CS626: Research Talk
Computational Sarcasm

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

13th October, 2022

Sentiment Analysis Definition (Liu 2010)

Sarscasm is a part of the general Sentiment analysis problem which is defined by the 5-tuple

< E, F, S, H, T>, where

- *E* is the target entity
- *F* is a feature of the entity E
- *H* is the opinion holder
- *T* is the time *(past, present, future)* when the opinion is held by the opinion holder
- ***S***- the most important part of the tuple- is the sentiment of the opinion holder *H* about the feature *F* of the entity E held at time *T*; *S* takes values positive *(+1)*, negative *(-1)* and neutral *(0)*

- *E.g, E: Movie, F: Acting*

Wheel of emotions

Plutchik (1982)

Multimodality is important

Pax: ***thank you for sending me to Delhi and my luggage to Mumbai ! Brilliant service!!!***

Chatbot: ***Thanks for the appreciation*** 🙇

sarcasm

Empathy reduces communciation barrier: multilingual chatbot built by us

Conversational History	Generic Response	Polite Response
Somebody please help me, I'm having the worst luck with your customer service	How can we help?	Help has arrived! We are sorry to see that you are having trouble, how can we help?
मेरा फोन एक मिनट पहले पूरी तरह से काम कर रहा था और फिर अचानक काम करना बंद कर दिया। (My phone was perfectly working a minute back and then it stopped working suddenly.)	आपको क्या परेशानी हो रही है? (What trouble are you having?)	हम आपके साथ इस पर गौर करेंगे, कृपया हमें बताएं कि आपको क्या परेशानी हो रही है? (We'll look into this with you, please tell us what trouble are you having?)
Is the Radio X radio within the app branded station, or just a station called Radio X?	Can you send us a screenshot of what you're seeing?	No worries at all! Help's here! Please send us a screenshot of what you're seeing? We'd like to take a close look.
पिछले हफ्ते ही घर आया था और मुझे फिर से अपने इंटरनेट से समस्या हो रही है। (Just came home last week and I'm having problems with my internet again.)	आपके इंटरनेट के साथ क्या हो रहा है? (What is happening with your internet?)	अरे नहीं यह अच्छा नहीं है। मैं मदद कर सकता हूँ! कृपया बताएं कि आपके इंटरनेट में क्या समस्या है? (Oh no that's not good. I can help! Please tell whats the problem with your internet?)

Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya; *Incorporating Politeness across Languages in Customer Care Responses: Towards building a Multilingual Empathetic Dialogue Agent*. **LREC 2020**, Marseille, France; 2020.


Homepage
Inbox (105,041) - pushpakbh
Inbox (18,664) - pushpakdir
Resource Centre for Indian L
www.cfilt.iitb.ac.in
Search
Government of India
Center For Indian Language Technology
Indian Institute of Technology, Bombay
Home
Join us
People
Downloads
News
Publications
Tools
Different Guidelines
Resources
Events
ML - Linguistics Lectures
Local
Events
Google Custom Sea
CFILT, IIT Bombay
Center for Indian Language Technology (CFILT) was set up with a generous grant from the Department of Information Technology (DIT), Ministry of Communication and Information Technology, Government of India in 2000 at the Department of Computer Science and Engineering, IIT Bombay. Prior to this the Natural Language Processing (NLP) activity of the CSE Department, IIT Bombay took off in 1996 with a grant from the United Nations University, Tokyo to create a multilingual information exchange system for the web. The project called Universal Networking Language (UNL; www.undl.org) was participated in by 15 research groups across continents.
At any point of time about 30 research members work in CFILT, which includes PhD , masters and bachelor students, faculty members, linguists and lexicographers.
Deep semantics and multilinguality has throughout played a pivotal role in the activities of CFILT. The stress on semantics has led to research in the following fronts:
• Lexical Resources: Multilingual wordnets and ontologies and their linking
• Lexical and Structural Disambiguation: Resolve word and attachment ambiguities
• Shallow Parsing: Identifying correct parts of speech, named entities and non-recursive noun phrases for Marathi and Hindi
• Cross Lingual Information Retrieval: Indian language query to English and Hindi Retrieval
• Machine Translation: Automatic translation involving Marathi, Hindi and English
• Text Entailment: Testing if a piece text (hypothesis) is inferable from another (text)
100%
ENG
IN
3:20 PM
11-Jul-17

Since 2000

A specific problem: Sarcasm Detection

Sarcasm Detection: a sub-problem of Sentiment and Emotion Analysis

Sentiment Analysis: The task of identifying if a certain piece of text contains any opinion, emotion or other forms of affective content.

Sarcasm

- Used as a tool to display wits, whimper or evade

- 6 tuple representation : (S, H, C, u, p, p’) [1]

Speaker

Hearer (Listener)

Context

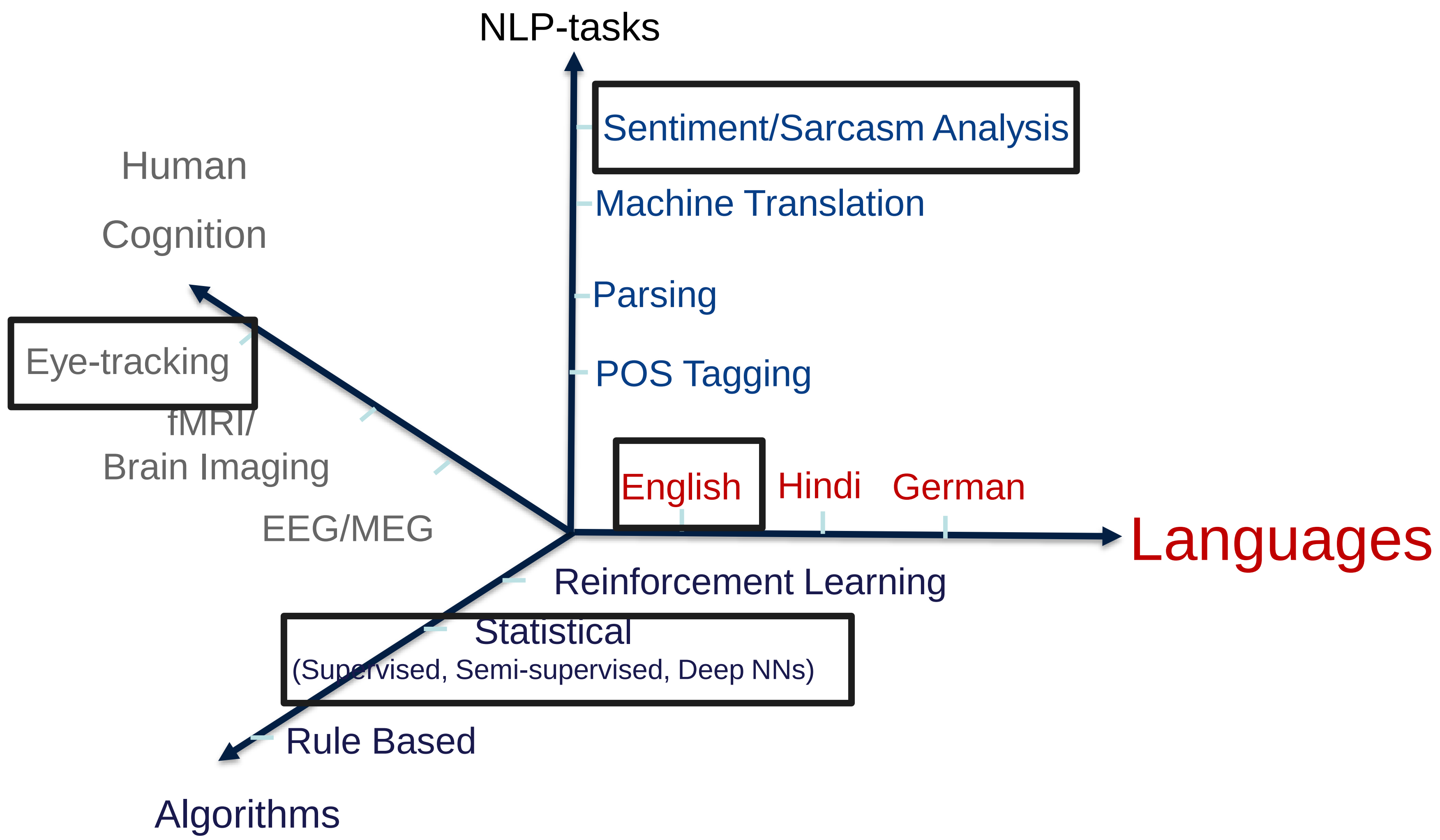
Utterance

Literal proposition

Intended proposition

- Known to contain negative implicit emotion

NLP-trinity (augmented)



Sarcasm: Etymology

- Greek: *'sarkasmós'*: 'to tear flesh with teeth'

- Sanskrit: *'vakrokti'*: 'a twisted (vakra) utterance (ukti)'

Foundation: Irony

Mean opposite of what is on surface

"A form of irony that is intended to express contempt or ridicule." The Free Dictionary	"Verbal irony that expresses negative and critical attitudes toward persons or events." (Kreuz and Glucksberg, 1989)
"The use of irony to mock or convey contempt." Oxford Dictionary	"Irony that is especially bitter and caustic" (Gibbs, 1994)

Allied concept: **Humble Bragging**- "Oh my life is miserable, have to sign 500 autographs a day!!

Types of Sarcasm

Sarcasm (Camp, 2012)			
Propositional	**Embedded**	**Like-prefixed**	**Illocutionary**
A proposition that is intended to be sarcastic. *'This looks like a perfect plan!'*	Sarcasm is embedded in the meaning of words being used. *'I love being ignored'*	'Like/As if' are common prefixes to ask rhetorical questions. *'Like you care'*	Non-speech acts (body language, gestures) contributing to the sarcasm *'(shrugs shoulders) Very helpful indeed!'*

Illocutionary sarcasm

Impact of Sarcasm on Sentiment Analysis (SA) (1/2)

Two SA systems:

- *MeaningCloud: https://www.meaningcloud.com/*
- *NLTK (Bird, 2006)*

Two datasets:

- *Sarcastic tweets by Riloff et al (2013)*
- *Sarcastic utterances from our dataset of TV transcripts (Joshi et al 2016b)*

Impact of Sarcasm on Sentiment Analysis (2/2)

	Precision (Sarc)	Precision (Non-sarc)
Conversation Transcripts		
MeaningCloud[1]	20.14	49.41
NLTK (Bird, 2006)	38.86	81
Tweets		
MeaningCloud[1]	17.58	50.13
NLTK (Bird, 2006)	35.17	69

[1] www.meaningcloud.com

Clues for Sarcasm

- Use of laughter expression
  - *haha, you are very smart xD*
  - *Your intelligence astounds me. LOL*
- Heavy Punctuation
  - *Protein shake for dinner!! Great!!!*
- Use of emoticons
  - *i LOVE it when people tweet yet ignore my text X-(*
- Interjections
  - *3:00 am work YAY. YAY.*
- Capital Letters
  - *SUPER EXCITED TO WEAR MY UNIFORM TO SCHOOL TOMORROW ! ! :D lol.*

Incongruity: at the heart of things!

- *I love being ignored*
- *3:00 am work YAY. YAY.*
- *Up all night coughing. yeah me!*
- *No power, Yes! Yes! Thank you storm!*
- *This phone has an awesome battery back-up of 2 hour (Sarcastic)*

Two kinds of incongruity

- **Explicit incongruity**
  - Overtly expressed through sentiment words of both polarities
  - Contribute to almost 11% of sarcasm instances

    'I <u>love</u> being <u>ignored</u>'
- **Implicit incongruity**
  - Covertly expressed through phrases of implied sentiment

    'I <u>love</u> this paper so much that I <u>made a doggy bag out of</u> it'

Sarcasm and Sense Ambiguity

Oh! Its so nice of you to give me a ring early in the morning!

Good to see you help dog bite victim!

Sarcasm Detection Using Semantic Incongruity

Aditya Joshi, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya and Mark Carman, *Are Word Embedding-based Features Useful for Sarcasm Detection?*, **EMNLP 2016**, Austin, Texas, USA, November 1-5, 2016.

Also covered in: How Vector Space Mathematics Helps Machines Spot Sarcasm, MIT Technology Review, 13th October, 2016.

www.cfilt.iitb.ac.in/sarcasmsuite/

Feature Set

Lexical	
Unigrams	Unigrams in the training corpus
Pragmatic	
Capitalization	Numeric feature indicating presence of capital letters
Emoticons & laughter expressions	Numeric feature indicating presence of emoticons and ‘lol’s
Punctuation marks	Numeric feature indicating presence of punctuation marks
Implicit Incongruity	
Implicit Sentiment Phrases	Boolean feature indicating phrases extracted from the implicit phrase extraction step
Explicit Incongruity	
#Explicit incongruity	Number of times a word is followed by a word of opposite polarity
Largest positive /negative subsequence	Length of largest series of words with polarity unchanged
#Positive words	Number of positive words
#Negative words	Number of negative words
Lexical Polarity	Polarity of a tweet based on words present

Datasets

Name	Text-form	Method of labeling	Statistics
Tweet-A	Tweets	Using sarcasm-based hashtags as labels	5208 total, 4170 sarcastic
Tweet-B	Tweets	Manually labeled (Given by Riloff et al(2013))	2278 total, 506 sarcastic
Discussion-A	Discussion forum posts (IAC Corpus)	Manually labeled (Given by Walker et al (2012))	1502 total, 752 sarcastic

Results

Features	P	R	F
Original Algorithm by Riloff et al. (2013)			
Ordered	0.774	0.098	0.173
Unordered	0.799	0.337	0.474
Our system			
Lexical (**Baseline**)	0.820	0.867	0.842
Lexical+Implicit	0.822	0.887	0.853
Lexical+Explicit	0.807	0.985	0.8871
All features	0.814	0.976	**0.8876**

Tweet-A

Approach	P	R	F
Riloff et al. (2013) (**best reported**)	0.62	0.44	0.51
Maynard and Greenwood (2014)	0.46	0.38	0.41
Our system (all features)	**0.77**	**0.51**	**0.61**

Tweet-B

Features	P	R	F
Lexical (**Baseline**)	0.645	0.508	0.568
Lexical+Explicit	0.698	0.391	0.488
Lexical+Implicit	0.513	0.762	0.581
All features	0.489	0.924	**0.640**

Discussion-A

Capturing Incongruity Using Word Vectors

Use **similarity** of word embeddings

”A man needs a woman like a fish needs bicycle”

Word2Vec *similarity(man,woman)*= 0.766
Word2Vec *similarity(fish, bicycle)*= 0.131

Word embedding-based features

Unweighted similarity features (S):

Maximum score of most similar word pair

Minimum score of most similar word pair

Maximum score of most dissimilar word pair

Minimum score of most dissimilar word pair

Distance-weighted similarity features (WS):

4 S features weighted by linear distance between the two words

Both (S+WS): 8 features

Experiment Setup

- Dataset: 3629 Book snippets (759 sarcastic) downloaded from GoodReads website
- Labelled by users with tags
- Five-fold cross-validation
- Classifier: SVM-Perf optimised for F-score
- Configurations:
  - Four prior works (augmented with our sets of features)
  - Four implementations of word embeddings (Word2Vec, LSA, GloVe, Dependency weights-based)

Results (1/2)

Features	P	R	F
	Baseline		
Unigrams	67.2	78.8	72.53
S	64.6	75.2	69.49
WS	67.6	51.2	58.26
Both	67	52.8	59.05

	LSA			GloVe			Dependency Weights			Word2Vec		
	P	R	F	P	R	F	P	R	F	P	R	F
L	73	79	75.8	73	79	75.8	73	79	75.8	73	79	75.8
+S	81.8	78.2	**79.95**	81.8	79.2	**80.47**	81.8	78.8	80.27	80.4	80	**80.2**
+WS	76.2	79.8	77.9	76.2	79.6	77.86	81.4	80.8	81.09	80.8	78.6	79.68
+S+WS	77.6	79.8	78.68	74	79.4	76.60	82	80.4	**81.19**	81.6	78.2	79.86
G	84.8	73.8	78.91	84.8	73.8	78.91	84.8	73.8	**78.91**	84.8	73.8	**78.91**
+S	84.2	74.4	**79**	84	72.6	77.8	84.4	72	77.7	84	72.8	78
+WS	84.4	73.6	78.63	84	75.2	**79.35**	84.4	72.6	78.05	83.8	70.2	76.4
+S+WS	84.2	73.6	78.54	84	74	78.68	84.2	72.2	77.73	84	72.8	78
B	81.6	72.2	76.61	81.6	72.2	76.61	81.6	72.2	76.61	81.6	72.2	76.61
+S	78.2	75.6	**76.87**	80.4	76.2	**78.24**	81.2	74.6	**77.76**	81.4	72.6	76.74
+WS	75.8	77.2	76.49	76.6	77	76.79	76.2	76.4	76.29	81.6	73.4	77.28
+S+WS	74.8	77.4	76.07	76.2	78.2	77.18	75.6	78.8	77.16	81	75.4	**78.09**
J	85.2	74.4	79.43	85.2	74.4	79.43	85.2	74.4	79.43	85.2	74.4	79.43
+S	84.8	73.8	78.91	85.6	74.8	79.83	85.4	74.4	79.52	85.4	74.6	**79.63**
+WS	85.6	75.2	**80.06**	85.4	72.6	78.48	85.4	73.4	78.94	85.6	73.4	79.03
+S+WS	84.8	73.6	78.8	85.8	75.4	**80.26**	85.6	74.4	**79.6**	85.2	73.2	78.74

Table 3: Performance obtained on augmenting word embedding features to features from four prior works, for four word embeddings; L: Liebrecht et al. (2013), G: González-Ibánez et al. (2011a), B: Buschmeier et al. (2014) , J: Joshi et al. (2015)

Results (2/2)

	Word2Vec	**LSA**	**GloVe**	**Dep. Wt.**
+S	0.835	0.86	0.918	**0.978**
+WS	**1.411**	0.255	0.192	1.372
+S+WS	**1.182**	0.24	0.845	0.795

Table 4: Average gain in F-Scores obtained by using intersection of the four word embeddings, for three word embedding feature-types, augmented to four prior works; Dep. Wt. indicates vectors learned from dependency-based weights

Word Embedding	**Average F-score Gain**
LSA	0.452
Glove	0.651
Dependency	1.048
Word2Vec	1.143

Table 5: Average gain in F-scores for the four types of word embeddings; These values are computed for a subset of these embeddings consisting of words common to all four

Numerical Sarcasm

Illustrates *need* for

Rule Based → Classical ML → Deep Learning

Abhijeet Dubey, Lakshya Kumar, Arpan Somani, Aditya Joshi and Pushpak Bhattacharyya, *"When Numbers Matter!": Detecting Sarcasm in Numerical Portions of Text*, 10th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (**WASSA 2019**), Minneapolis, USA, 7 June, 2019.

About 17% of sarcastic tweets have origin in number

1- This phone has an awesome battery back-up of 38 hours (Non-sarcastic)

2- This phone has a terrible battery back-up of 2 hours (Non-sarcastic)

3- This phone has an awesome battery back-up of 2 hour (Sarcastic)

Interesting question: why people use sarcasm?

- Dramatization, Forceful Articulation, lowering defence and then attack!

Numerical Sarcasm Dataset

Dataset-1	100000 (Sarcastic)	250000 (Non-Sarcastic)
Dataset-2	8681 (Num Sarcastic)	8681 (Non-Sarcastic)
Dataset-3	8681 (Num Sarcastic)	42107 (Non-Sarcastic)
Test Data	1843 (Num Sarcastic)	8317 (Non-Sarcastic)

- To create this dataset, we extract tweets from Twitter-API (https://dev.twitter.com).
- Hashtags of the tweets served as labels *#sarcasm #sarcastic* etc.
- Dataset-1 contains normal sarcastic + numeric sarcastic and non-sarcastic tweets.
- Rest all the other dataset contains numeric sarcastic and non-sarcastic tweets only.

Rule-based System (NP-Exact Matching) (Cont'd)

- Test Tweet: 'I love writing this paper at 9 am
- Matched Sarcastic Tweet: 'I love writing this paper daily at 3 am'
- 9 ***NOT*** close to 3

*test tweet is **non-sarcastic***

Example (sarcastic case)

- Test Tweet: ‘I am so productive when my room is 81 degrees‘
- Matched Non-sarcastic Tweet: ‘I am very much productive in my room as it has 21 degrees‘
- Absolute difference between 81 and 21 is high

Hence test tweet is

Sarcastic

Comparison of results (1: sarcastic, 0: non-sarcastic)

Approaches	Precision			Recall			F-score		
	P(1)	P(0)	P(avg)	R(1)	R(0)	R(avg)	F(1)	F(0)	F(avg)
Past Approaches									
Buschmeier et.al.	0.19	0.98	0.84	0.99	0.07	0.24	0.32	0.13	0.16
Liebrecht et.al.	0.19	1.00	0.85	1.00	0.07	0.24	0.32	0.13	0.17
Gonzalez et.al.	0.19	0.96	0.83	0.99	0.06	0.23	0.32	0.12	0.15
Joshi et.al.	0.20	1.00	0.86	1.00	0.13	0.29	0.33	0.23	**0.25**
Rule-Based Approaches									
Approach-1	0.53	0.87	0.81	0.39	0.92	0.83	0.45	0.90	**0.82**
Approach-2	0.44	0.85	0.78	0.28	0.92	0.81	0.34	0.89	0.79

Machine Learning based approach: classifiers and features

- SVM, KNN and Random Forest classifiers
- Sentiment-based features
  - Number of
    - positive words
    - negative words
    - highly emotional positive words,
    - highly emotional negative words.
- Positive/Negative word is said to be highly emotional if it’s POS tag is one amongst : ’JJ', ‘JJR', ‘JJS', ‘RB', ‘RBR', ‘RBS', ‘VB', ‘VBD',

Emotion Features

- Positive emoticon
- Negative emoticon
- Boolean feature that will be one if both positive and negative words are present in the tweet.
- Boolean feature that will be one when either positive word and negative emoji is present or vice versa.

Punctuation features

- number of exclamation marks.
- number of dots
- number of question mark.
- number of capital letter words.
- number of single quotations.

- Number in the tweet: This feature is simply the number present in the tweet.
- Number unit in the tweet : This feature is a one hot representation of the type of unit present in the tweet. Example of number unit can be hour, minute, etc.

Comparison of results (1: sarcastic, 0: non-sarcastic)

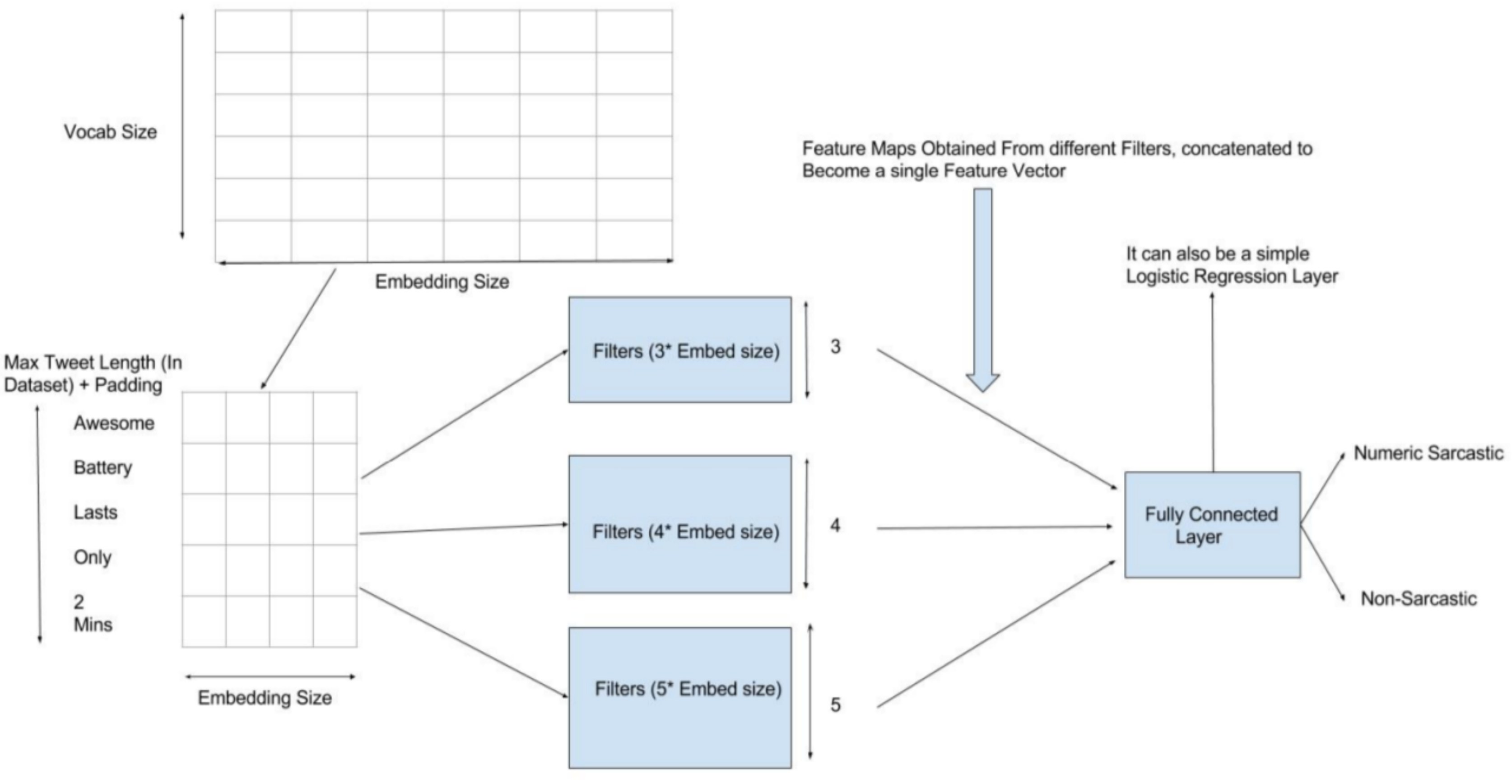
Approaches	Precision			Recall			F-score		
	P(1)	**P(0)**	**P(avg)**	**R(1)**	**R(0)**	**R(avg)**	**F(1)**	**F(0)**	**F(avg)**
Past Approaches									
Buschmeier et.al.	0.19	0.98	0.84	0.99	0.07	0.24	0.32	0.13	0.16
Liebrecht et.al.	0.19	1.00	0.85	1.00	0.07	0.24	0.32	0.13	0.17
Gonzalez et.al.	0.19	0.96	0.83	0.99	0.06	0.23	0.32	0.12	0.15
Joshi et.al.	0.20	1.00	0.86	1.00	0.13	0.29	0.33	0.23	**0.25**
Rule-Based Approaches									
Approach-1	0.53	0.87	0.81	0.39	0.92	0.83	0.45	0.90	**0.82**
Approach-2	0.44	0.85	0.78	0.28	0.92	0.81	0.34	0.89	0.79
Machine-Learning Based Approaches									
SVM	0.50	0.95	0.87	0.80	0.82	0.82	0.61	0.88	**0.83**
KNN	0.36	0.94	0.84	0.81	0.68	0.70	0.50	0.79	0.74
Random Forest	0.47	0.93	0.85	0.74	0.81	0.80	0.57	0.87	0.82

Deep Learning based

- Very little feature engg!!

- EmbeddingSize of 128
- Maximum tweet length 36 words
- Padding used
- Filters of size 3, 4, 5 used to extarct features

Deep Learning based approach: CNN-FF Model



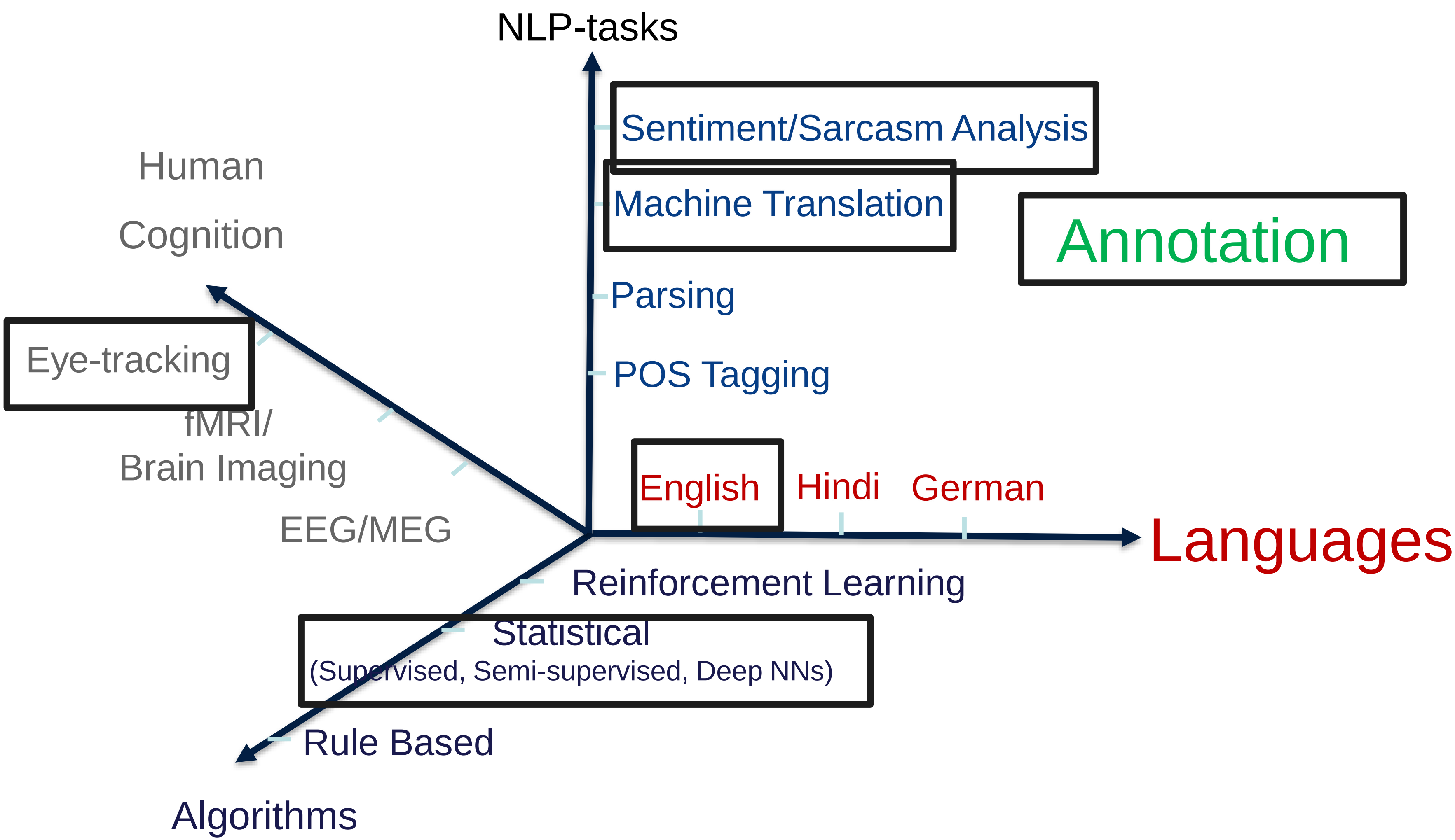
Comparison of results (1: sarcastic, 0: non-sarcastic)

Approaches	Precision			Recall			F-score		
	P(1)	**P(0)**	**P(avg)**	**R(1)**	**R(0)**	**R(avg)**	**F(1)**	**F(0)**	**F(avg)**
Past Approaches									
Buschmeier et.al.	0.19	0.98	0.84	0.99	0.07	0.24	0.32	0.13	0.16
Liebrecht et.al.	0.19	1.00	0.85	1.00	0.07	0.24	0.32	0.13	0.17
Gonzalez et.al.	0.19	0.96	0.83	0.99	0.06	0.23	0.32	0.12	0.15
Joshi et.al.	0.20	1.00	0.86	1.00	0.13	0.29	0.33	0.23	**0.25**
Rule-Based Approaches									
Approach-1	0.53	0.87	0.81	0.39	0.92	0.83	0.45	0.90	**0.82**
Approach-2	0.44	0.85	0.78	0.28	0.92	0.81	0.34	0.89	0.79
Machine-Learning Based Approaches									
SVM	0.50	0.95	0.87	0.80	0.82	0.82	0.61	0.88	**0.83**
KNN	0.36	0.94	0.84	0.81	0.68	0.70	0.50	0.79	0.74
Random Forest	0.47	0.93	0.85	0.74	0.81	0.80	0.57	0.87	0.82
Deep-Learning Based Approaches									
CNN-FF	**0.88**	**0.94**	**0.93**	**0.71**	**0.98**	**0.93**	**0.79**	**0.96**	**0.93**
CNN-LSTM-FF	0.82	0.94	0.92	0.72	0.96	0.92	0.77	0.95	0.92
LSTM-FF	0.76	0.93	0.90	0.68	0.95	0.90	0.72	0.94	0.90

back

Enter cognition

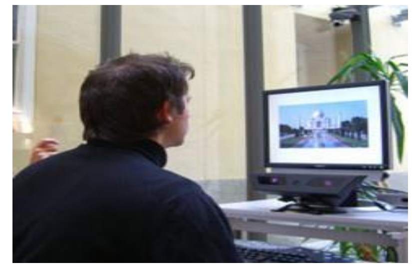
NLP-trinity



NLP-tasks
Sentiment/Sarcasm Analysis
Machine Translation
Annotation
Parsing
POS Tagging
Human
Cognition
Eye-tracking
fMRI/
Brain Imaging
EEG/MEG
English
Hindi
German
Languages
Reinforcement Learning
Statistical
(Supervised, Semi-supervised, Deep NNs)
Rule Based
Algorithms

Eye-tracking Technology

Invasive and non-invasive eye-trackers



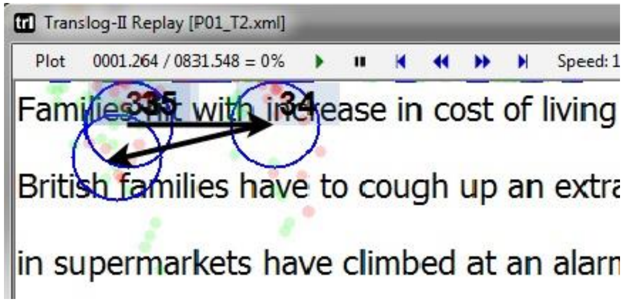
(image - sources: http://www.tobii.com/)

For linguistic studies non-invasive eye-trackers are used

- **Data delivered by eye-trackers**
  - Gaze co-ordinates of both eyes (binocular setting) or single eye (monocular setting)
  - Pupil size
- **Derivable data**
  - Fixations, Saccades, Scanpaths, Specific patterns like progression and regression.

Nature of Gaze Data

- **Gaze Point:** Position (co-ordinate) of gaze on the screen
- **Fixations :** A long stay of the gaze on a particular object on the screen
- **Saccade:** A very rapid movement of eye between the positions of rest.
  - Progressive Saccade / Forward Saccade / Progression
  - Regressive Saccade / Backward Saccade / Regression
- **Scanpath:** A path connecting a series of fixations.



Eye-movement and Cognition

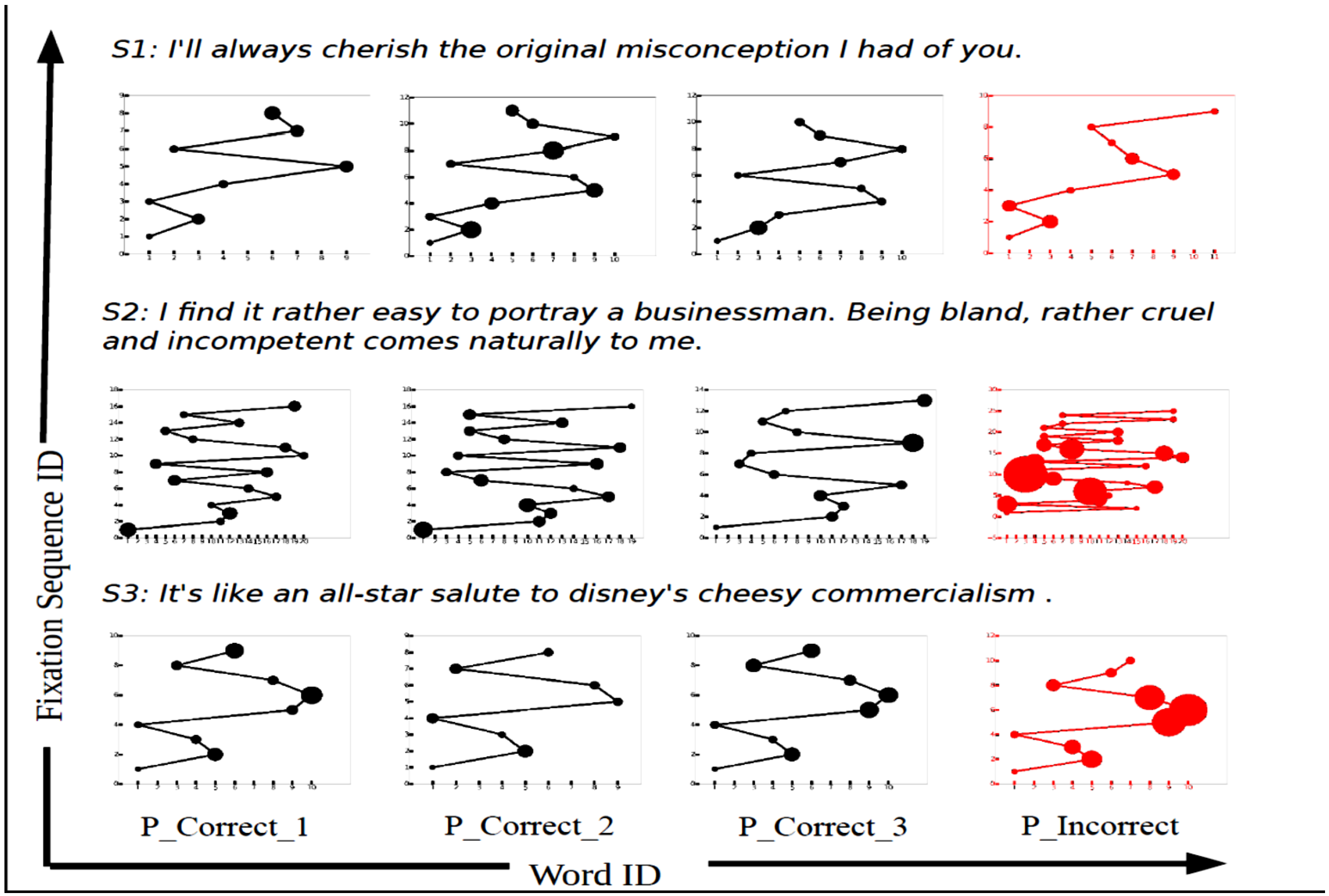
- Eye-Mind Hypothesis (Just and Carpenter, 1980)

When a subject is views a word/object, he or she also processes it cognitively, for approximately the same amount of time he or she fixates on it.

- Considered useful in explaining theories associated with reading (Rayner and Duffy,1986; Irwin, 2004; von der Malsburg and Vasishth, 2011)

- Linear and uniform-speed gaze movement is observed over texts having simple concepts, and often non-linear movement with non-uniform speed over more complex concepts (Rayner, 1998)

Sarcasm Understandability – Scanpath Representation



Harnessing Cognitive Features for Sarcasm Detection (Mishra and Bhattacharyya, ACL 2016)

Features for Sarcasm: Augmented with cognitive

Textual

(1) Unigrams (2) Punctuations
(3) Implicit incongruity
(4) Explicit Incongruity
(5) Largest +ve/-ve subsequences
(6) +ve/-ve word count
(7) Lexical Polarity
(8) Flesch Readability Ease,
(9) Word count

Complex gaze

(1) Edge density,
(2) Highest weighted degree
(3) Second Highest weighted degree
(With different edge-weights)

Simple gaze

(1) Average Fixation Duration,
(2) Average Fixation Count,
(3) Average Saccade Length,
(4) Regression Count,
(5) Number of words skipped,
(6) Regressions from second half to first half,
(7) Position of the word from which the largest regression starts

Experiment Setup

- **Dataset**:
  - 994 text snippets : 383 positive and 611 negative, 350 are sarcastic/ironic
  - Mixture of Movie reviews, Tweets and sarcastic/ironic quotes
  - Annotated by 7 human annotators
  - Annotation accuracy: **70%-90%** with Fleiss kappa IAA of **0.62**

- **Classifiers**:
  - Naïve Bayes, SVM, Multi Layered Perceptron
  - Feature combinations:
    - Unigram Only
    - Gaze Only (Simple + Complex)
    - Textual Sarcasm Features (Joshi et., al, 2015) (Includes unigrams)
    - Gaze+ Sarcasm

- **Compared with** : **Riloff, 2013 and Joshi, 2015**

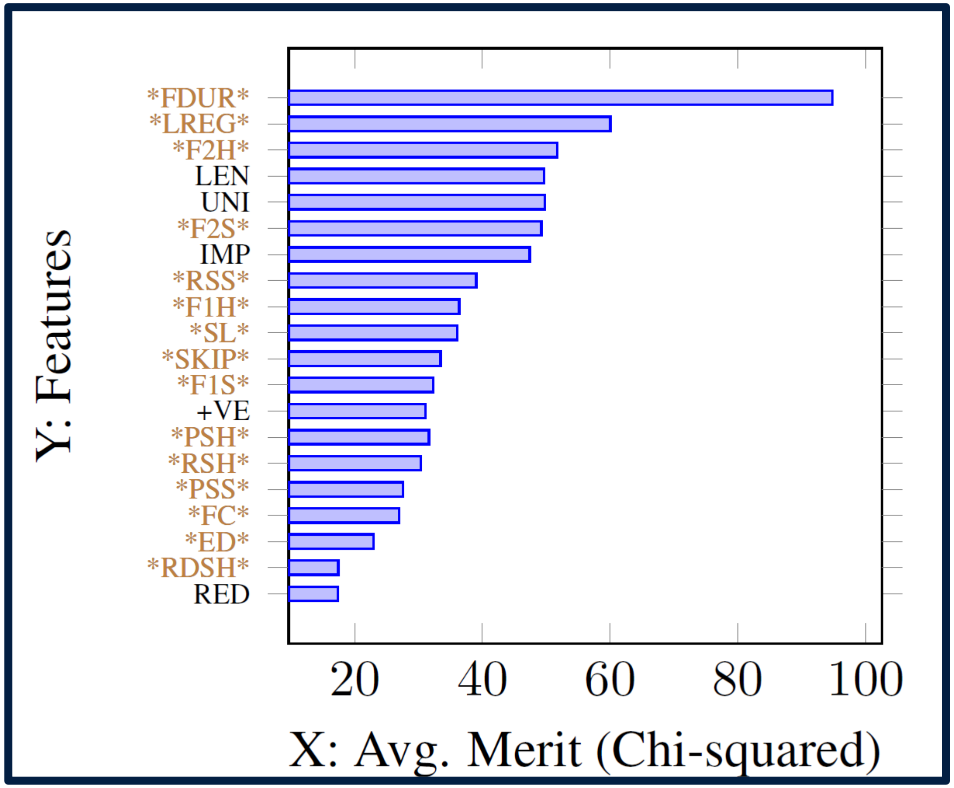
Results

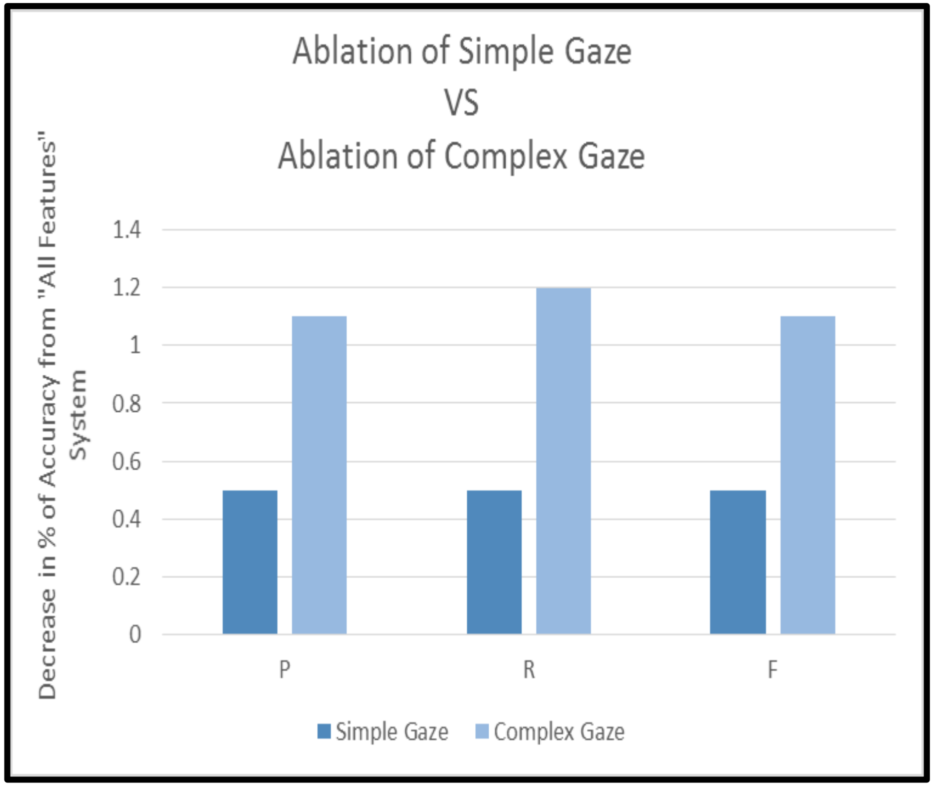
Features	P(1)	P(-1)	P(avg)	R(1)	R(-1)	R(avg)	F(1)	F(-1)	F(avg)
Multi Layered Neural Network									
Unigram	53.1	74.1	66.9	51.7	75.2	66.6	52.4	74.6	66.8
Sarcasm (Joshi et. al.)	59.2	75.4	69.7	51.7	80.6	70.4	55.2	77.9	69.9
Gaze	62.4	76.7	71.7	54	82.3	72.3	57.9	79.4	71.8
Gaze+Sarcasm	63.4	75	70.9	48	84.9	71.9	54.6	79.7	70.9
Näive Bayes									
Unigram	45.6	82.4	69.4	81.4	47.2	59.3	58.5	60	59.5
Sarcasm (Joshi et. al.)	46.1	81.6	69.1	79.4	49.5	60.1	58.3	61.6	60.5
Gaze	57.3	82.7	73.8	72.9	70.5	71.3	64.2	76.1	71.9
Gaze+Sarcasm	46.7	82.1	69.6	79.7	50.5	60.8	58.9	62.5	61.2
Original system by Riloff et.al. : Rule Based with implicit incongruity									
Ordered	60	30	49	50	39	46	54	34	47
Unordered	56	28	46	40	42	41	46	33	42
Original system by Joshi et.al. : SVM with RBF Kernel									
Sarcasm (Joshi et. al.)	73.1	69.4	70.7	22.6	95.5	69.8	34.5	80.4	64.2
SVM Linear: with default parameters									
Unigram	56.5	77	69.8	58.6	75.5	69.5	57.5	76.2	69.6
Sarcasm (Joshi et. al.)	*59.9*	*78.7*	*72.1*	*61.4*	*77.6*	*71.9*	*60.6*	*78.2*	*72*
Gaze	**65.9**	75.9	72.4	49.7	86	73.2	56.7	80.6	72.[illegible]
Gaze+Sarcasm	63.7	79.5	74	61.7	80.9	74.1	62.7	80.2	74
Multi Instance Logistic Regression: Best Performing Classifier									
Gaze	65.3	77.2	73	53	**84.9**	73.8	58.5	**80.8**	73.[illegible]
Gaze+Sarcasm	62.5	**84**	**76.5**	**72.6**	76.7	**75.3**	**67.2**	80.2	**75.[illegible]**

p=0.01

p=0.03

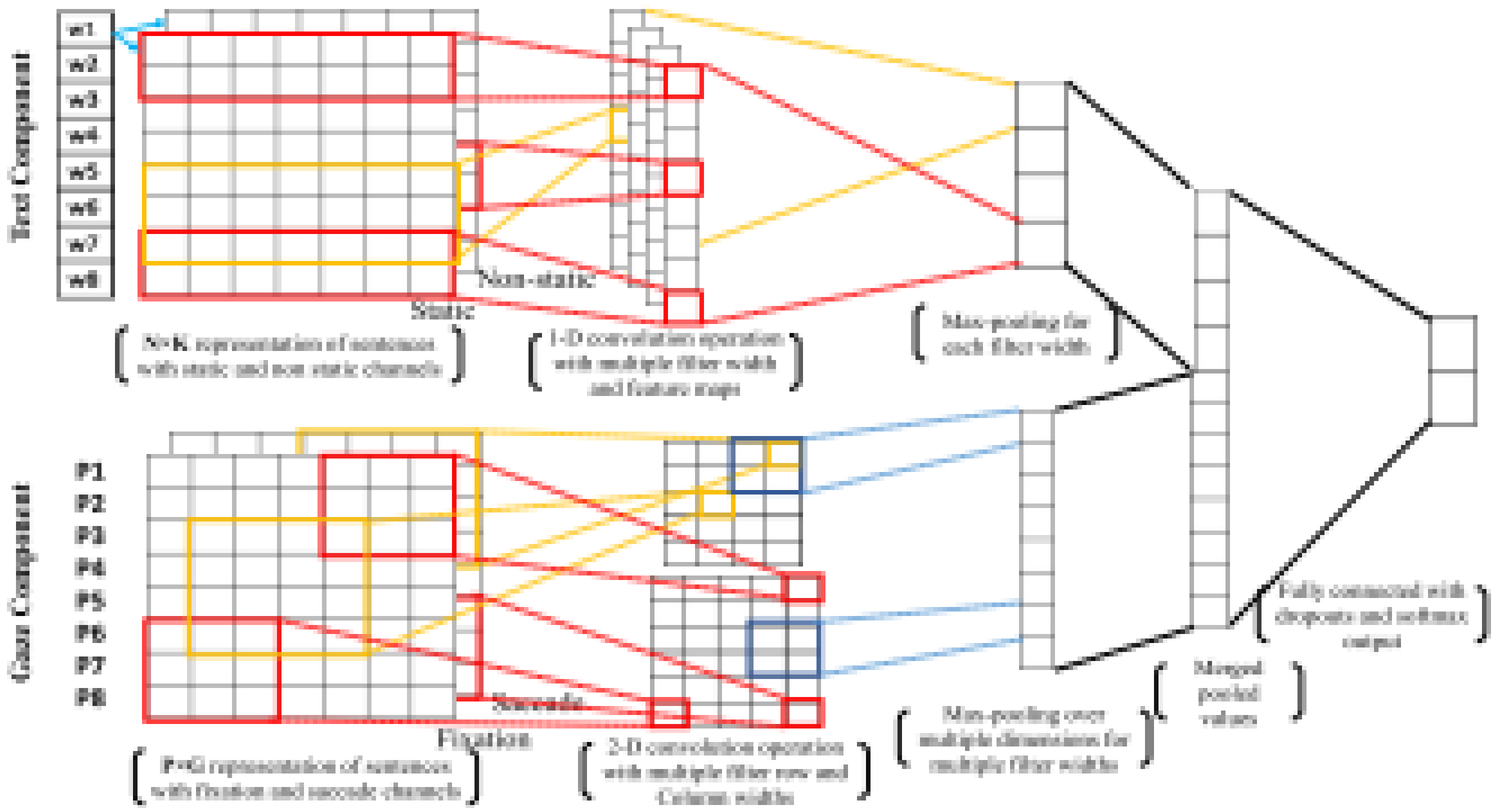
Feature Significance





Abhijit Mishra, Kuntal Dey and Pushpak Bhattacharyya, *Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification Using Convolutional Neural Network*, **ACL 2017**, Vancouver, Canada, July 30-August 4, 2017.

CNN-FF combination



Results: Sarcasm Detection

	Configuration	Precision	Recall	F_Score
Gaze	Gaze-Fixation	74.39	69.62	**71.93**
	Gaze-Saccade	68.58	68.23	68.40
	Gaze-Multi-channel	67.93	67.72	67.82
Text	Text-static	67.17	66.38	66.77
	Text-non-static	84.19	87.03	**85.59**
	Text-Multi-channel	84.28	87.03	**85.63**
Gaze & Text	Text-static_Gaze-Fixation	72.38	71.93	72.15
	Text-static_Gaze-Saccade	73.12	72.14	72.63
	Text-static_Gaze-Multi-channel	71.41	71.03	71.22
	Text-non-static_Gaze-Fixation	87.42	85.2	**86.30**
	Text-non-static_Gaze-Saccade	84.84	82.68	83.75
	Text-non-static_Gaze-Multi-channel	84.98	82.79	83.87
	Text-Multi-channel_Gaze-Fixation	87.03	86.92	**86.97**
	Text-Multi-channel_Gaze-Saccade	81.98	81.08	81.53
	Text-Multi-channel_Gaze-Multi-channel	83.11	81.69	82.39

(a) Results with Deep CNNs

Configuration	Precision	Recall	F_Score
Gaze_NB	73.8	71.3	71.9
Gaze_SVM	**72.4**	**73.2**	**72.2**
Gaze_MLP	71.7	72.3	71.8

(b) CoNLL systems with Gaze Features

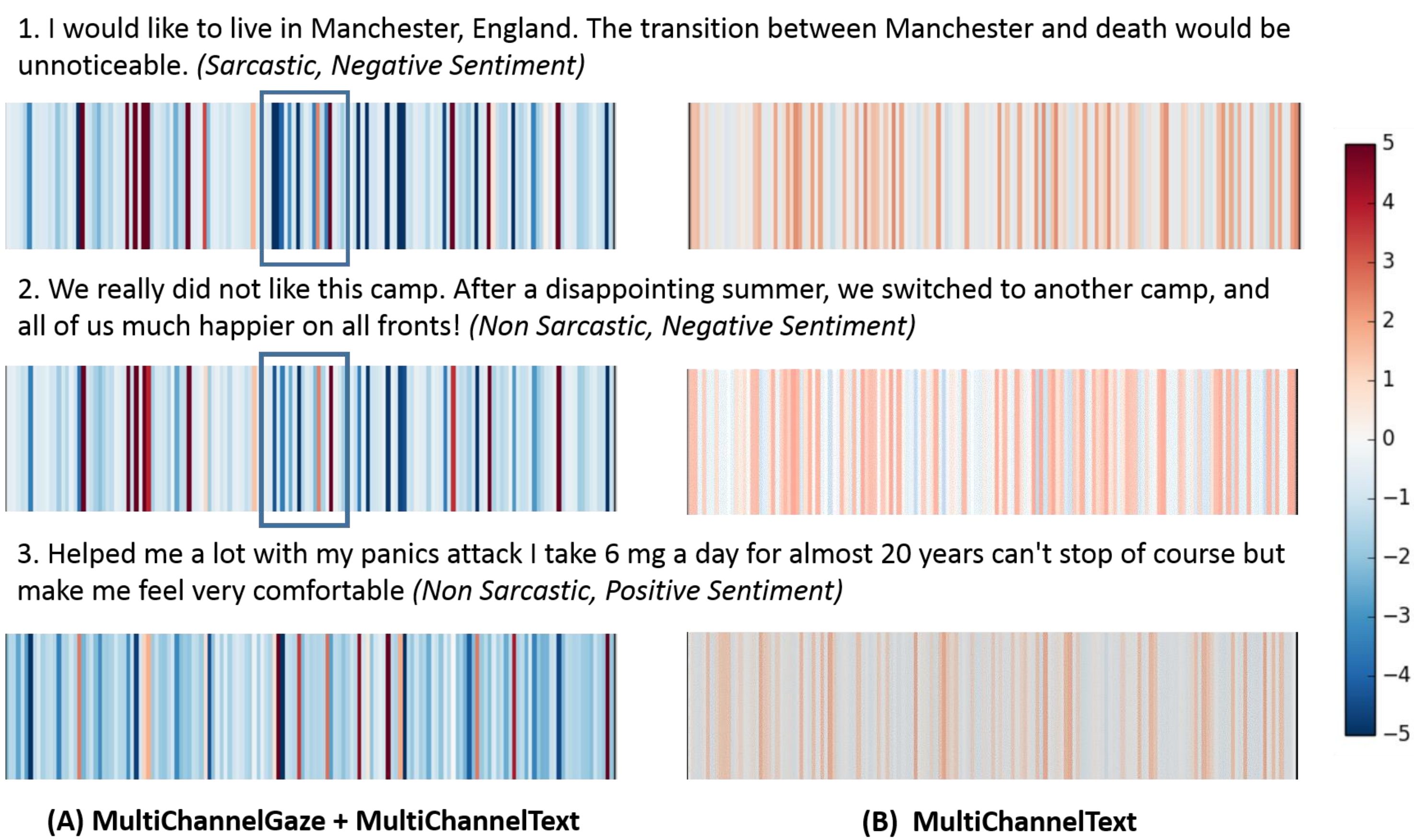
Configuration	Precision	Recall	F_Score
Gaze_Text_NB	**70.9**	71.9	71.2
Gaze_Text_SVM	**74**	**74.1**	**74**
Gaze_Text_MLP	70.9	71.9	70.9

(c) CoNLL systems with Gaze+Text Features

Observations - Sarcasm

- **Higher classification accuracy**
  - Clear differences between vocabulary of sarcasm and no-sarcasm classes in our dataset., Captured well by non-static embeddings.
- **Effect of dimension variation**
  - Reducing embedding dimension improves accuracy by a little margin.
- **Effect of fixation / saccade channels:**
  - Fixation and saccade channels perform with similar accuracy when employed separately.
  - Accuracy reduces with gaze multichannel (may be because the higher variation of both fixations and saccades across sarcastic and non-sarcastic classes, unlike sentiment classes).

Analysis of Features



- Visualization of representations learned by two variants of the network. The output of the Merge layer (of dimension 150) are plotted in the form of colour-bars following Li et al. (2016)

Ongoing work

Work with masters students Apoorva, Divyank, and IBM Researchers Anupama and Rudramurthy

Sarcasm Detection Using Gaze Features

- Detect the presence of sarcasm in a Multimodal Input having Audio, Video, Text and Gaze features.

- Input: Fused Input vector of Audio, Video and Text features along with gaze features of an utterance.

- Output: Tag representing presence / absence of Sarcasm.

DATASET

- **Mustard++ dataset** - 1202 Scenes (Dialogue Conversations) taken from popular sitcoms like Bigbang theory, Friends.
  - 601 Sarcastic and 601 Non-Sarcastic samples.
- Each scene has a set of dialogues as context followed by a single utterance sentence.
- For each context and utterance, Video, Text and Audio features were extracted.
- 231 samples from MUSTARD++ dataset (116 sarcastic + 115 nonsarcastic) were manually selected which had audio visual and text modalities and these were added with gaze fetures.

DATA INSTANCE EXAMPLE

- BERNADETTE: So, what are you working
- on these days?
- AMY: I'm studying one-celled organisms
- to try and find the neurochemicals that
- lead to the feeling of shame.
- BERNADETTE: What would a
- one-celled organism have to be
- embarrassed about?

Context

- PENNY: Same as all of us, getting out
- of a car without clothes.

Utterance

Annotation

- 5 annotators annotated 230 samples
- Fleiss Kappa 0.4 (good agreement)
- During Annotation, the annotators eye was tracked and various gaze features like Fixation duration, Regression Path duration were captured.
- blue circles are fixation points

Gaze Features Involved

IA_FIRST_FIXATION_DURATION: Duration of the first fixation event that was within the current interest area.

IA_REGRESSION_PATH_DURATION: The summed fixation duration from when the current interest area is first fixated until the eyes enter an interest area with a higher IA_ID.

IA_REGRESSION_OUT_FULL_COUNT: Number of times interest area was exited to a lower IA_ID (to the left in English).

IA_REGRESSION_IN: Whether the current interest area received at least one regression from later interest areas (e.g., laterparts of the sentence). 1 if interest area was entered from a higher IA_ID (from the right in
English); 0 if not.

IA_RUN_COUNT: Number of times the Interest Area was entered and left (runs)

....

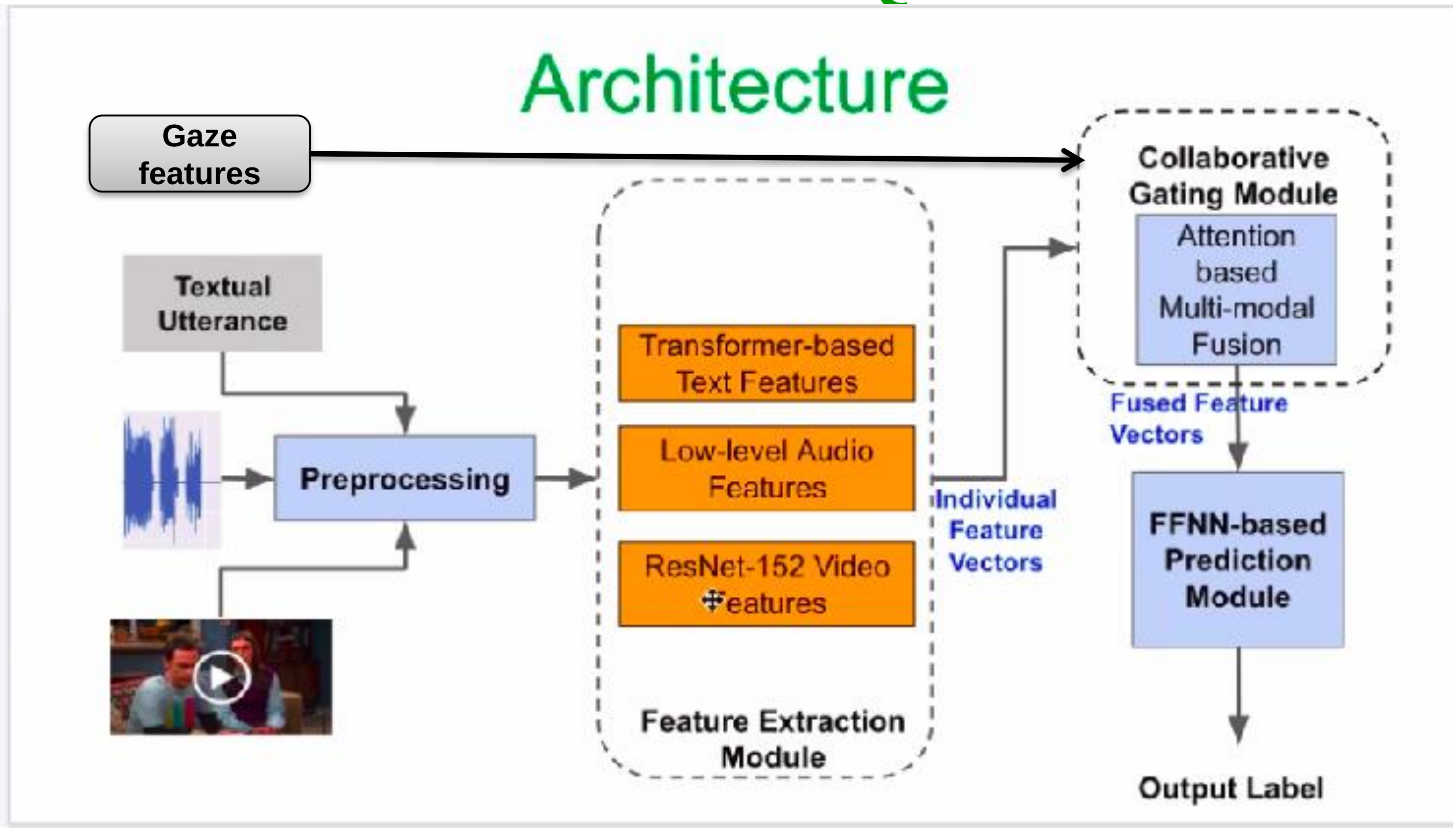
Multimodal Feature Extractions

Text Feature extraction: 1024 sized text feature vectors generated using BART model.

Video Feature extractions: 2048 sized video feature vectors were generated using ResNet-152 model.

Audio Feature extractions: MFCC features along with prosodic features extracted using OpenSMILE toolkit .

TECHNIQUE



Along with the indivisual multimodal feature vetors, the feature vector of the gaze features is also passed in the fusion module.

The FFNN now also uses the gaze feature to predict output label.

Baseline Model

Technique:

-Gaze is not used in this setting

- The 230 samples were split into (171, 59) samples.

- The 171 samples were replicated 5 times since eye tracking data of 5 annotators is being used for same set of sentences.
- The remaining 59 new instances were used for testing.

- Feed forward neural network is being used which takes in the fused inputs from text, audio, video.

Baseline Model Results

	BASELINE WITHOUT GAZE		
	macro-Precision	macro-Recall	macro-F1 score
Vid+Aud+Text	**0.71**	**0.71**	**0.71**
Vid+Text	0.69	0.69	0.69
Vid+Aud	0.66	0.66	0.66
Aud+Text	0.706	0.706	0.706
Vid	0.59	0.59	0.59
Aud	0.65	0.65	0.65
Text	0.69	0.69	0.69

With Gold Gaze Data

230 samples were split into (171, 59) samples.

The 171 samples were replicated 5 times since eye tracking data of 5 annotators is being used for same set of sentences.

The remaining 59 new instances were used for testing.

Feed forward neural network with fused inputs from text, audio, video along with gaze.

Gold gaze Model Results

WITH GOLD GAZE DATA(231 SAMPLES)			
	macro-Precision	macro-Recall	macro-F1 score
Vid+Aud+Text +	0.82	0.82	0.82
Vid+Text + Gaze	0.8	0.8	0.8
Vid+Aud + Gaze	0.79	0.78	0.78
Aud+Text + Gaz	0.83	0.83	0.83
Vid + Gaze	**0.86**	**0.86**	**0.86**
Aud + Gaze	0.77	0.77	0.77
Text + Gaze	0.83	0.83	0.83
Gaze	**0.92**	**0.92**	**0.91**

Gaze Feature Prediction

We predict gaze features for rest 971 samples using Feed forward Neural Networks for each gaze feature.

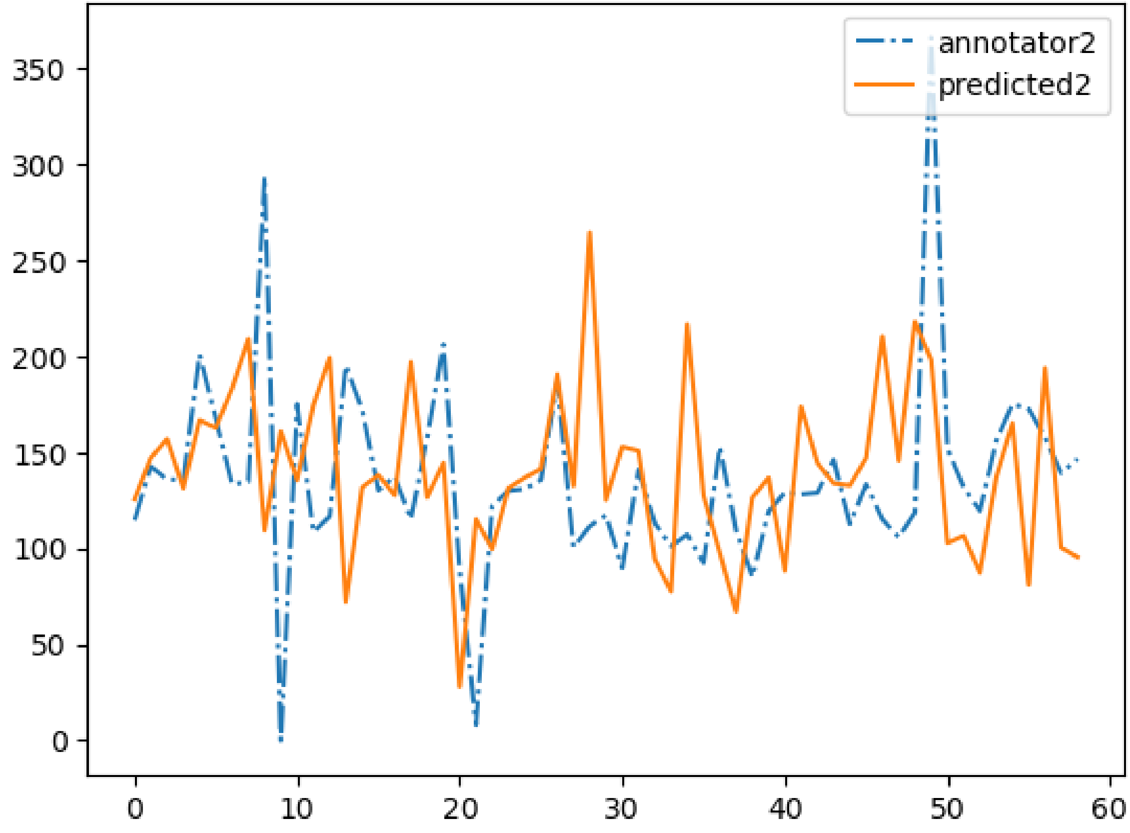
Trained a deep learning model for each of 25 features prediction.

Complete 1024 sized text feature vectors were used as input to the NN and gaze feature value was the output.

Ground truth gaze features were used as labels for the prediction task of each feature

Predicted and Gold gaze Comparison

The predicted Average fixation duration is compared here with the original Average Fixation duration for annotator2.



X axis : Test Sentence id

Y axis: feature value

Performance using Predicted gaze

-Predicted Gaze data for 971 samples is being used in this setting

- The 971 samples were split into (750,221) samples.

- The 750 samples are used for training.
- The remaining 221 new instances were used for testing.

- Feed forward neural network is being used which takes in the fused inputs from text, audio, video along with predicted gaze.

Predicted gaze Model Results

WITH PREDICTED GAZE FEATURES			
	macro-P	macro-R	macro-F1
Vid+Aud+Text + PredGaze	**0.684**	**0.684**	**0.684**
Vid+Text + PredGaze	0.646	0.646	0.646
Vid+Aud + PredGaze	0.641	0.641	0.641
Aud+Text + PredGaze	0.678	0.678	0.678
Vid + PredGaze	0.597	0.597	0.596
Aud + PredGaze	0.643	0.643	0.643
Text + PredGaze	0.663	0.663	0.664

Performance using Predicted and Gold gaze combined

- The training set has 1600 samples, completely shuffled and having samples with both gold and predicted gaze features

- The test set has 500 samples.

- Feed forward neural network is being used which takes in the fused inputs from text, audio, video along with predicted gaze.

Predicted + Gold gaze Model Results

WITH PREDICTED AND GOLD GAZE FEATURES			
	macro-P	macro-R	macro-F1
Vid+Aud+Text + PredGaze + Ga	**0.7215**	**0.7215**	**0.7215**
Vid+Text + PredGaze + Gaze	0.7105	0.7105	0.7105
Vid+Aud + PredGaze + Gaze	0.695	0.695	0.695
Aud+Text + PredGaze + Gaze	0.704	0.704	0.704
Vid + PredGaze + Gaze	0.6615	0.6595	0.659
Aud + PredGaze + Gaze	0.6745	0.6745	0.674
Text + PredGaze + Gaze	0.709	0.709	0.709
PredGaze + Gaze	0.63	0.6295	0.6295

Test of Significance: Two Sample students T-test

NULL Hypothesis: The group means are equal(samples represent same population)

Alternative Hypothesis: The groups have unequal means

In case of two sample independent T test: *p* value is a probability that represents how similar or different the two samples are from each other.

If the p value < Significance level, then the two samples are significantly different.

Test of Significance: Gaze makes a difference

T- Test performed on 18 samples each of sarcastic and non sarcastic data instances for the following Feature :

Average Fixation Duration : Average duration (in milliseconds) of --
-all fixations in the trial(ET experiment)

Significan--ce value	Annotator	Mean(sarcastic)	Mean(non sarcastic)	p- value
0.05	P1	258.3042	235.8321	0.02912
	P2	221.9011	205.6732	0.03229
	P3	218.7231	199.2187	0.03842
	P4	243.86	211.7013	0.01286
	P5	228.276	209.2134	0.03982

Conclusions

- AI→NLP→SA→Sarcasm chain
- General SA does not work well for Sarcasm
- General Sarcasm does not work well for numerical sarcasm
- Rich feature set needed: surface to deeper intent incongruity
- Success from Deep Learning
- Cognition signals help boost accuracy; is realistic as eye tracking is integrated with smart phones
- Has societal applications in mental health monitoring and creating "agony-aunt-bots" (our work in NAACL 2022, COLING 2022)
- Need for zero-shot, few-shot and meta learning

Towards End-to-end Motivational Dialogue System: An Application of Sentiment Analysis and Natural Language Generation to Mental Health

Tulika Saha, Saichethan Miriyala Reddy, Anindya Sundar Das, Sriparna Saha, Pushpak Bhattacharyya, *A Shoulder to Cry on: Towards A Motivational Virtual Assistant for Assuaging Mental Agony*, **NAACL 2022**, Seattle, July 10-15, 2022.

Motivation

- Global burden of mental illness
- Shortage of Mental Health Professionals
- Need Virtual Agents (VAs)
- Challenge: Lack of high quality conversational data

Contributions

- Datasets: *MotiVAte* and *Counsel-VA* consisting of dyadic conversations for depression/multiple mental illnesses between users and the VA prepared with actual conversations collected from mental health forums

- Mental disorders from counselling conversations: Dual attention (self and cross) based DNN classifier built on top of BERT for modelling conversations

- Sentiment driven response generation for motivational conversations on top of the GPT-2 model

MotiVAte and Counsel-VA Datasets

- MotiVAte: 5k dyadic conversations amounting to a total of 18,750 utterances between the depressed users and the VA imparting appropriate suggestion, hope and motivation

- Counsel-VA: 4046 dyadic conversations belonging to three mental disorder categories- Major Depressive Disorder (MDD), Anxiety, Obsessive Compulsive Disorder (OCD)

MotiVAte and Counsel-VA Datasets

Sample Conversation-1

unwanted > My 17 year old son got made and throw some thing at me. I did not call the police. I don't want him to go to jail…….If you know what I am going through I would love it if you would talk to me and help me understand. I feel like such a bad person all of the time.

VA > You are not a bad person! You did what you felt was best for you and your son. Don't let what your husband tell you get to you……Let him know how you feel. It's going to be tough but it is what needs to be done! Take good care of you! *hugs*.

unwanted > Thanks. I have tried telling him how I feel……..I don't know what else to do to let him know that I am just tired of the fighting.

***VA* >** Have you tried family therapy? Maybe that might help you to get through this. Your family needs to know when to stop………Hope your day gets better!

.

.

***VA* >** You should stop doing those things for him! He needs to learn to take care of himself…..It's called tough love, we all have to do it at some point and time. Hang in there!

Sample Conversation-2

pngindreamr > Is excessive guilt associated with OCD? I feel guilty about things that happened 3+ years ago……….. …….but I treat my mistakes like They are the most evil horrible things! Please help if you have any advice.

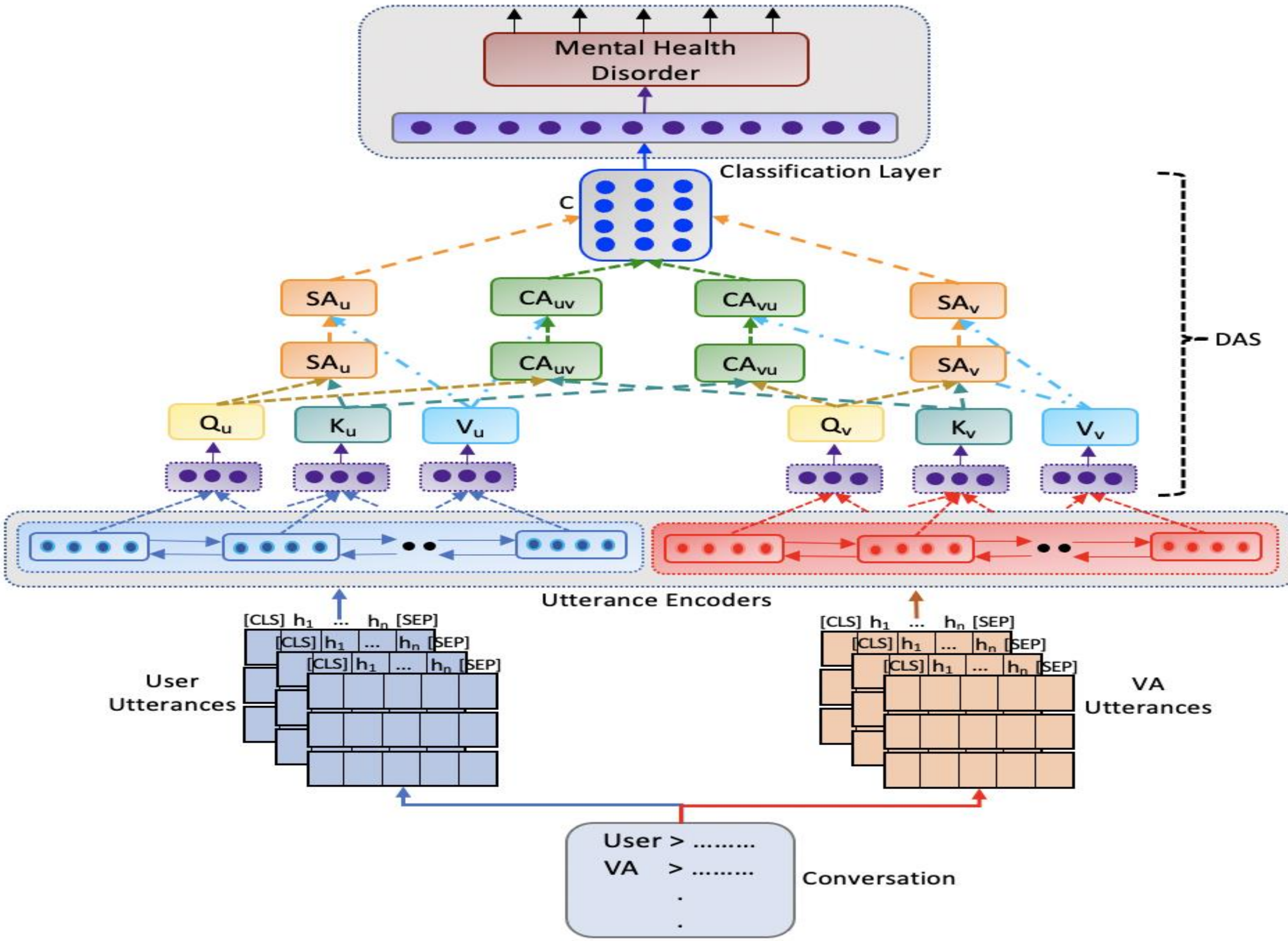
VA > Yes, people with OCD often feel an excessive guilt.Have you been diagnosed with OCD? one possible thing to do is to ignore our annoying thoughts, but if it doesn't work for you maybe you could try other methods… do you know a therapist who could suggest you something?

pngindreamr > I have panic attacks and ocd. According to 2 therapists I've seen. They did recommend…………I guess I just really want it to be gone. :/

***VA* >** You should be consistent with your efforts. It is going to take time, you have to be consistent. Be regular with therapy, try as much as you can to ignore the intrusive thoughts. Remember that they are just some random……….. Good luck!

Sample conversations from MotiVAte and Counsel-VA datasets

Classification of mental health disorders: Cross Attention and Self Attention



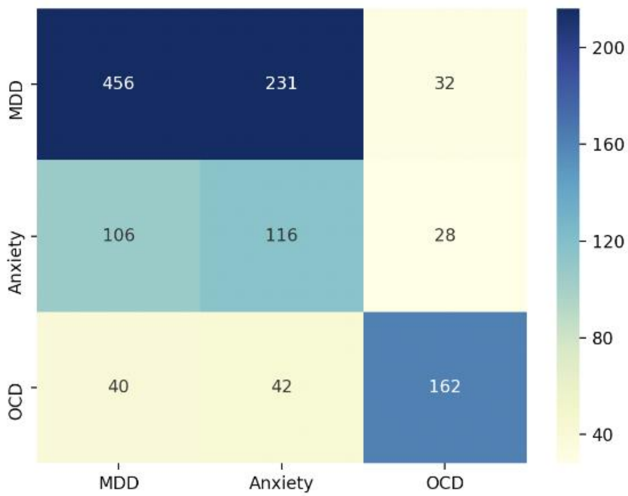
Response Generation

- The GPT-2 model is initially fine-tuned in a supervised manner to generate semantically plausible responses
- Next, with the help of the learnt parameters, the model is tuned again to learn a policy that maximizes long-term future rewards
- BLEU: n-gram matching
- ROUGE-L: matching of the longest common sub-sequence
- Sentiment Score

Results: Mental Disorder Classification

Model	Accuracy	Precision	Recall	F1-score
CNN	52.46	0.4550	0.5515	0.4008
Bi-GRU	53.72	0.4612	0.5576	0.4052
Bi-LSTM	54.27	0.4685	0.5647	0.4135

Table 1 : Results of simple baseline models with GloVe embeddings without any DAS or sentiment based scores



Confusion Matrix for the proposed model

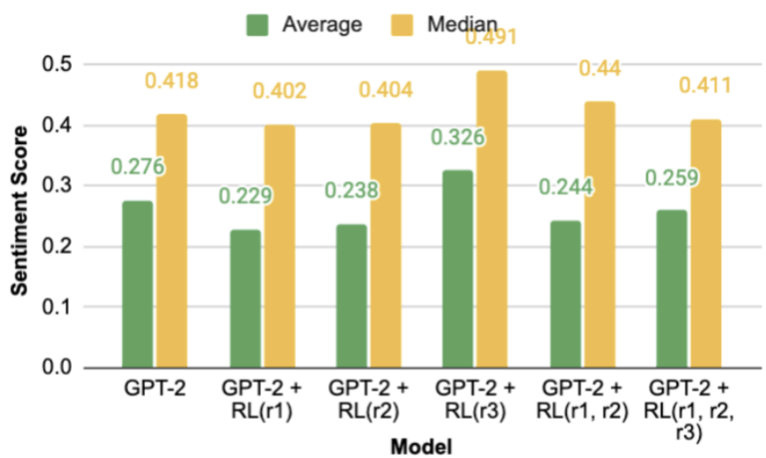
Model	Accuracy	Precision	Recall	F1-score
BERT+CNN (NA+NSenti)	55.83	0.4720	0.5625	0.4112
BERT+CNN+Senti (NA)	58.54	0.5562	0.5688	0.5390
BERT+Bi-GRU (NA+NSenti)	56.02	0.4785	0.5680	0.4211
BERT+Bi-GRU+Senti (NA)	57.33	0.5270	0.5753	0.4677
BERT+Bi-LSTM (NA+NSenti)	56.68	0.4872	0.5745	0.4281
BERT+Bi-LSTM+Senti (NA)	59.21	0.5601	0.5703	0.5427
BERT+Bi-LSTM+Senti (only SA)	59.61	0.5617	0.5725	0.5436
BERT+Bi-LSTM+Senti (only CA)	59.43	0.4800	0.5888	0.5411
Proposed Model	**60.49**	**0.5730**	**0.6016**	**0.5640**

Table 2 : Results of the proposed model and its varying combinations

Results: Response Generation

Model	Embedding Metric			Perplexity	BLEU-1	ROUGE-L
	Average	Extrema	Greedy			
SEQ2SEQ	0.610	0.314	0.403	53.81	0.076	0.059
HRED	0.681	0.327	0.418	67.20	0.093	0.077
GPT-2	0.731	0.371	0.470	68.03	0.139	0.104
GPT-2 + RL(r_{1})	0.726	0.370	0.468	69.12	0.136	0.103
GPT-2 + RL(r_{2})	0.734	0.368	0.476	55.70	0.141	0.110
GPT-2 + RL(r_{3})	0.723	0.367	0.467	60.34	0.137	0.104
GPT-2 + RL(r_{1}, r_{2})	0.732	0.374	0.474	64.62	0.142	0.109
GPT-2 + RL(r_{1}, r_{2}, r_{3})	**0.733**	**0.377**	**0.478**	**50.90**	**0.142**	**0.111**

Table 3 : Automatic evaluation results of the baselines and the proposed models



Link

Sentiment polarity results of the generated VA utterances for different models

AI-ML for Mental Distress

Psychology Types

https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/home.htm

Psychological Types

- Introduced in the 1921 by Carl G. Jung, one of the famous psychologists-trio: Adler, Jung and Freud
- During World War II, in 1940s and 50s, two American women, Isabel Briggs Myers and her mother Katharine Cook Briggs, developed the MBTI tool
- Millions of people worldwide have taken the Indicator each year since its first publication in 1962.

Jung's Theory

- Essence: Much seemingly random variation in the behavior is actually quite orderly and consistent, being due to basic differences in the ways individuals prefer to use their perception and judgment.
- "Perception involves all the ways of becoming aware of things, people, happenings, or ideas. Judgment involves all the ways of coming to conclusions about what has been perceived. If people differ systematically in what they perceive and in how they reach conclusions, then it is only reasonable for them to differ correspondingly in their interests, reactions, values, motivations, and skills."

MBTI and 16 Personality Types (1/2)

- The Myers Briggs Type Indicator (or MBTI for short) is a personality type system
- It divides everyone into 16 distinct personality types across 4 axis:
  - **Introversion (I) – Extroversion (E)**
  - **Intuition (N) – Sensing (S)**
  - **Thinking (T) – Feeling (F)**
  - **Judging (J) – Perceiving (P)**

MBTI and 16 Personality Types (2/2)

- So for example, someone who prefers **introversion, intuition, thinking and perceiving** would be labelled an **INTP** in the MBTI system
- There are lots of personality based components that would model or describe this person's preferences or behaviour based on the label.

Elaboration of Types (1/2)

- Favorite world: Do you prefer to focus on the outer world or on your own inner world? This is called **Extraversion (E) or Introversion (I).**

- Information: Do you prefer to focus on the basic information you take in or do you prefer to interpret and add meaning? This is called **Sensing (S) or Intuition (N).**

Elaboration of Types (1/2)

- Decisions: When making decisions, do you prefer to first look at logic and consistency or first look at the people and special circumstances? This is called **Thinking (T) or Feeling (F)**.
- Structure: In dealing with the outside world, do you prefer to get things decided or do you prefer to stay open to new information and options? This is called **Judging (J) or Perceiving (P).**

Elaboration of Types (1/2)

- Decisions: When making decisions, do you prefer to first look at logic and consistency or first look at the people and special circumstances? This is called **Thinking (T) or Feeling (F)**.
- Structure: In dealing with the outside world, do you prefer to get things decided or do you prefer to stay open to new information and options? This is called **Judging (J) or Perceiving (P).**

Comparison of Personality Types (1/2)

Extraversion

1. Outwardly directed energy needed to move into action

2. Responsiveness to what is going on in the environment

3. A natural inclination to converse and to network

Introversion

1. Inwardly directed energy needed for focused reflection

2. Stability from attending to enduring ideas

3. A natural tendency to think and work alone

Sensing

1. A mastery of the facts

2. Knowledge of what materials and resources are available

3. Appreciation of knowing and doing what works

Intuition

1. Insight and attention to meanings

2. A grasp of what is possible and what the trends are

3. Appreciation of doing what hasn't been tried before

Comparison of Personality Types (2/2)

Thinking
Analysis of the pros and cons of situations, even when they have a personal stake
An ability to analyze and solve problems
Want to discover the “truth” and naturally notice logical inconsistencies

Feeling
Knowledge of what is important to people and adhere to that in the face of opposition
The ability to build relationships and to be persuasive
Desire to uncover the greatest good in a situation and notice when people may be harmed

Judging
Organization, planning, and follow through on projects
Push to get things settled and decided
Appreciation of well-organized efficiency

Perceiving
Quickly and flexibly responding to the needs of the moment
Strive to keep things open so new information may be gathered
Appreciation of the need for spontaneity and exploration

16 Personality Types

Managers, Administrators, and Supervisors

N = 4808

		Sensing		Intuition	
		Thinking	Feeling	Feeling	Thinking
Introversion	Judgment	ISTJ N = 935 % = 19.45	ISFJ N = 261 % = 5.43	INFJ N = 124 % = 2.58	INTJ N = 392 % = 8.15
Introversion	Perception	ISTP N = 175 % = 3.64	ISFP N = 80 % = 1.66	INFP N = 130 % = 2.70	INTP N = 280 % = 5.82
Extraversion	Perception	ESTP N = 158 % = 3.29	ESFP N = 93 % = 1.93	ENFP N = 203 % = 4.22	ENTP N = 285 % = 5.93
Extraversion	Judgment	ESTJ N = 786 % = 16.35	ESFJ N = 218 % = 4.53	ENFJ N = 177 % = 3.68	ENTJ N = 511 % = 10.63

Note: ■ = 1% of sample

Source: Gerald P. Macdaid, CAPT Data Bank, 1997, Center for Applications of Psychological Type, Inc.

EM-PERSONA: EMotion-assisted Deep Neural Framework for PERSONAlity Subtyping from Suicide Notes

-Ghosh, Ekbal, Bhattacharyya, COLING 2022

Contributions

- **Data**: Existing suicide note corpora (Ghosh et al., 2020, 2022) are annotated at the sentence level with personality types.

- **Model**: End-to-End multi-task emotion-assisted system for simultaneous detection of types from suicide notes.

Task Definition

- **Input**: Suicide note (N) with each sentence annotated with an emotion class
- **Output**: classify the author of the note into one of the two categories for each of the following personality dichotomies: (I/E), (N/S), (F/T), (J/P).

Model Architecture

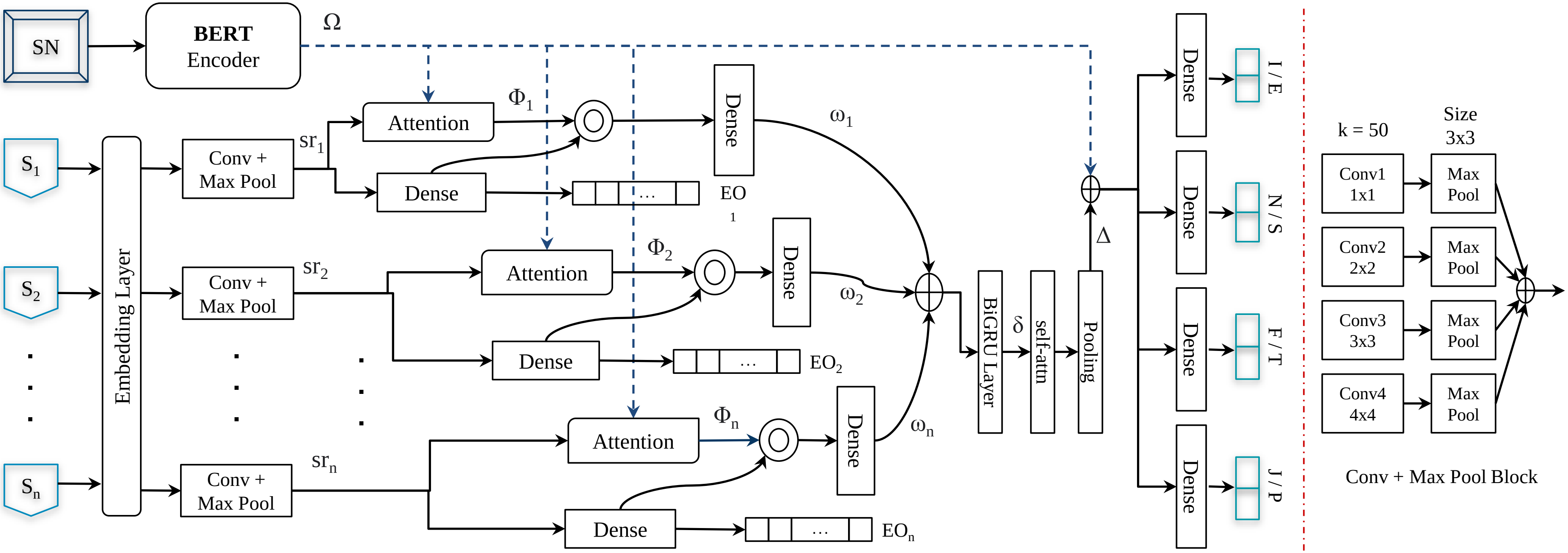


Figure 1: Architecture of the EMotion-assisted deep neural framework for PERSONAlity Subtyping.

Dataset (1/2)

- Dataset: CEASE-v2.0 dataset (Ghosh et al., 2022)
- 4932 sentences collected from 325 real-life suicide notes
- Average Kappa Agreement over the four personality dichotomies is 0.61

Results

Models	F1$^{\text{I-E}}$	F1$^{\text{N-S}}$	F1$^{\text{F-T}}$	F1$^{\text{J-P}}$
Single-task Baselines				
HAN	45.4	48.1	43.87	36.6
CNN+cLSTM	44.5	43	44.5	36.1
BERT	44.87	43	39.88	49.36
RoBERTa	44.46	39.88	42.69	50.58
Multi-task Baselines				
MT-BERT	44.35	42.68	39.90	47.31
Proposed Multi-task Approach				
EM-PERSONA	**47.44**	**51.79**	**49.02**	**54.00**
Ablation (Removing Emotion Task)				
EM-PERSONA [-Emo]	45.53	50.27	46.96	51.40

Table 3: Scores from 10-fold cross-validation experiments.
MT: Multi-task learning.

Sample Predictions

Category	*Note Excerpts*	*Actual*	MT-BERT	*EM-PERSONA*
BL & PP: FC	After many hours of thought and meditation, I have made a decision that should not be an example to anyone else ... Please tell my story on every radio and television station and in every newspaper and magazine ... to those of you who are shallow the events of this morning will be that story ... <NAME>, love you	E S F P	E S F P	E S F P
BL: PC PP: FC	.... If we had a problem it is because I loved her so much. ... we came to the understanding that for now we were not right for each other ... Unlike what has been written in the press, <NAME> and I had a great relationship for most of our lives together ... most of it ... is totally made up.	I S F J	E S F P	I S F J
BL: IC PP: PC	You have always been my soul mate and I want you to love life and know I am always with you. ... your characteristic is that of a true angel and the definition of god's love! This was the supreme Almighty's plan not mine! Look after <NAME> and <NAME> for me they are my boys you are rich. ...	E N T P	I S F J	E S F P
BL & PP: PC	Dear Mum, I am really sorry that I did this. Do not you ever think it was your fault. ... I love you so much and I could not ask for a better mum. Thank you for caring and feeding and loving me for 14 years. ... my heart cannot take this pain . I am going to miss you so much. ... I will be waiting at heaven's gates for you. ...	I S F J	E S T P	E S F P

Table 4: Sample predictions by the MT-BERT and EM-PERSONA systems over various categories.

BL: baseline MT-BERT, PP: proposed EM-PERSONA, PC: partially correct, FC: fully correct, IC: fully incorrect.

I: Introversion, E: Extraversion, N: Intuition, S: Sensing, T: Thinking: F: Feeling, J: Judging, P:

Use of Cognitive NLP

Introduce cognitive features

- Derive and augment cognitive features with traditional textual features.

- **Why?:** Textual nuances affect gaze (Just and Carpenter, 1979; Rayner, 1998)

- **Feasibility:** Inexpensive eye-tracking hardware available and integrated with handheld gadgets (e.g.,http://www.sencogi.com)

Eye tracking

Saccades

Fixations

Image Courtesy: http://www.smashingmagazine.com/2007/10/09/30-usability-issues-to-be-aware-of/

Eye Tracking Machines

Most comfortable technique to measure gaze based on infrared light

A bit more complicated way to measure gaze using electric potential around the eye.

The eye tracking glasses are used for broad range of mobile eye tracking studies.

The ergonomic chin rest eye tracking device for high speed and accurate measurements with a large visual field.

Image courtesy: www.smivision.com

Eye tracking on mobile phones

- Samsung Galaxy S4 comes with eye tracking capability
- The software umoove (http://www.umoove.me/) runs on mobile phones, tracking eyes

- **MIT Technogy Review, June 2015:**
  - **"Eye-tracking system uses ordinary cellphone camera"**

Eye Tracking: basic parameters

- **Gaze points**:
  - Position of eye-gaze on the screen
- **Fixations:**
  - A long stay of the gaze on a particular object on the screen. Fixations have both Spatial (coordinates) and Temporal (duration) properties.
- **Saccade:**
  - A very rapid movement of eye between the positions of rest.
- **Scanpath:**
  - A path connecting a series of fixations.
- **Regression:**
  - Revisiting a previously read segment

Use of eye tracking

- Used extensively in Psychology
  - Mainly to study reading processes
  - Seminal work: Just, M.A. and Carpenter, P.A. (1980). A theory of reading: from eye fixations to comprehension. Psychological Review 87(4):329–354
- Used in flight simulators for pilot training
- Website developers use eye tracking to improve look and feel of websites

Eye tracking usage

Our contribution:

(a) Better measures of Readability

(b)Use of eye tracking in NLP- cognitive NLP

NLP-ML and Eye Tracking

- Kliegl (2011)- Predict word frequency and pattern from eye movements
- Doherty et. al (2010)- Eye-tracking as an automatic Machine Translation Evaluation Technique
- Stymne et al. (2012)- Eye-tracking as a tool for Machine Translation (MT) error analysis
- Dragsted (2010)- Co-ordination of reading and writing process during translation.

Relatively new and open research direction

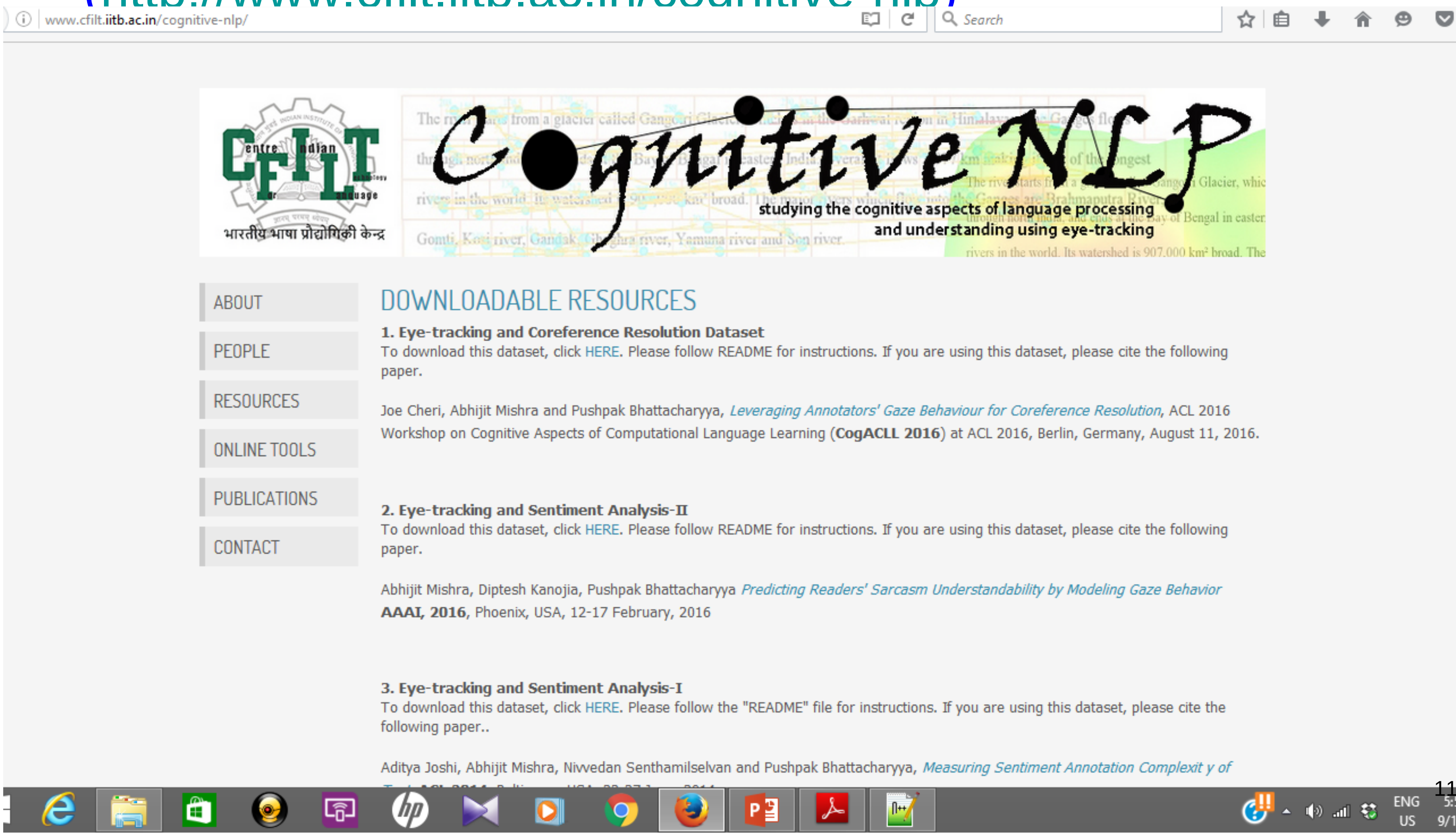
Our lab (CFILT@IITB) has been Contributing

- Joshi, Aditya and Mishra, Abhijit and S., Nivvedan and Bhattacharyya, Pushpak. 2014. Measuring Sentiment Annotation Complexity of Text. Association for Computational Linguistics, **(ACL 2014)** Baltimore, USA.
- Mishra, Abhijit and Bhattacharyya, Pushpak and Carl, Michael. 2013. Automatically Predicting Sentence Translation Difficulty.Association for Computational Linguistics **(ACL 2013)**, Sofia, Bulgaria

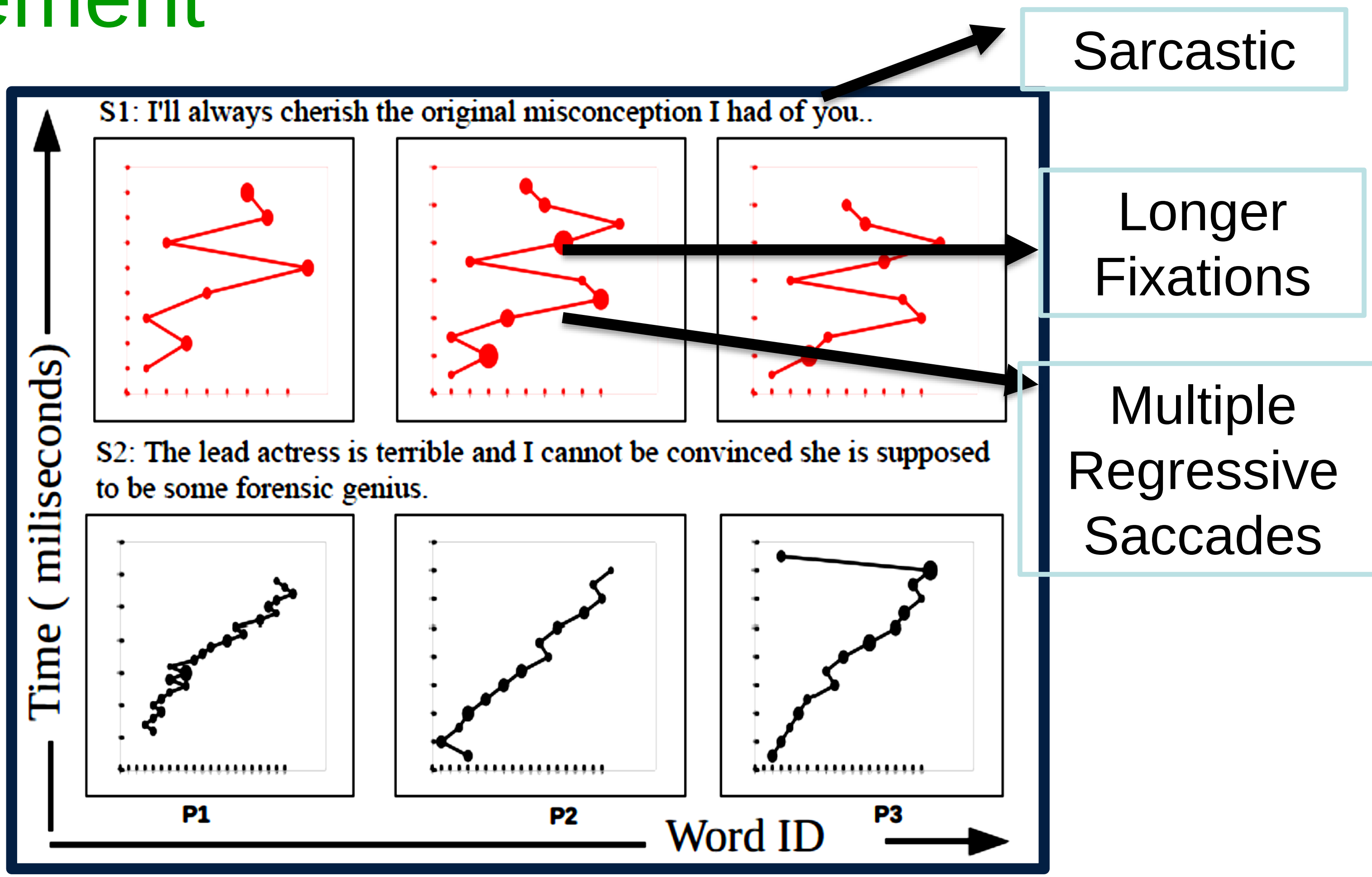
Contribution to NLP Community

Publicly available datasets and tools

(http://www.cfilt.iitb.ac.in/cognitive-nlp)



Sentiment Annotation and Eye Movement



Datasets

- Two publicly available datasets released by us (Mishra et al, 2016; Mishra et al., 2014)
- **Dataset 1: (Eye-tracker: Eyelink-1000 Plus)**
  - 994 text snippets : 383 positive and 611 negative, 350 are sarcastic/ironic
  - Mixture of Movie reviews, Tweets and sarcastic/ironic quotes
  - Annotated by 7 human annotators
  - Annotation accuracy: **70%-90%** with Fleiss kappa IAA of **0.62**
- **Dataset 2: (Eye-tracker: Tobi TX300)**
  - 843 snippets : 443 positive and 400 negative
  - Annotated by 5 human subjects
  - Annotation accuracy: **75%-85%** with Fleiss kappa IAA of **0.68**

Accuracy of Traditional Classifiers on our Datasets

- Trained Naïve Bayes and SVM using 10662 short text and traditional features (**Liu and Zhang, 2012)**
- Classifiers tried: Naïve Bayes, SVM and Rule Based

	NB			SVM			RB		
	P	R	F	P	R	F	P	R	F
D1	66.15	66	**66.15**	64.5	65.3	**64.9**	56.8	60.9	**53.5**
D2	74.5	74.2	**74.3**	77.1	76.5	**76.8**	75.9	53.9	**63.02**

Features for SA (Textual)

Presence of Unigrams (NGRAM_PCA)
Count of Subjective Words (Positive_words, Negative_words)
Subjective Score from SentiWordNet (PosScore, NegScore)
Sentiment Flip Count (FLIP)
Part of Speech Ratios (VERB, NOUN, ADJ, ADV)
Count of Named Entities (NE)
Count of Discourse Connectors – *e.g., however , although* (DC)

Features for SA (Textual)

- **Sarcasm, Irony and Thwarting related Features (Joshi et al, 2015; Ramteke et al. 2013)**

Presence of Implicitly Incongruous Phrases – Riloff et al. (IMP)
Longest pos/neg subsequence (LAR)
Resultant Lexical Word Polarity of Text (LP)
Punctuations and Inrerjections (PUNC)

- **Features related to reading difficulty**

Flesch Readability Ease (RED)
Total word count (LEN)
Average syllable per word (SYL)

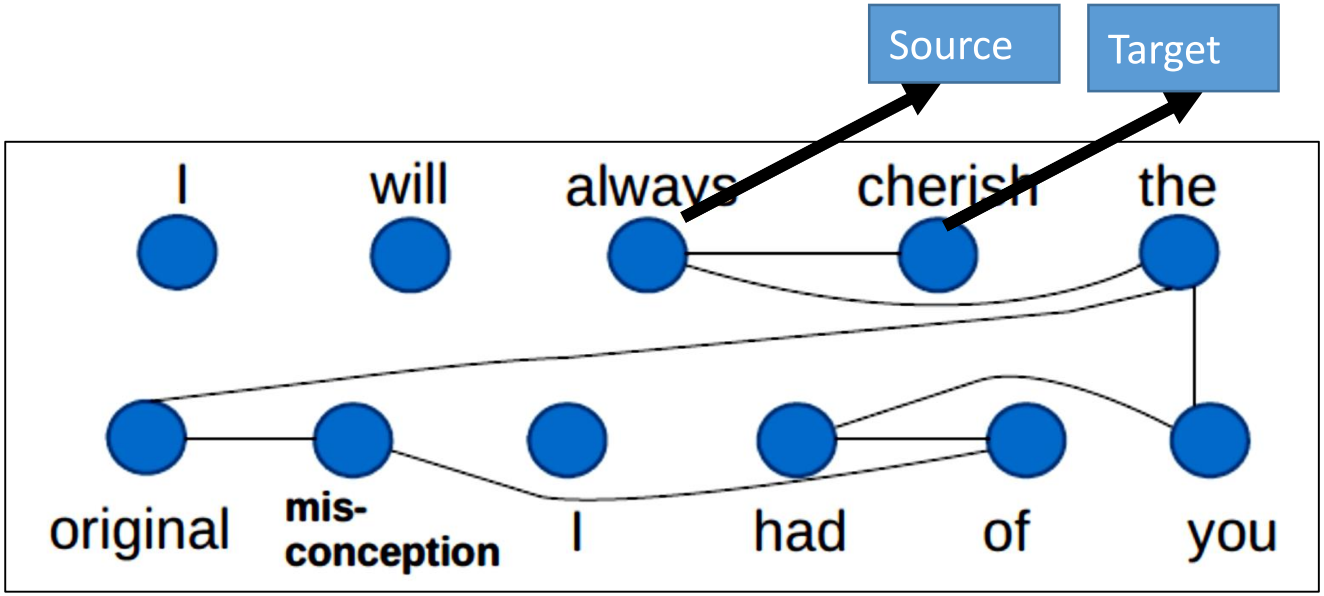
Features for SA (Cognitive)

- Simple Features from Eye-movement (extracted directly from recorded eye-movement data)

Average First Fixation Duration per Word (FDUR)
Average Fixation Count (FC)
Average Saccade Length (SL)
Total Regressive Saccade Count (REG)
Count of Number of Words Skipped (SKIP)
Count of Regressive Saccades from Second Half to First Half of the Text (RSF)
Position of the word from which the largest regression starts (LREG)

Features for SA (Cognitive)

Complex Gaze Features derived from Gaze-saliency graph



Features for SA (Cognitive)

- Features from the Gaze Salency Graph

Edge Density (ED) of the Gaze Saliency Graph
Highest , 2nd Highest Weighted Degree With **Fixation Duration** at Source Node, Target Node as Edge Weight (F1H, F1S, F2H, F2S)
Highest , 2nd Highest Weighted Degree With **Forward Saccade Count** as Edge Weight (FSH, FSS)
Highest , 2nd Highest Weighted Degree With **Forward Saccade Distance** as Edge Weight (FSDH, FSDS)
Highest , 2nd Highest Weighted Degree With **Reverse Saccade Count** as Edge Weight (RSH, RSS)
Highest , 2nd Highest Weighted Degree With **Reverse Saccade Distance** as Edge Weight (RSDH, RSDS)

Why these Gaze features?

- **Key observation from dataset:** Negative sentiment bearing texts are more linguistically subtle (irony, sarcasm, implicit-sentiment)
- **Why simple gaze features?:** Significant variation in gaze attributes (fixation duration, regression count, skip count and observed when text has such subtleties (observed through t-tests). So, our simple gaze features contain important information regarding subtleties.
- **Why complex gaze features?:** When the text has distinct phrases pointing to situational disparities (like incongruity in sarcasm), a lot of regressive saccades around these phases observed, making the gaze saliency graph **Dense** (Captured by Edge Density) and modular (with a few nodes having very large degrees).

Experiment

- **Sentiment Polarity prediction of Snippets :** Binary Classification Problem
- **Classifiers:** Naïve Bayes, Support Vector Machine (With Linear Kernel), Multi-layered Perceptron
- **Evaluation Mode**: 10-Fold Cross validation
- **Feature Combination**
  - Unigram Only (Uni)
  - Sentiment [Includes Unigram Presence] (Sn)
  - Sarcasm, Irony and Thwarting Features [Include Unigram Presence](Sr)
  - Gaze and readability (Gz)

Results

p=0.006

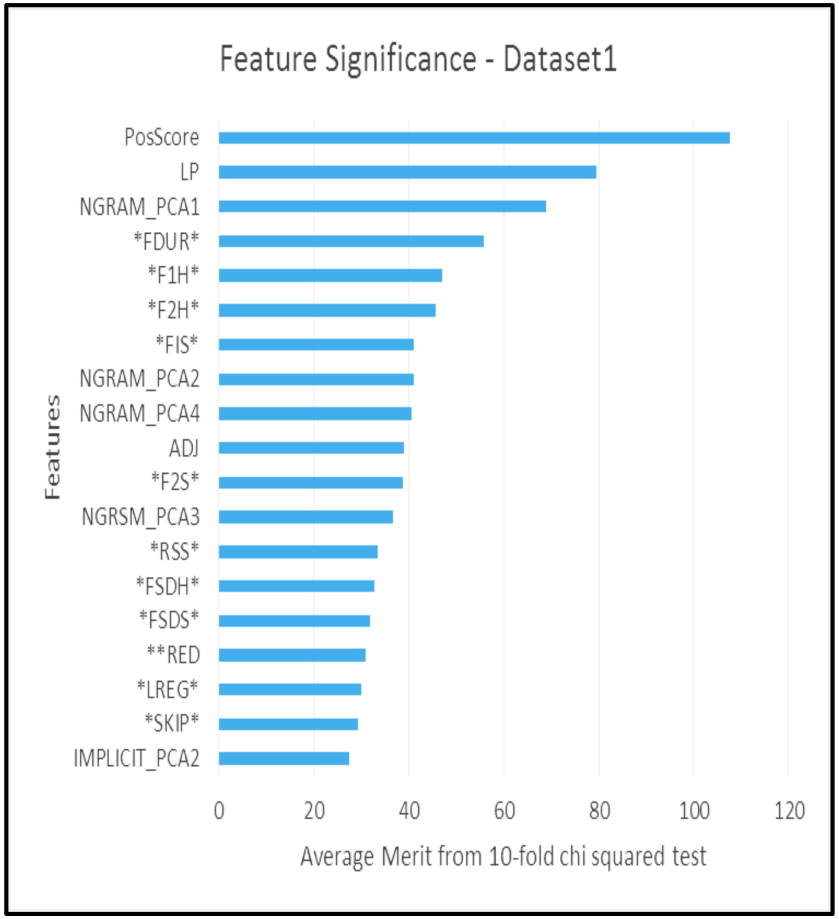
p = 2e-5

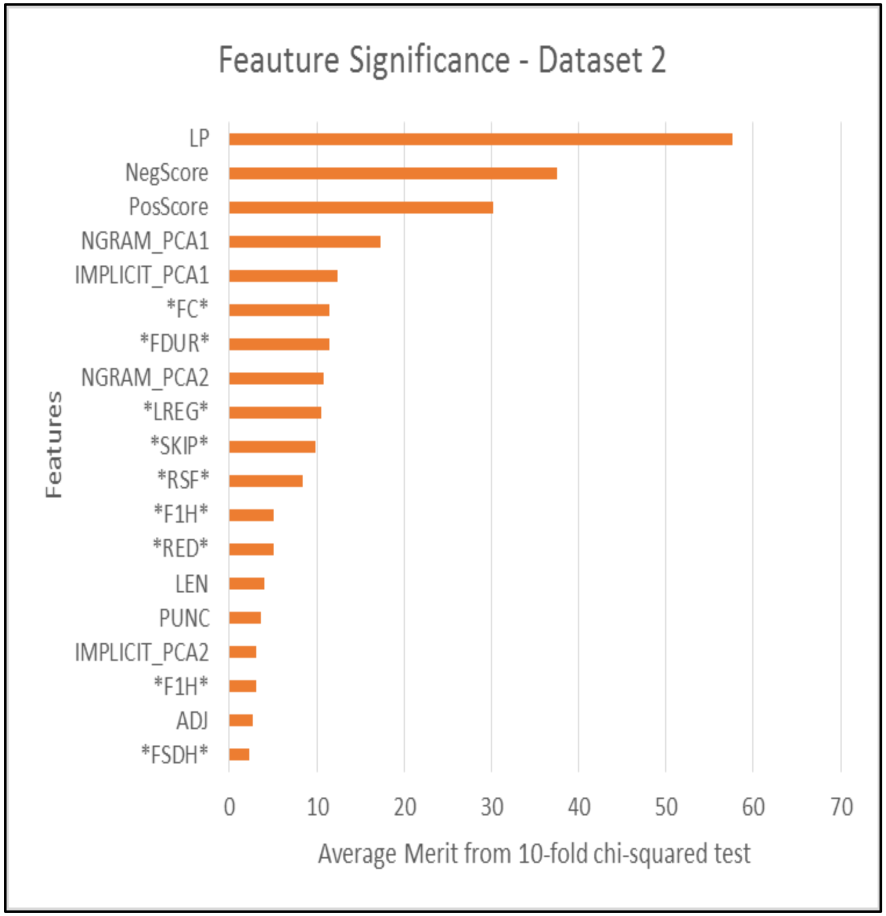
p = ***0.0003,***

p =***0.21***

Classifier	Näive Bayes			SVM			Multi-layer NN		
	Dataset 1								
	P	R	F	P	R	F	P	R	F
Uni	58.5	57.3	57.9	67.8	68.5	68.14	65.4	65.3	65.34
Sn	58.7	57.4	58.0	69.6	70.2	69.8	67.5	67.4	67.5
Sn + Sr	63.0	59.4	61.14	72.8	73.2	72.6	69.0	69.2	69.1
Gz	61.8	58.4	60.05	54.3	52.6	53.4	59.1	60.8	60
Sn+Gz	60.2	58.8	59.2	69.5	70.1	69.6	70.3	70.5	70.4
Sn+ Sr+Gz	**63.4**	**59.6**	**61.4**	**73.3**	**73.6**	**73.5**	**70.5**	**70.7**	**70.6**
	Dataset 2								
Uni	**51.2**	**50.3**	**50.74**	57.8	57.9	57.8	53.8	53.9	53.8
Sn	51.1	50.3	50.7	62.5	62.5	62.5	58.0	58.1	58.0
Sn+Sr	50.7	50.1	50.39	70.3	70.3	70.3	66.8	66.8	66.8
Gz	49.9	50.9	50.39	48.9	48.9	48.9	53.6	54.0	53.3
Sn+Gz	51	50.3	50.6	62.4	62.3	62.3	59.7	59.8	59.8
Sn+ Sr+Gz	50.2	49.7	50	**71.9**	**71.8**	**71.8**	**69.1**	**69.2**	**69.1**

How good are Cognitive Features? – Chi squared test





**Ablation test: No significant differences observed by ablating one feature at a time*

How good are Cognitive Features?- Heldout accuracy

- Dataset-1 split into a train-test split of 760:234 (Out of 234, 131 contain irony/sarcasm)
- We checked how our best performing classifier with different feature combinations perform for both Irony and Non-irony parts.

	Irony	Non-Irony
Sn	58.2	75.5
Sn+Sr	60.1	75.9
Gz+Sn+Sr	64.3	77.6

p = *0.001*

F-scores on texts containing Sarcasm/Irony in Held-out Dataset derived from dataset-1 (Train-test split of 760:234)

Example Sentences

Sentence	Gold	SVM_Ex.	NB_Ex.	RB_Ex.	Sn	Sn+Sr	Sn+Sr+Gz
1. I find television very educating. Every time somebody turns on the set, I go into the other room and read a book	-1	1	1	0	1	**-1**	**-1**
2. I love when you do not have two minutes to text me back.	-1	1	**-1**	1	1	1	**-1**

Discussions: Augmented features for Sarcasm

Textual

(1) Unigrams (2) Punctuations
(3) Implicit incongruity
(4) Explicit Incongruity
(5) Largest +ve/-ve subsequences
(6) +ve/-ve word count
(7) Lexical Polarity
(8) Flesch Readability Ease,
(9) Word count

Simple gaze

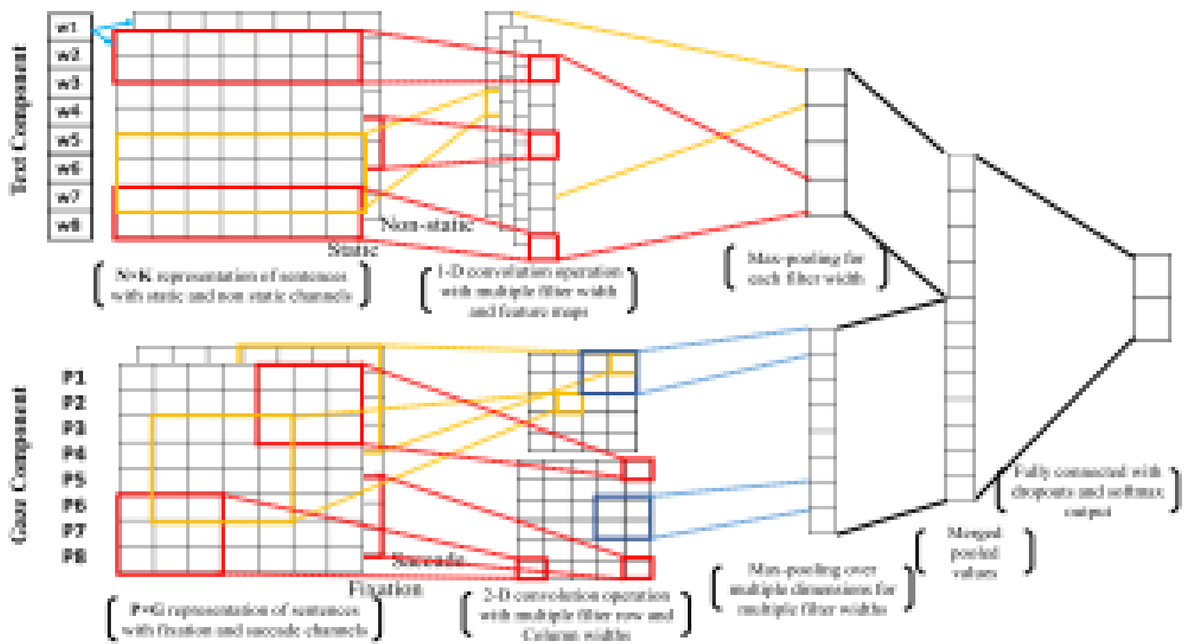
(1) Average Fixation Duration,
(2) Average Fixation Count,
(3) Average Saccade Length,
(4) Regression Count,
(5) Number of words skipped,
(6) Regressions from second half to first half,
(7) Position of the word from which the largest regression starts

Complex gaze

(1) Edge density,
(2) Highest weighted degree
(3) Second Highest weighted degree
(With different edge-weights)

Link-end

Abhijit Mishra, Kuntal Dey and Pushpak Bhattacharyya, *Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification Using Convolutional Neural Network*, **ACL 2017**, Vancouver, Canada, July 30-August 4, 2017.



Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification

- In complex classification tasks like sentiment analysis and sarcasm detection, even the extraction and choice of features should be delegated to the learning system
- The idea of channels in CNN is exploited, and CNN learns features from both gaze and text and uses them to classify the input text

Central Idea

- Learn features from Gaze sequences (fixation duration sequences and gaze-positions) and Text automatically using Deep Neural Networks.

- Deep NNs have proven to be good at learning feature representations for Image and Text classification tasks (Krizhevsky et al., 2012;Collobert et al., 2011).

- Use Convolutional Neural Network (already used for sentiment classification, Kim, 2014)

Summary

- Motivation for VAMH (Virtual Agent for Mental Health)
- NLP, Sentiment and Emotion
- A specific SA challenge: Sarcasm
- Techniques of Sarcasm Detection
- A VA for MH- an Agony Aunt
- Personality Typing from suicide notes
- Challenges:
  - Absence of Data in good quantity and Quality
  - “Physician Heal Thyself!!”- the agent should factual, consistent, non-hallucinating
  - Grapple with Pragmatic Ambiguity (Sarcasm, Irony, Metaphor, Hyperbole, Hate Speech, Profanity, Faking)

Conclusions and Future Work

- Compelling case for VAMH- need of the hour
- Model building to go hand in hand with data creation, curation, dissemination
- Theory and technique building by behavioural study
- Time not yet ripe for prescription: only spread motivation and positivity

- Close interaction with MH professionals needed
- Fine classification of disorders necessary
- Assistant and not competitor

Thank You

http://www.cse.iitb.ac.in/~pb

http://www.cfilt.iitb.ac.in

Related Work

- Qureshi et al., Intelligent System, 2019: Multi-modal depression detection from dialogue
- Gaur et al., CIKM, 2018: Study to understand the mental health of anonymous users of Reddit
- Yazdavar et al., Plos One, 2020: Multi-modal depression detection in Twitter
- Althoff, ACL, 2016: Investigative study on large-scale SMS based counseling conversation
- Dowling and Rickwood, Computers in Human Behavior, 2016: Explored how “hope” and “expectations” affect treatment outcomes in youths suffering from mental health issues
- Ji et al., 2020: Deep Neural Network for identifying users with higher suicide risk as well as identifying the mental illness behind it from Twitter
- Rao et al., Knowledge Science, Engineering and Management, 2020: BERT based ensemble learning classifier for detecting mental health disorders, namely depression and anorexia from daily posts of an online user in social media

Dataset (2/2)

Traits	Distribution	
Introversion (I) / Extraversion (E),	I: 285	E: 71
Intuition (N) / Sensing (S),	N: 90	S: 268
Thinking (T) / Feeling (F)	F: 238	T: 119
Judging (J) / Perceiving (P)	J: 145	P: 214

Table 1: Data distribution over various personality traits.

Emotions	**Sentences**
Forgiveness	44
happiness	100
Hopefulness	353
Love	266
Pride	28
Thankfulness	97
Abuse	34
Anger	154
Blame	208
Fear	62
Guilt	169
Hopelessness	151
Sorrow	720
Information	2180
Instruction	366

Table 2: Data distribution over the emotion classes.

Ode to Scientists and Engineers

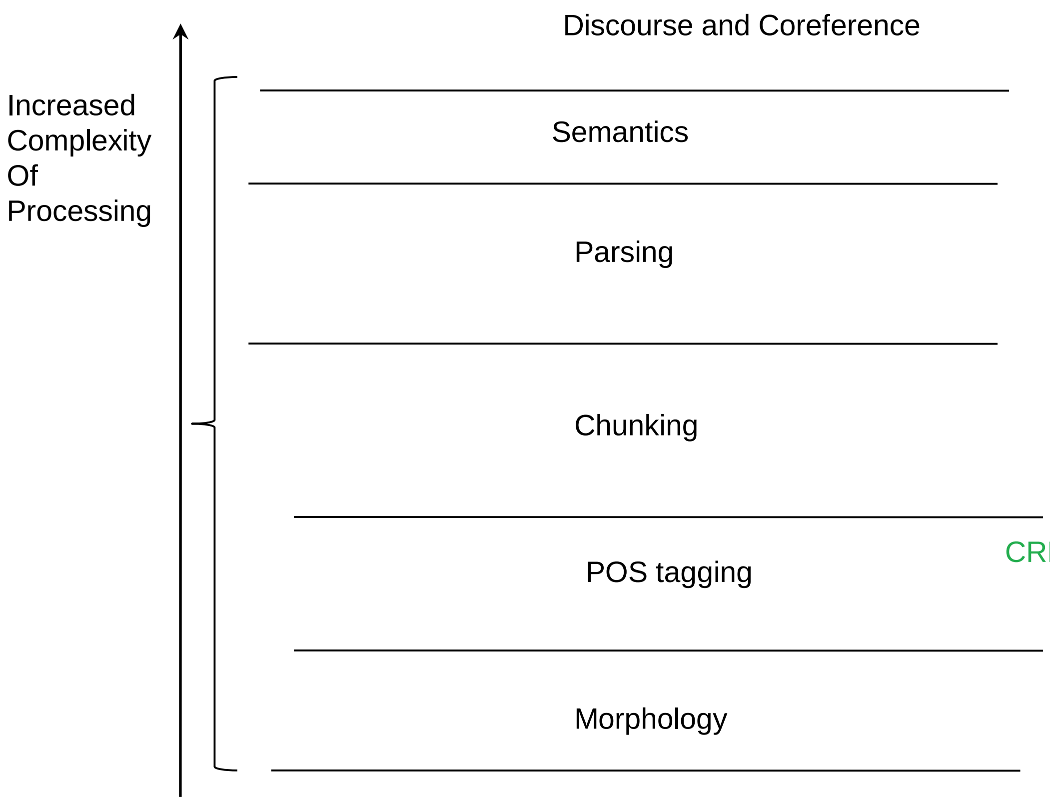
Scientists ask WHY
Engineers ask WHY NOT
Scientists wonder at WHAT-IS
Engineers wonder WHAT-COULD-BE
World couldn’t do without either.

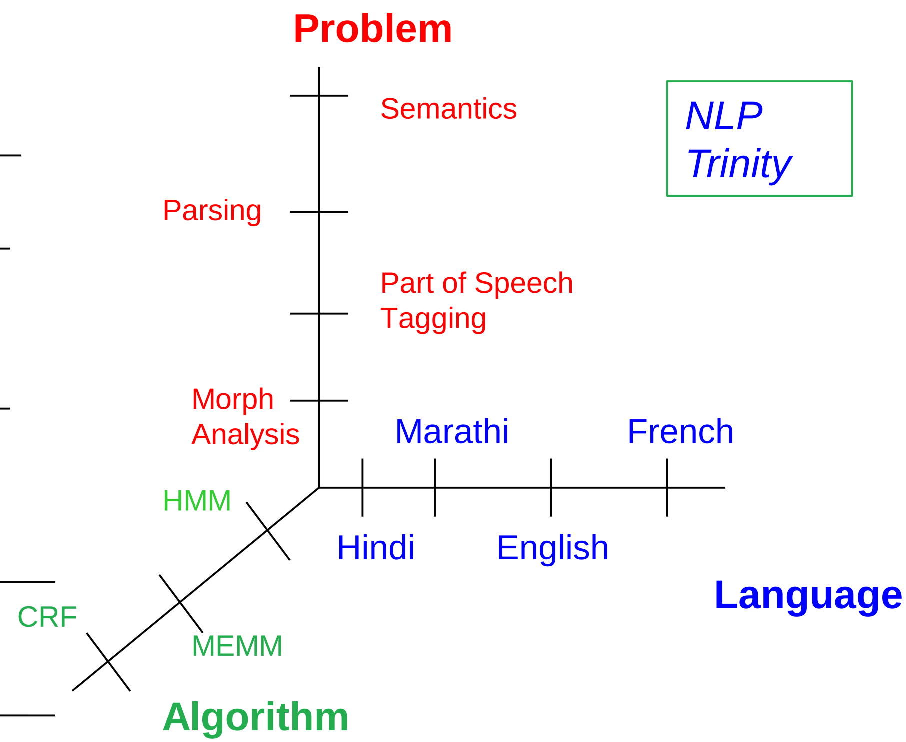
Scientists STUDY
Engineers MAKE
And ever the twain shall meet.

Natural Language Processing

Art, science and technique of making computers understand and generate language

NLP is layered Processing, Multidimensional too





Three Dimensions of NLP: *language, content, emotion*

- Has content, has empathy
- Difference in language leads to communication barrier
- Difference in emotion also leads to communication barrier

Ambiguity is the main challenge!!

<u>Lady A</u>: Yesterday you told me about shop that sells artificial jewellery

<bn>ki naam jeno?</bn> (what did you say the name was?)

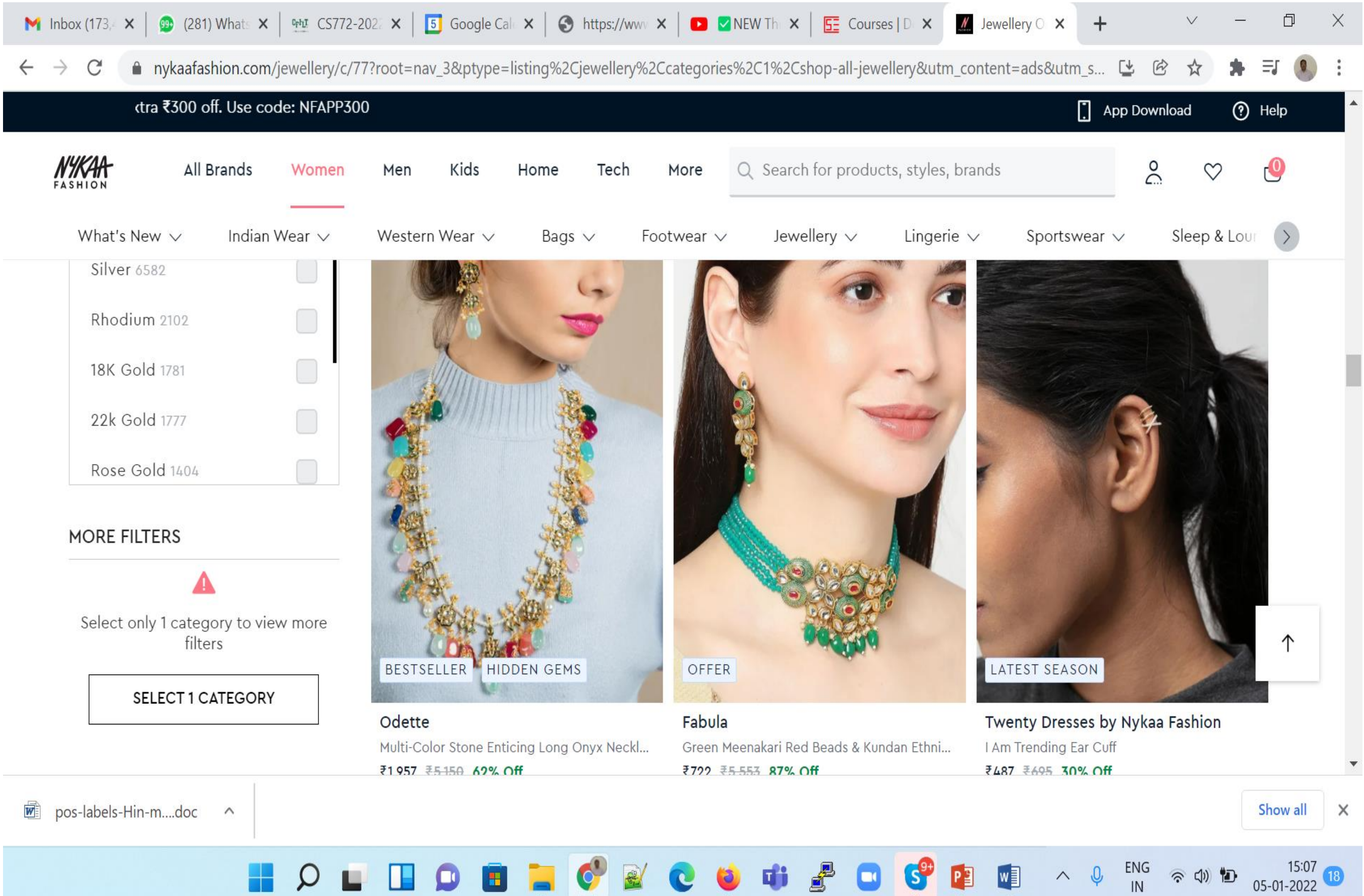
<u>Lady B</u>: nykaa

<u>Lady A</u> (offended): What do you mean Madam? Is this the way to talk?

<u>Lady B</u>: <bn> kena ki holo?</bn> (why what happened?)

… *Lady A did not reply; she was angry!!!*

NYKAA Fashion *(9 billion INR/100 million USD)*



Root cause of the problem: Ambiguity!

- NE-non NE ambiguity (proper noun-common noun)
- Aggravated by code mixing
- “Nykaa”: name of the shop
- Sounds similar to “ন্যাকা” (nyaakaa), meaning somebody “who feigns ignorance/innocence” in a derogatory sense
- An offensive word