Forward Neural Network Approach

- (u1-u2) is used as the input
- Interleaved linear and leakyReLU layers with reducing dimension



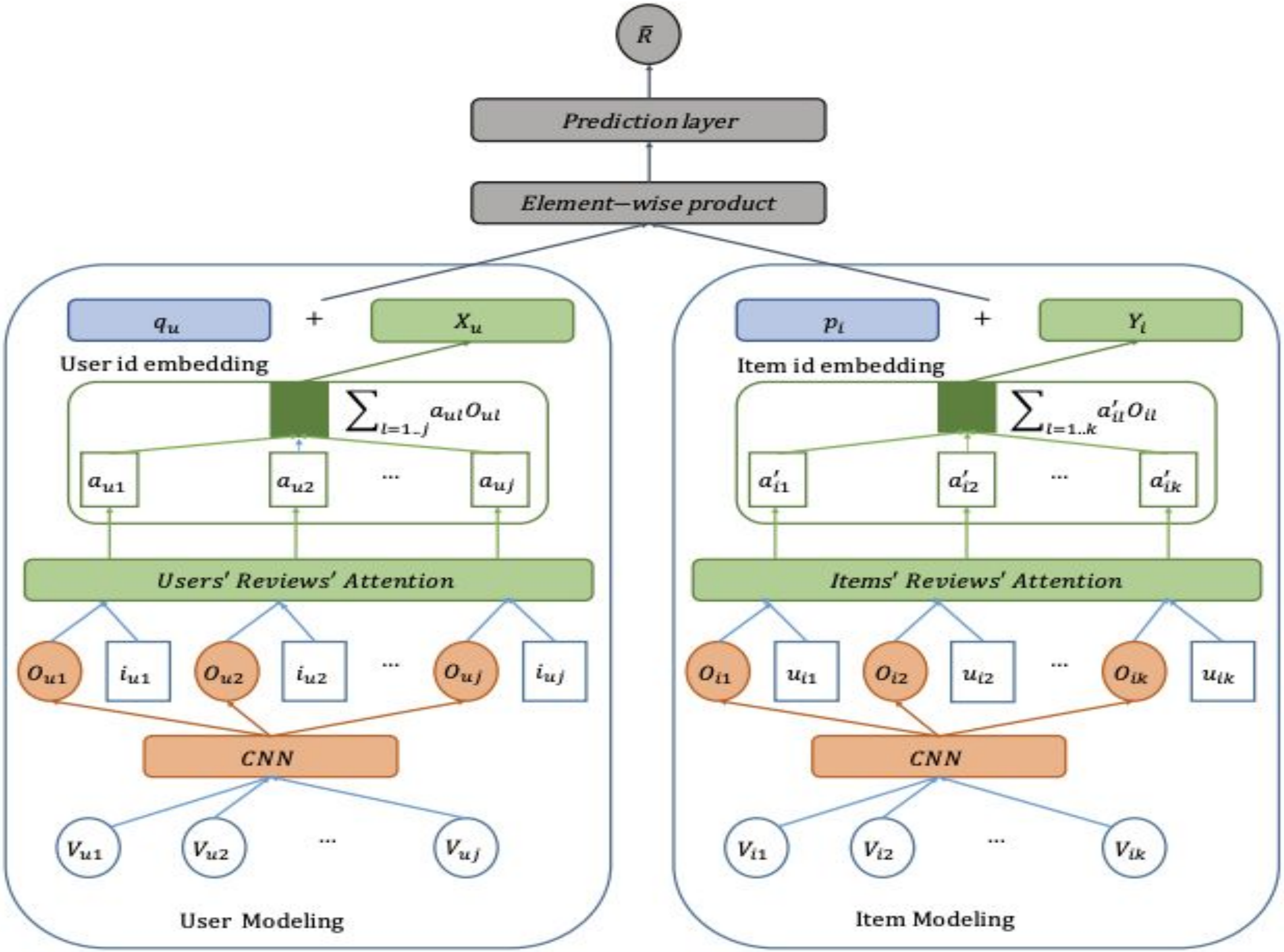
Results

Carnegie Mellon

Approach	Train Accuracy	Dev Accuracy	Test Accuracy
Word level Attention	**0.9266**	**0.8772**	**0.8869**
Comment level Attention	0.7042	0.7075	0.7193
HAN	0.7329	0.7199	0.7255
Handcrafted	0.9321	0.8791	0.8813

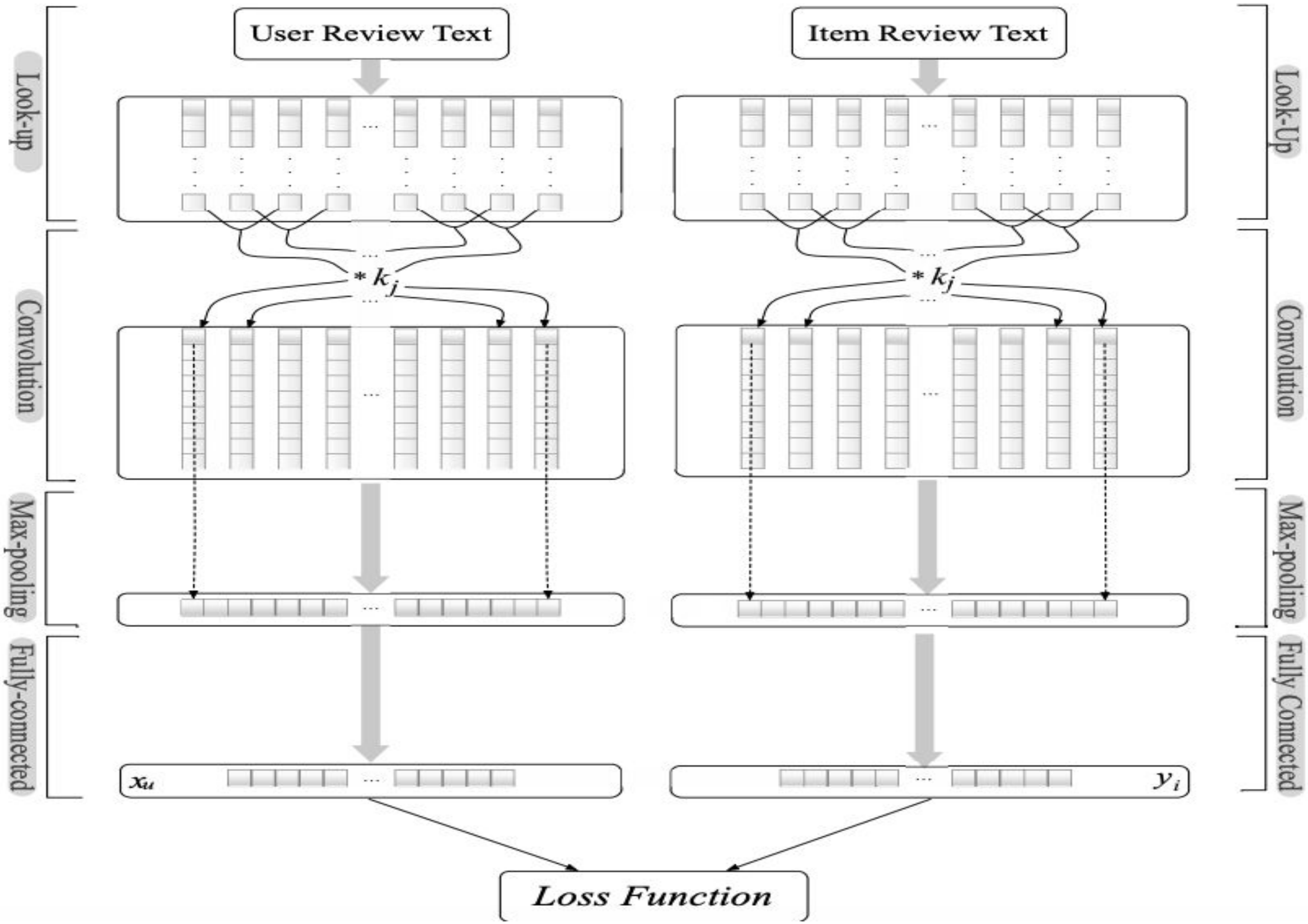
Word level attention performs best. User profiling with hand crafted features approach performs similar to the best model emphasizing the fact that categorizing users based on their topics, sentiment is indeed applicable in this scenario.

Other Architectures Experimented (NARRE)



Other Architectures Experimented (DeepCoNN)

Carnegie Mellon



Analysis

Carnegie Mellon

- We analyzed the predictions made by our model and visualized how the importance i.e., attention weight is distributed across different words in a sentence to get a representation.
- We noticed two major categories in which the model's predictions are interpretable

Users with similar journey scenario

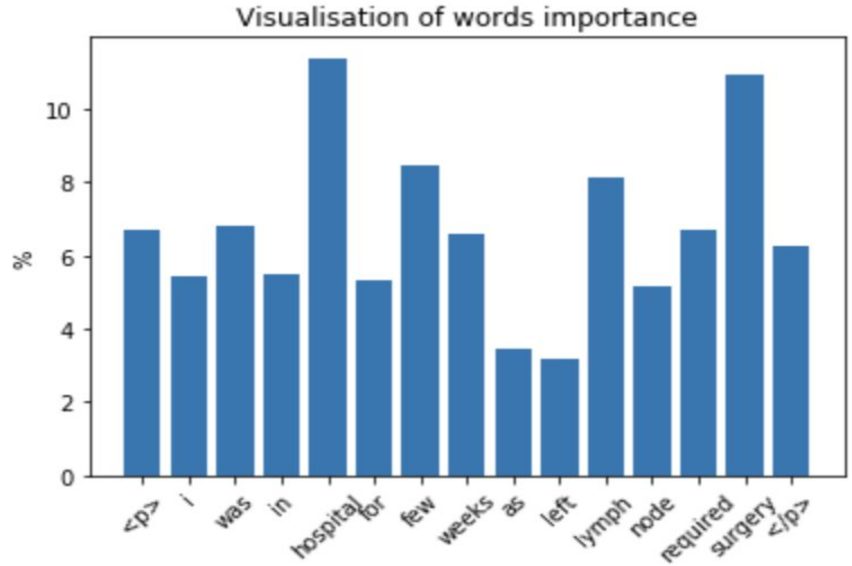
Carnegie Mellon

User 1 : “I was in hospital for few weeks as left lymph node required surgery.”

User 2 : “I too had surgery to remove more tissue and lymph nodes.”

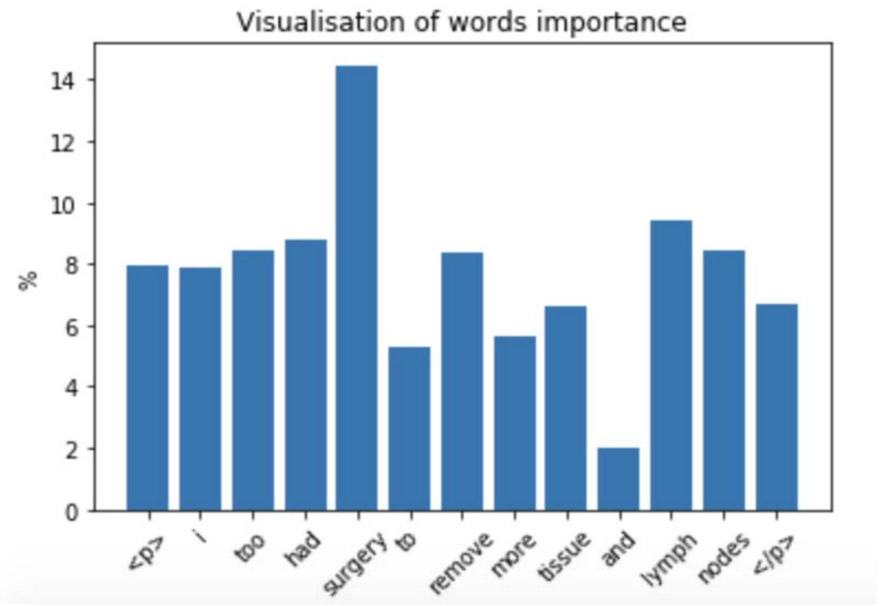
- This can be seen as an example of users who had undergone or who are undergoing similar surgery/therapy process

Similar User Attention Visualization



User 1

User 2



CareGiver - CareSeeker scenario

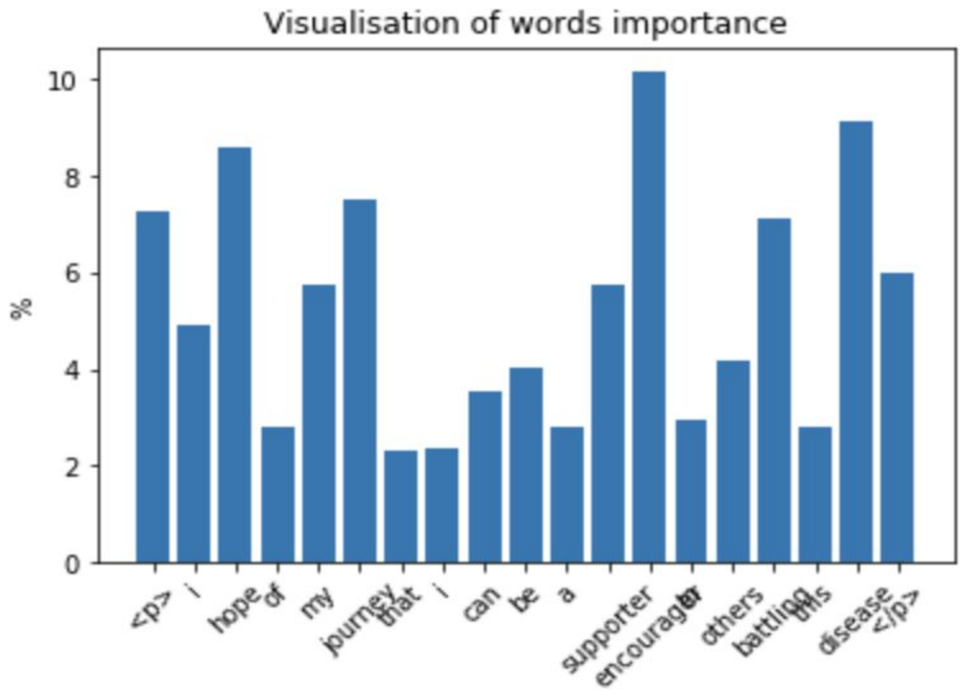
Carnegie Mellon

User 1 : "I hope of my journey I can be a supporter, encourager to others battling this disease. "

User 2 : " Your courage inspired me and I now have hope. "

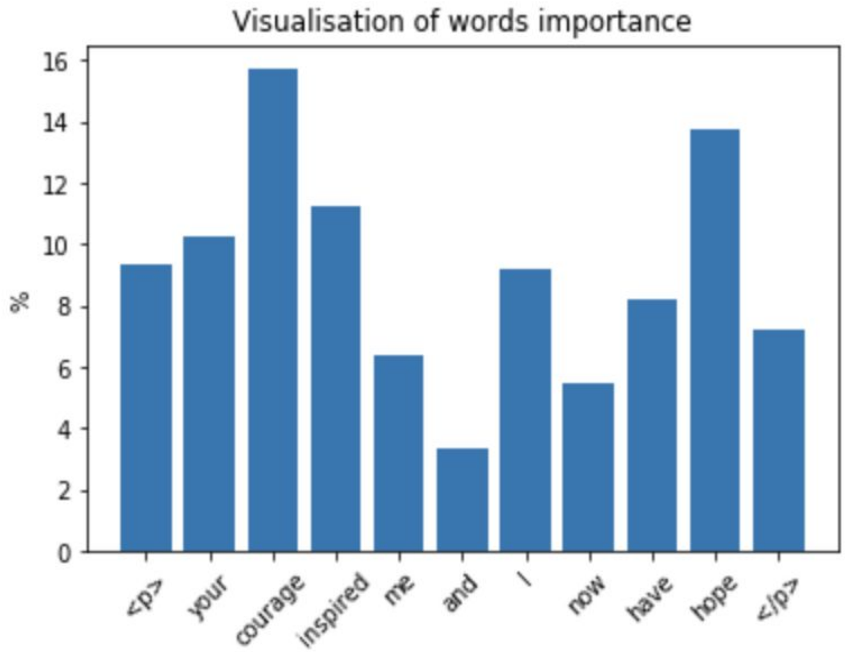
- This can be seen as an example of a user providing support (care-giver) and another user seeking support (care-seeker)

Caregiver - Careseeker Attention Visualization



User 1

User 2



Conclusion & Future Work

Carnegie Mellon

- We proposed a novel task of predicting conversation transitions from public discussion forum to private chats
- Experimented with various approaches to user profile creation
- Contrasted the approaches and analysed the same through attention visualization

- Track the users’ activity to use the feedback for further improvements
- Statistical tests to validate our hypothesis about word level attention model

Key Takeaways

- Application of techniques learnt through courses at CMU
- Handling real world large dataset.
- Dataset creation is not always trivial.
- Approaching a problem by framing it as a known task.
- Symbolic vs Neural learnings.
- Simple models often give strong baselines. Should always verify that first.

Timeline - On Track

Check point	Date	Action Item
0	1st September 2018	IRB Certification
1	15th September 2018	Set up and Replicate Existing System Baseline
2	25th September 2018	Text-based Feature Extraction
3	1st October 2018	Handcrafted feature based approach
4	15th October 2018	Deep Learning based approaches
5	1st November 2018	More Complex Neural Approaches
6	15th November 2018	More textual Features
7	25th November 2018	Matching
8	1st December 2018	Results , Analysis and Conclusion
9	**12th December 2018**	**Report and Final Presentation**

References

1) Alexis Conneau, Douwe Kiela, Holger Schwenk, Lo¨ic Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 670–680, Copenhagen, Denmark. Association for Computational Linguistics

2) Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North
American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1480–1489

3) Lei Zheng, Vahid Noroozi, and Philip S. Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM '17, pages 425–434, New York, NY, USA. ACM.

4) Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, pages 1583–1592. International World Wide Web Conferences Steering Committee.

Q&A?

Thank You