Frontiers in Sentiment Analysis

Pushpak Bhattacharyya
CSE Dept.,
IIT Patna and Bombay

Talk at IBM Research-IISc Workshop, Bangalore

7 Mar, 2018

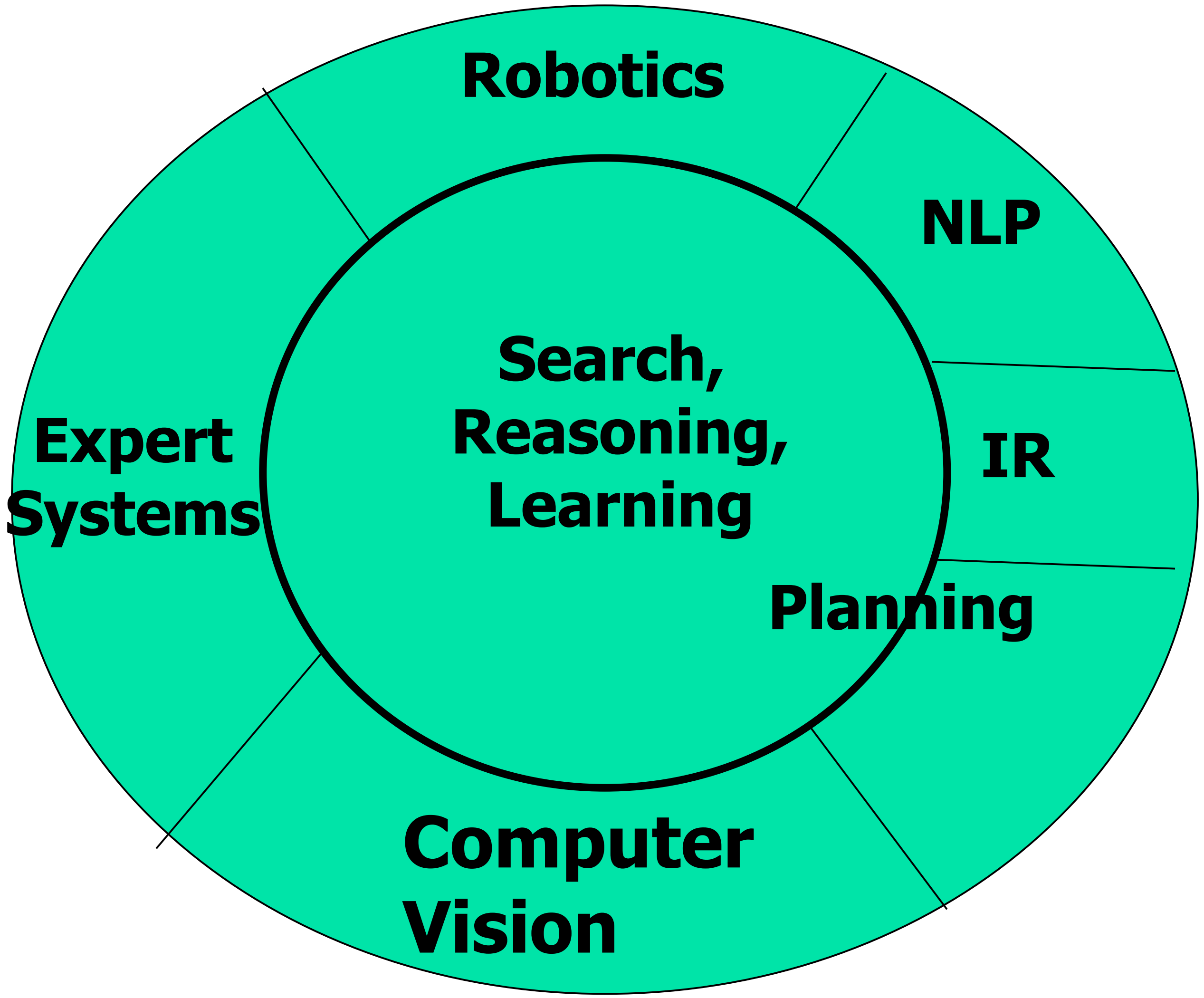
Acknowledgment: studens *Aditya, Raksha, Abhijit, Kevin, Lakshya, Arpan, Vabhav, Prerana, Vinita, Shad* and many, many others


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
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)
ENG
IN
3:20 PM
11-Jul-17

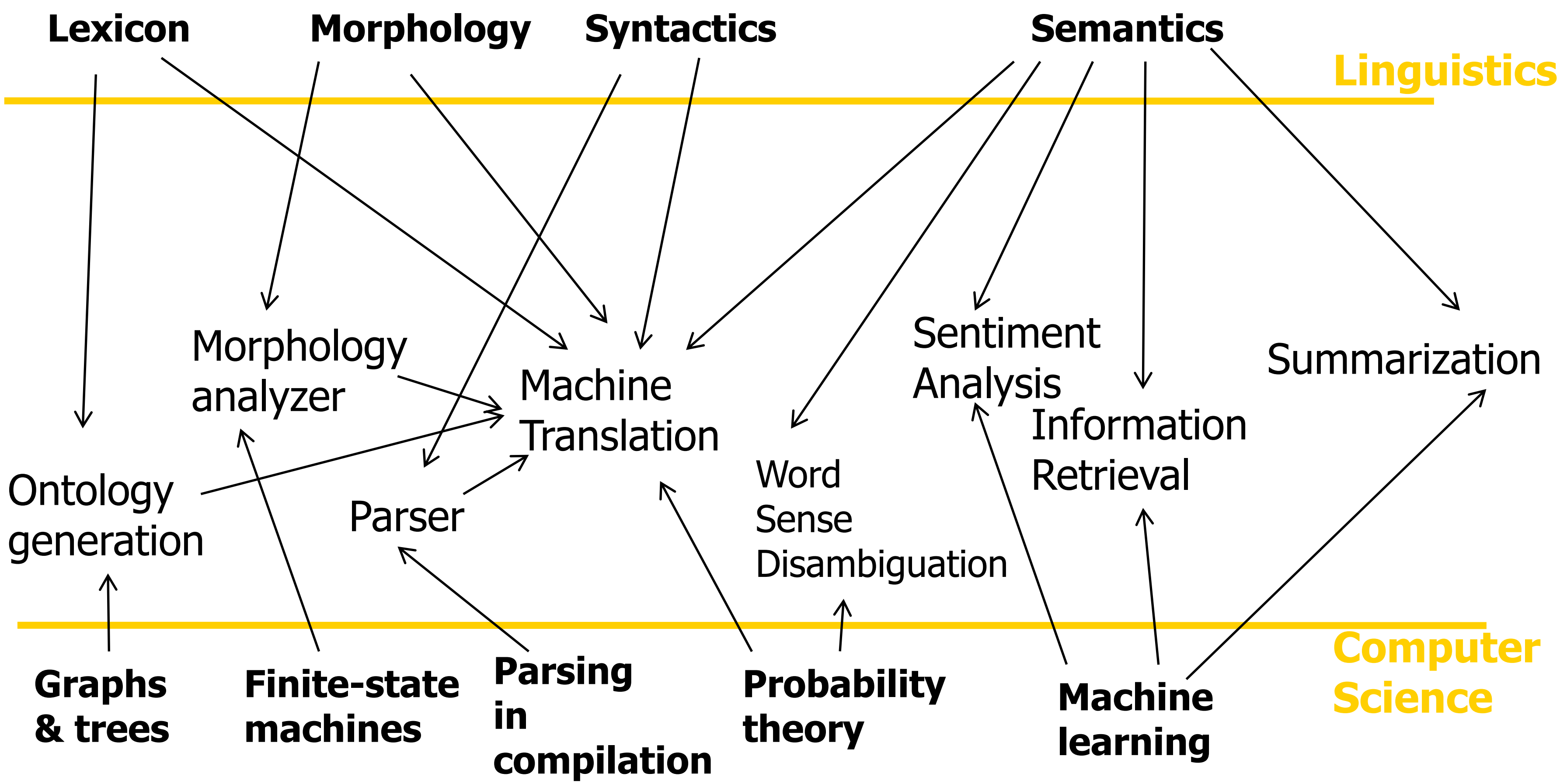

Homepage
Inbox (105,041) - pus
Inbox (18,664) - push
Resource Centre for I
Download Lectures: AI-NL
Welcome to AI-NLP-ML G
www.iitp.ac.in/~ai-nlp-ml/
Search
सत्यमेव जयते
AI-NLP-ML GROUP
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Home
People
Downloads
Publications
Events
Talks
Gallery
Contact Us
Custom Search
Search
Latest News
Current Notifications
ACM Transaction on Asian
One paper has been accepted in Knowledge-Based Systems, Elsevier,
Three papers have been accepted in NLDB-2017.
One paper has been accepted in CICLING-2017.
Shweta Yadav, Research
AI-NLP-ML Group
Department of CSE, IIT Patna
The Artificial Intelligence-Natural Language Processing-Machine Learning (AI-NLP-ML) group at Department of Computer Science and Engineering, IIT Patna has started its official journey in June, 2015. The group is dedicated to explore the frontiers of Artificial Intelligence, Machine Learning and Natural Language Processing under the able guidance of Prof. Pushpak Bhattacharyya. The group also consists of other two faculty members, Dr. Asif Ekbal and Dr. Sriparna Saha, and around 30 members including research scholars, research engineers, lexicographers, B.Tech & M.Tech students. Several industry sponsored projects are currently being undertaken.
Elsevier, the renowned scientific literature publishing company has set up the Elsevier Centre of Excellence for Natural Language Processing to
ENG
IN
3:23 PM
11-Jul-17

Nature of CL/NLP

AI Perspective (post-web)



NLP: At the confluence of linguistics & computer science



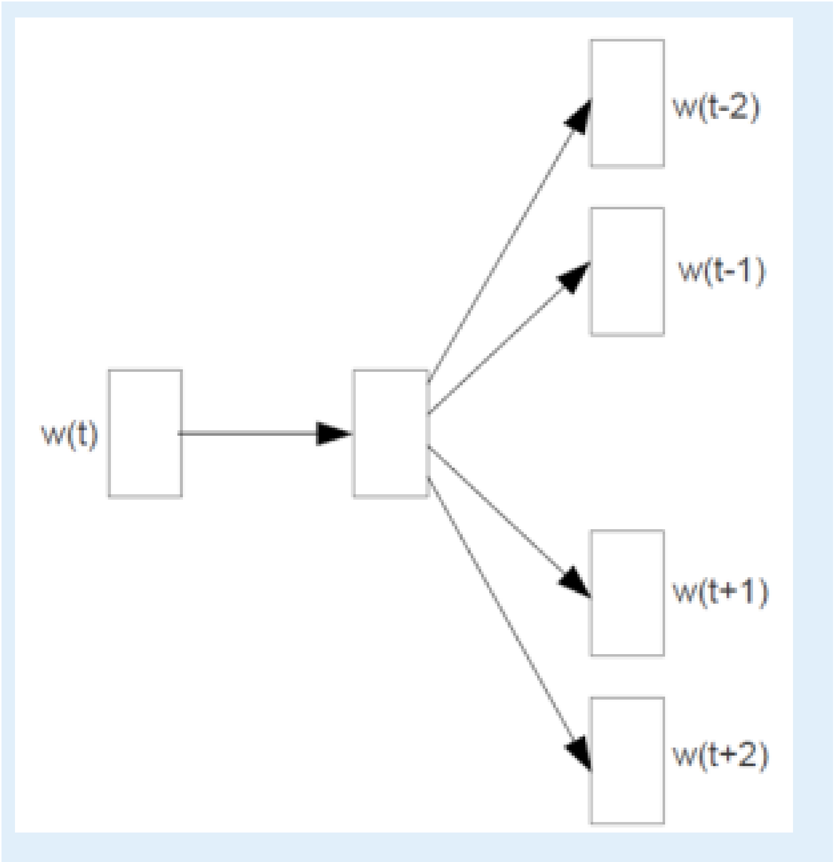
Linguistics is the Eye, Computation is the Body

"Linguistics is the eye": Harris Distributional Hypothesis

- Words with similar distributional properties have similar meanings. (Harris 1970)

- Model **differences** in meaning rather than the proper meaning itself

“Computation is the body”: Skip gram- predict context from word



CBOW:

Just reverse the Input-Ouput

Dog – Cat - Lamp

{bark, police, thief,
vigilance, faithful, friend,
animal, milk, carnivore)

{mew, comfort, mice, furry,
guttural, purr, carnivore, milk}

{candle, light, flash, stand, shade,
Halogen}

Test of representation

- **Similarity**
  - 'Dog' more similar to 'Cat' than 'Lamp', because
  - Input- vector('dog'), output- vectors of associated words
  - More similar to output from vector('cat') than from vector('lamp')

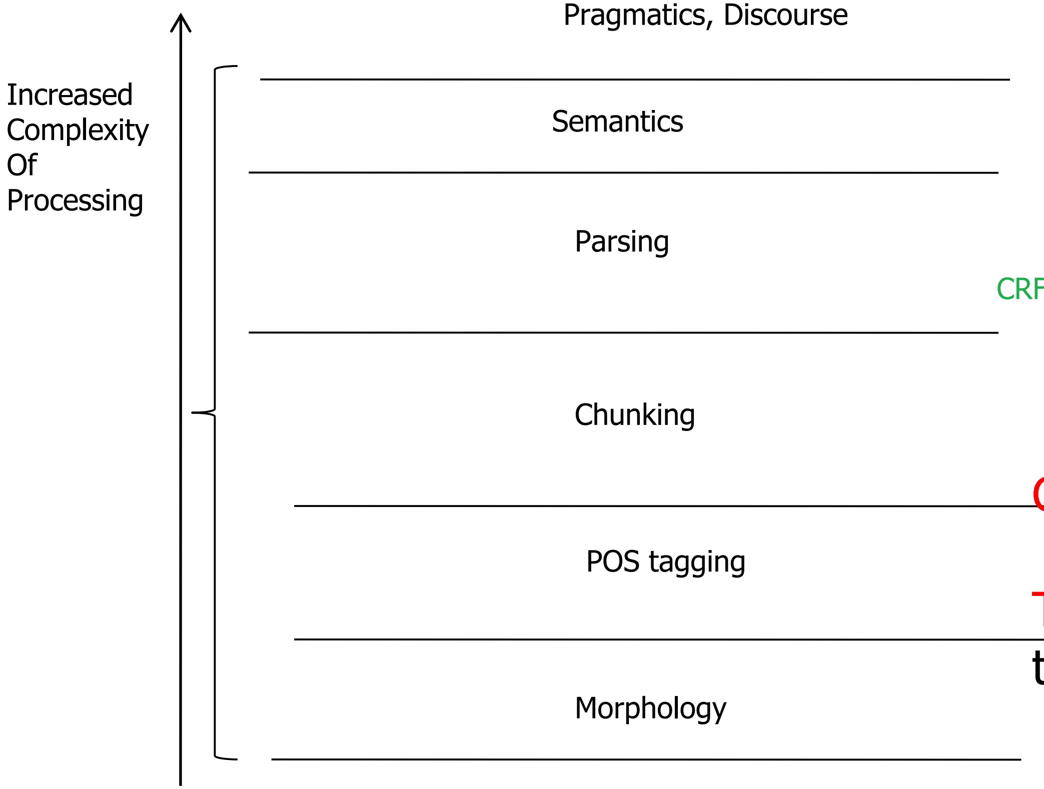
“Linguistics is the eye, Computation is the body”

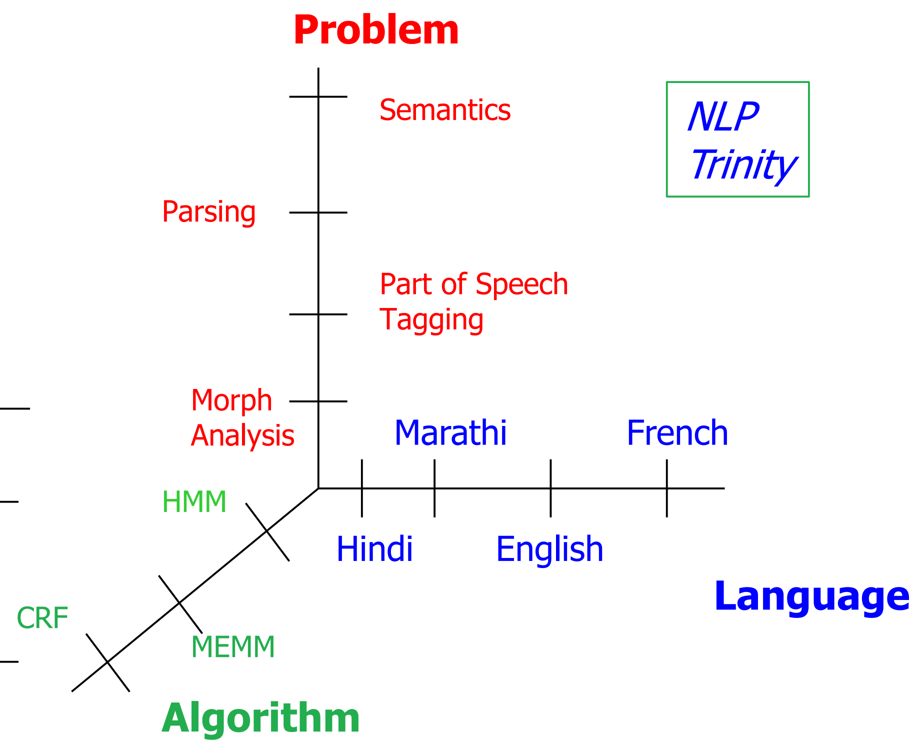
The encode-decoder deep learning network is nothing but

the ***implementation*** of

Harris’s Distributional Hypothesis

NLP: multilayered, Multi dimensional





GharaaSamorChyaaNe malaa sangitle

The one who is in front of the house told me

Need for NLP

- Humongous amount of language data in electronic form
- Unstructured data (like free flowing text) will grow to 40 zetabytes (1 zettabyte= 10^{21} bytes) by 2020.
- How to make sense of this huge data?
- Example-1: e-commerce companies need to know **sentiment** of online users, sifting through 1 lakh e-opinions per week: needs NLP
- Example-2: **Translation** industry to grow to $37 billion business by 2020

Machine Learning

- Automatically learning rules and concepts from data

Learning the concept of table.

What is “tableness”

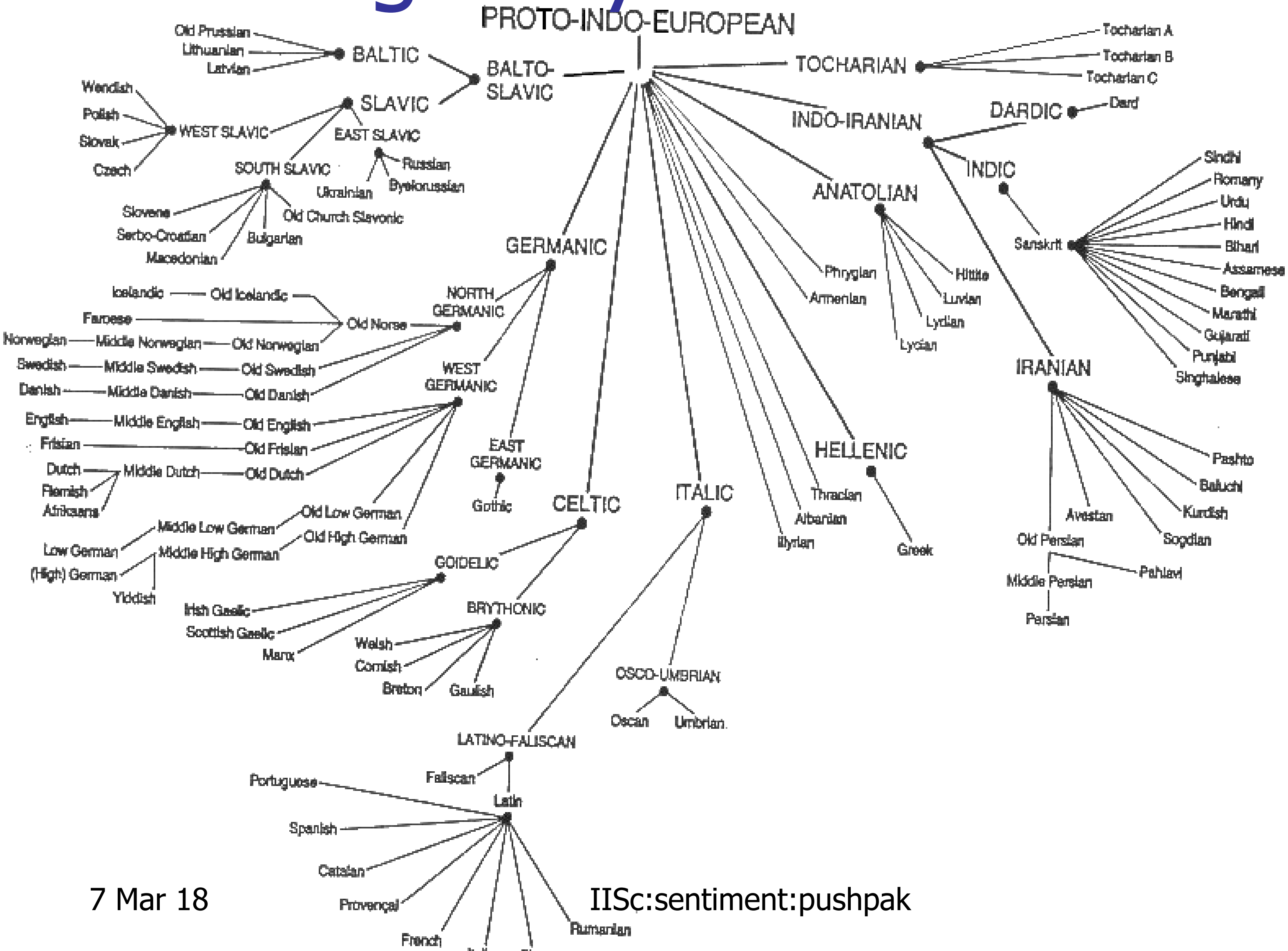
Rule: a flat surface with 4 legs (approx.: to be refined gradually)

NLP-ML marriage

NLP= Ambiguity Processing

- Lexical Ambiguity
  - *Present (Noun/Verb/Adjective; time/gift)*
- Structural Ambiguity
  - *1 and 2 bed room flats live in ready*
- Semantic Ambiguity
  - *Flying planes can be dangerous*
- Pragmatic Ambiguity
  - *I love being ignored* (after a party, while taking leave of the host)

Another challenge of NLP: multilinguality



Rules: when and when not

- When the phenomenon is understood AND expressed, rules are the way to go

- "Do not learn when you know!!"

- When the phenomenon "seems arbitrary" at the current state of knowledge, DATA is the only handle!
  - *Why do we say "Many Thanks" and not "Several Thanks"!*
  - *Impossible to give a rule*

Impact of probability: Language modeling

Probabilities computed in the context of corpora

1. P("The sun rises in the east")
2. P("The sun rise in the east")
   - Less probable because of grammatical mistake.
3. P(The svn rises in the east)
   - Less probable because of lexical mistake.
4. P(The sun rises in the west)
   - Less probable because of semantic mistake.

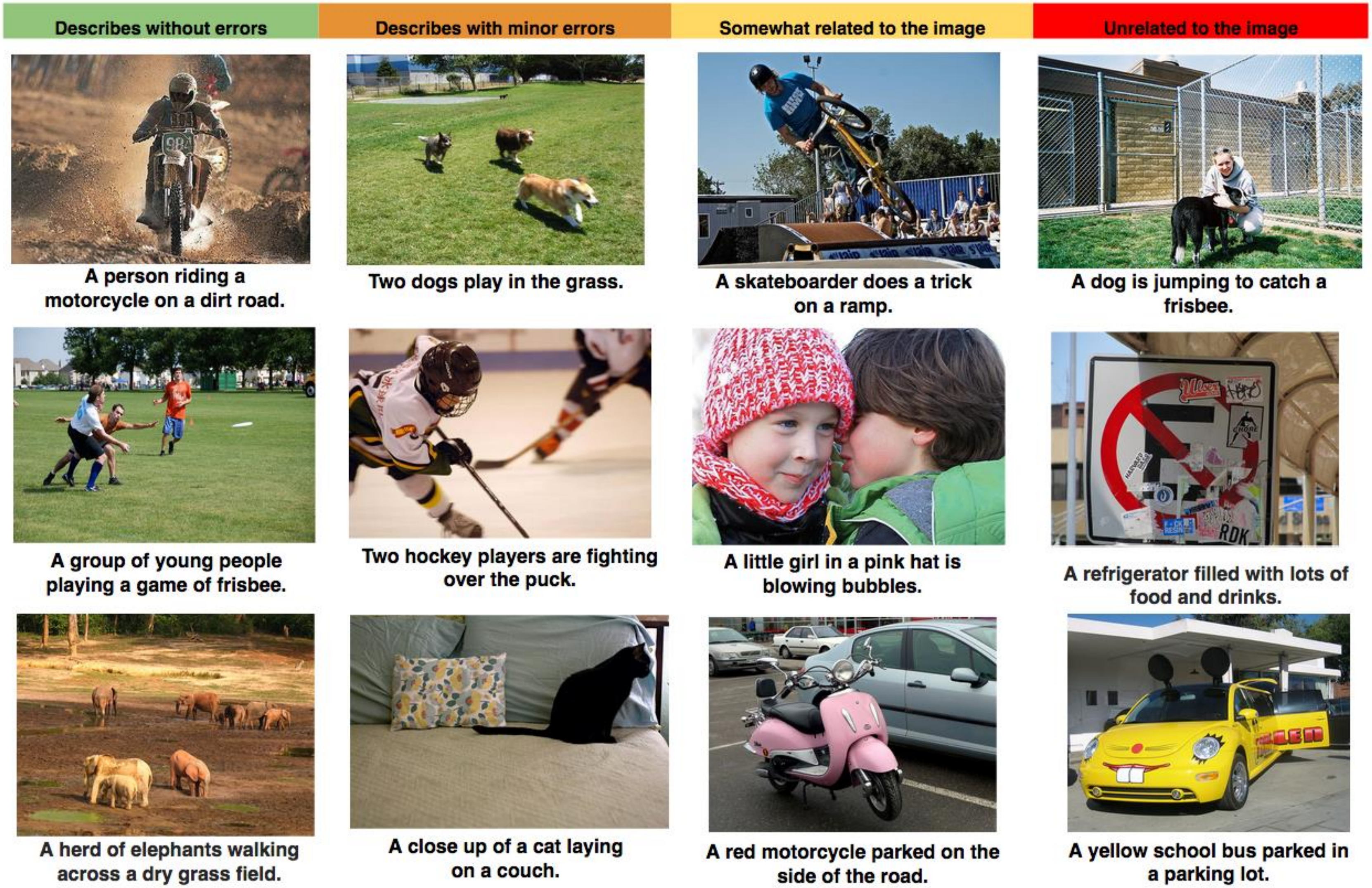
Probability Computation (quadrigram)

- ***P("sun rises in the east")= P(sun). P(rises|sun). P(in|sun, rises). P(the|sun, rises, in).P(east|rises, in, the)***
- ***P("sun rises in the east")= P(sun). P(rises|sun). P(in|sun, rises). P(the|sun, rises, in).P(east|rises, in, the)***
- ***#(rises, in, the , east) >> #(rises, in, the , east)*** in the **corpora**

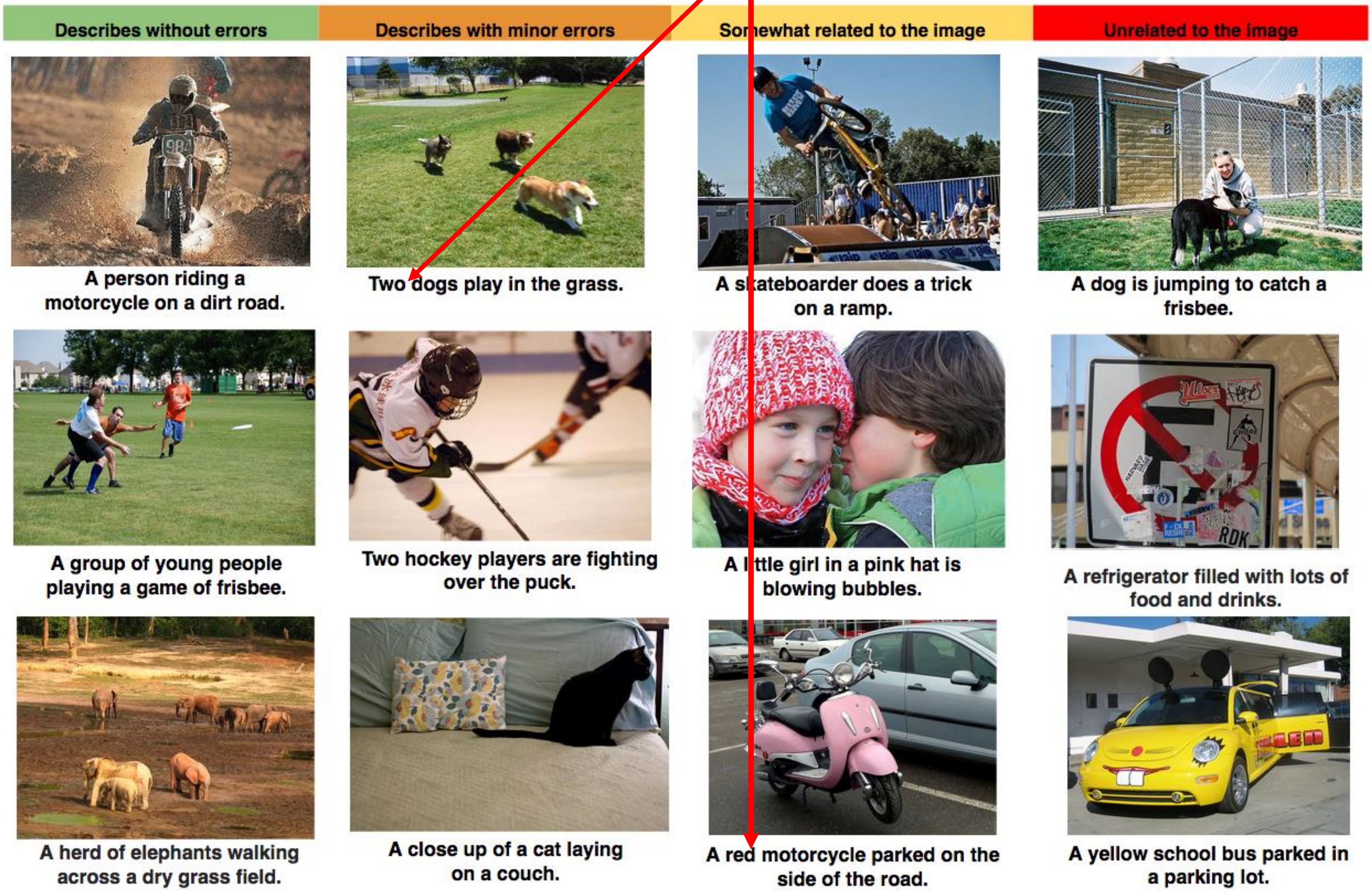
Power of Data- Automatic image labeling
(Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan, 2014)

Automatically captioned: "Two pizzas sitting on top of a stove top oven"

Automatic image labeling (cntd)



Shallow Understanding



Main methodology

- Object A: extract parts and features
- Object B which is in correspondence with A: extract parts and features
- LEARN mappings of these features and parts
- Use in NEW situations: called DECODING

New age NLP-ML-AI

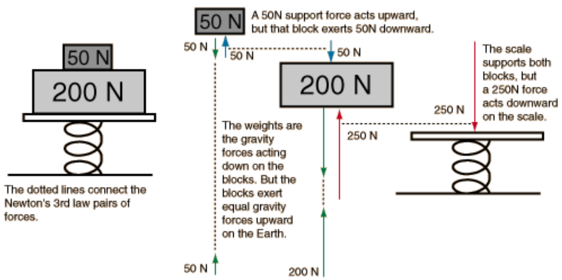
- **Deep Understanding=**

  Shallow Understanding

  +

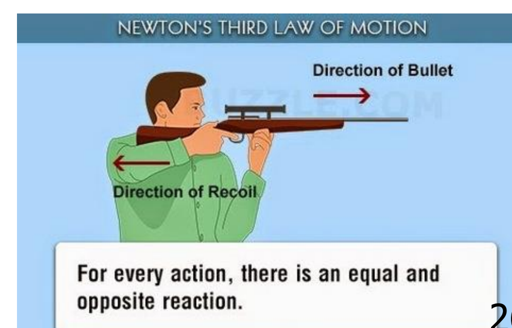
  Big Data

Grind methodology: Show umpteen number of problems



Newton's 3rd law

Subject to solving huge number of problems!!



Pattern driven learning

- Memorise the patterns
- MCQ
- Match pattern
- Eliminate choices
- Select from a few

Classification vs. Learning Distribution

- "I love being ignored" (after a party to the host)
  - Sarcastic- Yes, non-sarcastic- No
    - ***HARDMAX***
- S- "This movie is great for putting you to sleep"
  - P("sarcastic"|S)- 0.9; P("non-sarcastic"|S)- 0.1
    - ***SOFTMAX***

Example of new age NLP: MT

- Data playing a key role in machine translation
- Unexpected developments!
- For example, machine translation
  - Who could imagine that a machine with LEARN to translate from parallel corpora?

Word alignment is the crux of the matter

English	French
(1) three rabbits	(1) trois lapins
a b	w x
(2) rabbits of Grenoble	(2) lapins de Grenoble
b c d	x y z

Initial Probabilities:
each cell denotes *t(a ←→ w), t(a ←→ x) etc.*

	a	b	c	d
w	1/4	1/4	1/4	1/4
x	1/4	1/4	1/4	1/4
y	1/4	1/4	1/4	1/4
z	1/4	1/4	1/4	1/4

“counts”

a b* ←→ *w x	a	b	c	d
w	1/2	1/2	0	0
x	1/2	1/2	0	0
y	0	0	0	0
z	0	0	0	0

b c d* ←→ *x y z	a	b	c	d
w	0	0	0	0
x	0	1/3	1/3	1/3
y	0	1/3	1/3	1/3
z	0	1/3	1/3	1/3

Revised probabilities table

	a	b	c	d
w	1/2	1/4	0	0
x	1/2	5/12	1/3	1/3
y	0	1/6	1/3	1/3
z	0	1/6	1/3	1/3

“revised counts”

a b ←→ ***w x***	a	b	c	d
w	1/2	3/8	0	0
x	1/2	5/8	0	0
y	0	0	0	0
z	0	0	0	0

b c d ←→ ***x y z***	a	b	c	d
w	0	0	0	0
x	0	5/9	1/3	1/3
y	0	2/9	1/3	1/3
z	0	2/9	1/3	1/3

Re-Revised probabilities table

	a	b	c	d
w	1/2	3/16	0	0
x	1/2	**85/144**	1/3	1/3
y	0	1/9	1/3	1/3
z	0	1/9	1/3	1/3

Continue until convergence; notice that (b,x) binding gets progressively stronger; b=rabbits, x=lapins

Sentiment Analysis

Definition (Liu 2010)

(Liu, 2010) defines a sentiment or opinion as a quintuple-

$$< o_j , f_{jk} , so_{ijkl} , h_i , t_l >,$$

where

o_j *is a target object,*

f_{jk} *is a feature of the object* o_j,

so_{ijkl} *is the sentiment value of the opinion*

of the opinion holder h_i

on feature f_{jk}

of object o_j

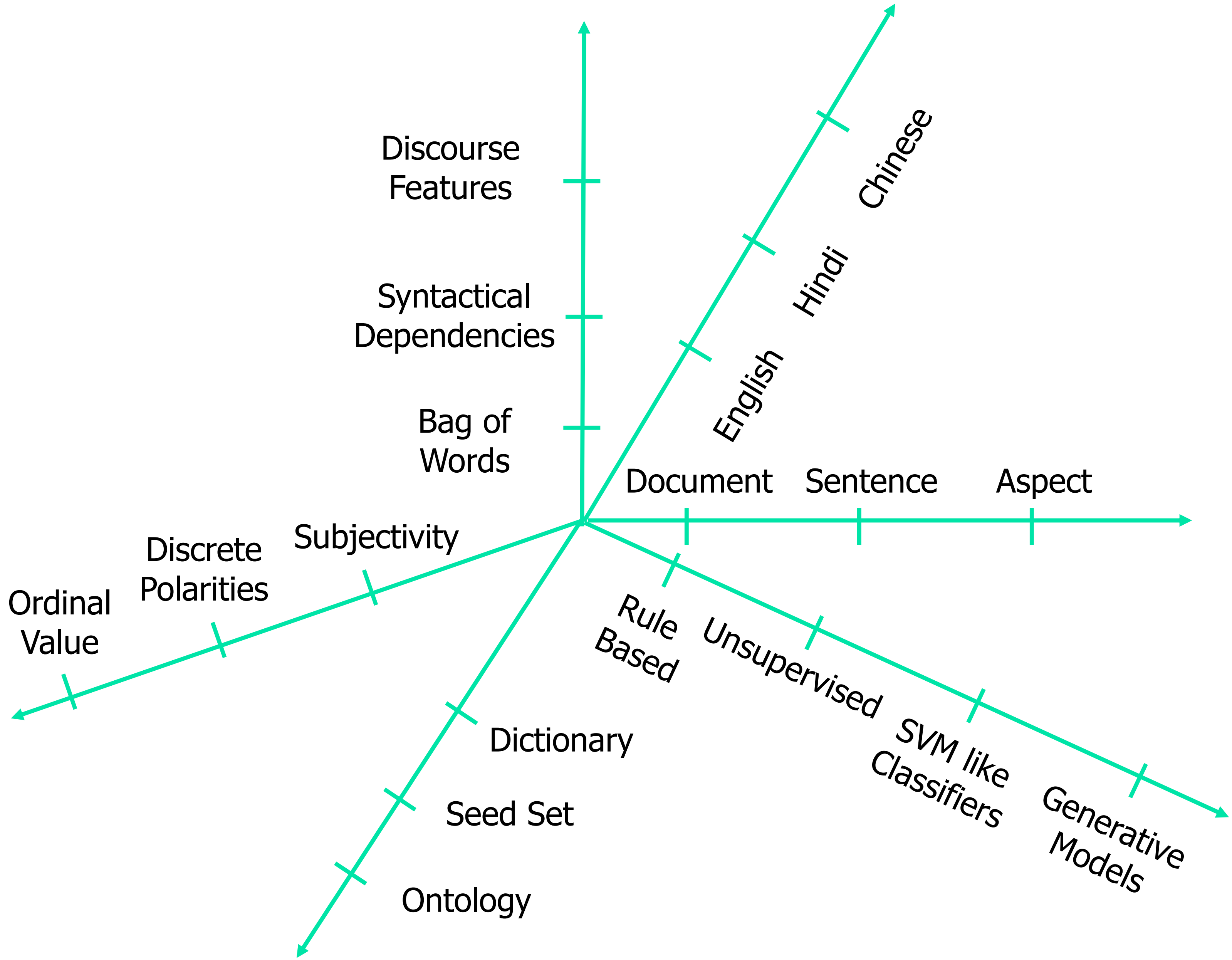
at time t_l

Example

- “I love the songs in the movie, though only the cast was liked by my brother who said the director was of the opinion that the story line which is from a novel by Shakespeare will be lapped up by the public”

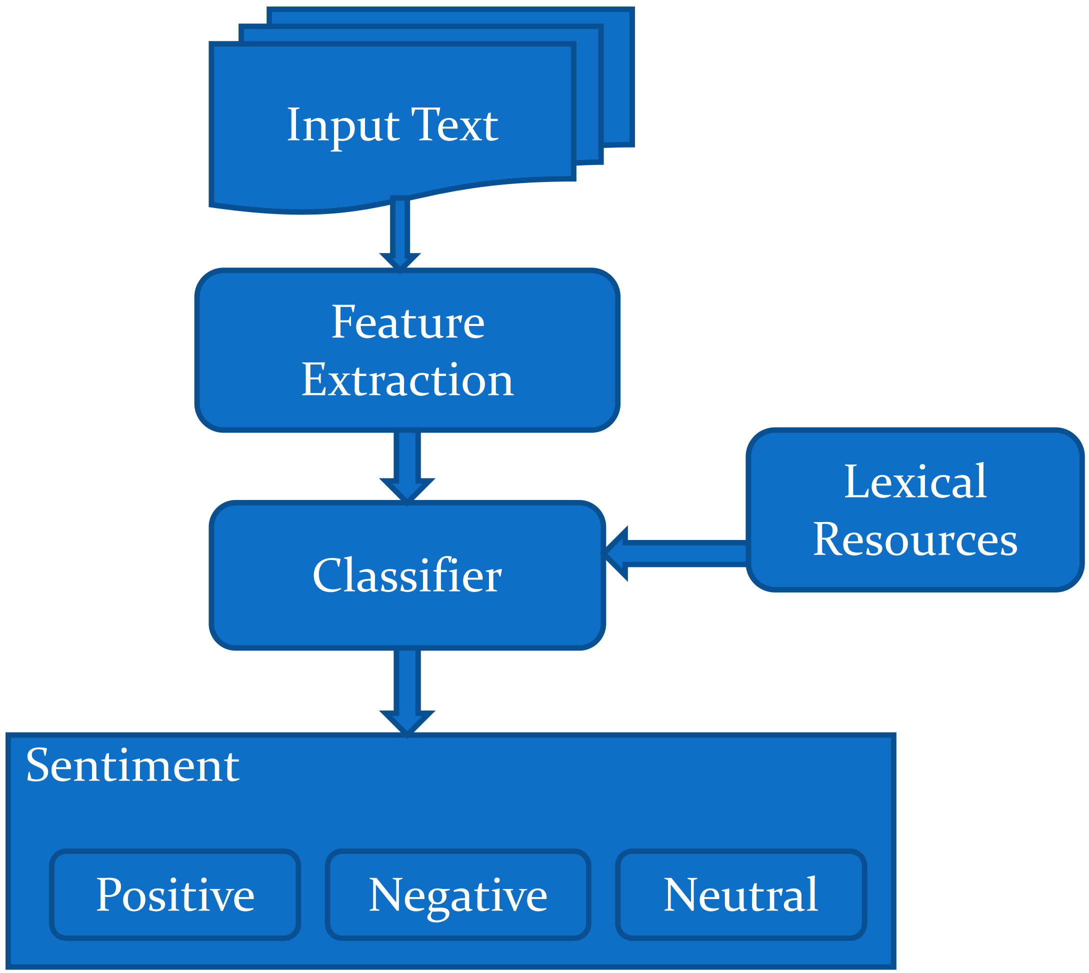
Example (cntd.)

- Entity: *movie*
- Aspects: *songs, cast, story line*
- Opinion holder: *I, brother, director, public* (not *Shakespeare!!*)
- Time: *present (I), past (brother), present (director), future (public)*
- Opinioner-sentiment-aspect: *I-love-song, brother-like-cast, director-like-story_line (indirectly), public-lap_up-story_line*



Dimensions of Sentiment Analysis

Block diagram



Challenges

`I suggest you wear your perfume with windows and doors shut! #sarcasm'

Sarcasm

`... falls 284 runs short of what would have been a fourth first-class triple-century'.

www.cricinfo.co

Implicit knowldege

`The movie may have the nicest actors, a talented music director of worldwide acclaim and the most expensive set one has ever seen but it fails to impress'.

Thwarting

`keeps you on the edge of your seat'

`Tim Tam. \m/'

Nature of text

`He is a deadly football player'

`You may have deadly snakes at the camp site at night`

Balamurali et al [2011

Domain specificity

Representative figures for SA Accuracy

Features	# of features	Frequency or Presence?	NB	ME	SVM
Unigrams	16165	Freq.	**78.7**	N/A	72.8
Unigrams	16165	Pres.	81.0	80.4	**82.9**
Unigrams+bigrams	32330	Pres.	80.6	80.8	**82.7**
Bigrams	16165	Pres.	77.3	**77.4**	77.1
Unigrams+POS	16695	Pres.	81.5	80.4	**81.9**
Adjectives	2633	Pres.	77.0	**77.7**	75.1
Top 2633 unigrams	2633	Pres.	80.3	81.0	**81.4**
Unigrams+position	22430	Pres.	81.0	80.1	**81.6**

Sarcasm

Etymology

- Greek: '*sarkasmós*': ‘to tear flesh with teeth’
- Sanskrit: '*vakrokti*': ‘a twisted (vakra) utterance (ukti)’

Definition- Foundation is *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.	Sarcasm is embedded in the meaning of words being used.	'Like/As if' are common prefixes to ask rhetorical questions.	Non-speech acts (body language, gestures) contributing to the sarcasm
'This looks like a perfect plan!'	*'I love being ignored'*	*'Like you care'*	*'(shrugs shoulders) Very helpful indeed!'*

Impact 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 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 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 (Based on Riloff et al	
Implicit Sentiment Phrases	Boolean feature indicating phrases extracted from the implicit phrase extraction step
Explicit Incongruity (Based on Ramteke et al	
#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

Incongruity and embeddings

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

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 examples

- waiting 45 min for the subway in the freezingcold is so much fun.
- well 3 hrs of sleep this is awesome.
- gotta read 50 pages and do my math before tomorrow i'm so excited.
- -28 c with the windchill fantastic 2 weeks.
- woooo when you're up to 12:30 finishing you're english paper.

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.

Example

“This phone has an awesome battery back-up of 2 hours”,

```
(S
  This/DT
  (NP (NBAR phone/NN))
  has/VBZ
  an/DT
  (NP (NBAR awesome/JJ battery/NN backup/NN))
  of/IN
  2/CD
  (NP (NBAR hours/NNS)))
```

Example (cntd.)

- Noun Phrases:

  ['phone', 'awesome', 'battery', 'backup', 'hours']

- Addition to sarcastic repository:

  (Tweet No., ['phone', 'awesome', 'battery', 'backup', 'hours'], 2, 'hours')

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', 'VBG', 'VBN', 'VBP', 'VBZ'.

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)

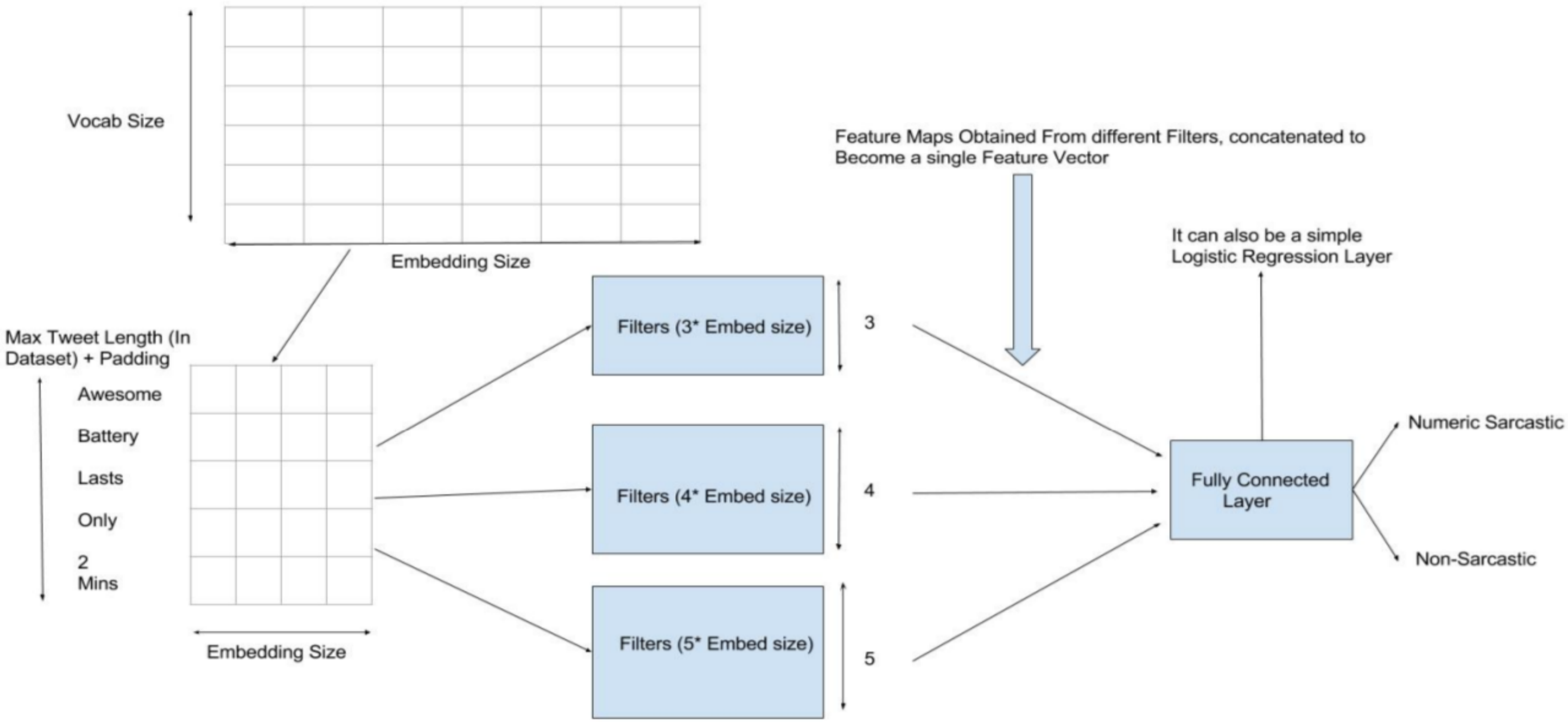
Approaches	Precision			Recall			F-score		
	P(1)	P(0)	P(avg)	R(1)	R(0)	R(avg)	F(1)	F(0)	F(avg)
Past Approaches									
Buschmeier et.al.	0.19	0.98	0.84	0.99	0.07	0.24	0.32	0.13	0.16
Liebrecht et.al.	0.19	1.00	0.85	1.00	0.07	0.24	0.32	0.13	0.17
Gonzalez et.al.	0.19	0.96	0.83	0.99	0.06	0.23	0.32	0.12	0.15
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

Insight

- Ad hocism in the decision
  - for sarcasic/non-sarcastic (9 close to 3, 81 not close to 21 etc.)
- We rely on the data to give us the decision threshold.
- SVM, KNN etc.- human intervention is in the form of features.
- Even this level of human intervention is removed by resorting to Deep Learning (accuracy goes to ~90%).

Message

- Rule based systems are great for intuition building and explainability.

- However, some human decisions seem ad hoc. So relegate that decision to come from data.

- In the final step resort to DL to have even feature engineering from data.

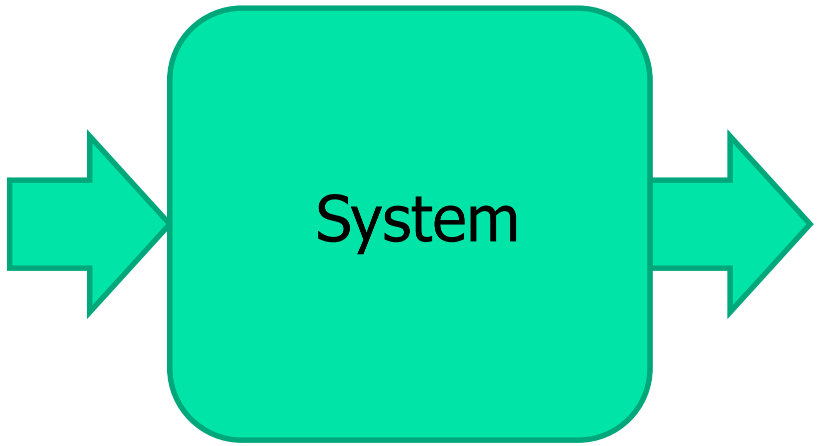
Thwarting

Ankit Ramteke, Akshat Malu, Pushpak Bhattacharyya and Saketha Nath, *Detecting Turnarounds in Sentiment Analysis: Thwarting*, **ACL 2013**, Sofia, Bulgaria, 4-9 August, 2013

Problem definition

- To detect Thwarting in text

Text Document



Thwarted/ Not Thwarted

Thwarted

The actors performed well. The music was enthralling. The direction was good. But, I still did not like the movie.

Not Thwarted

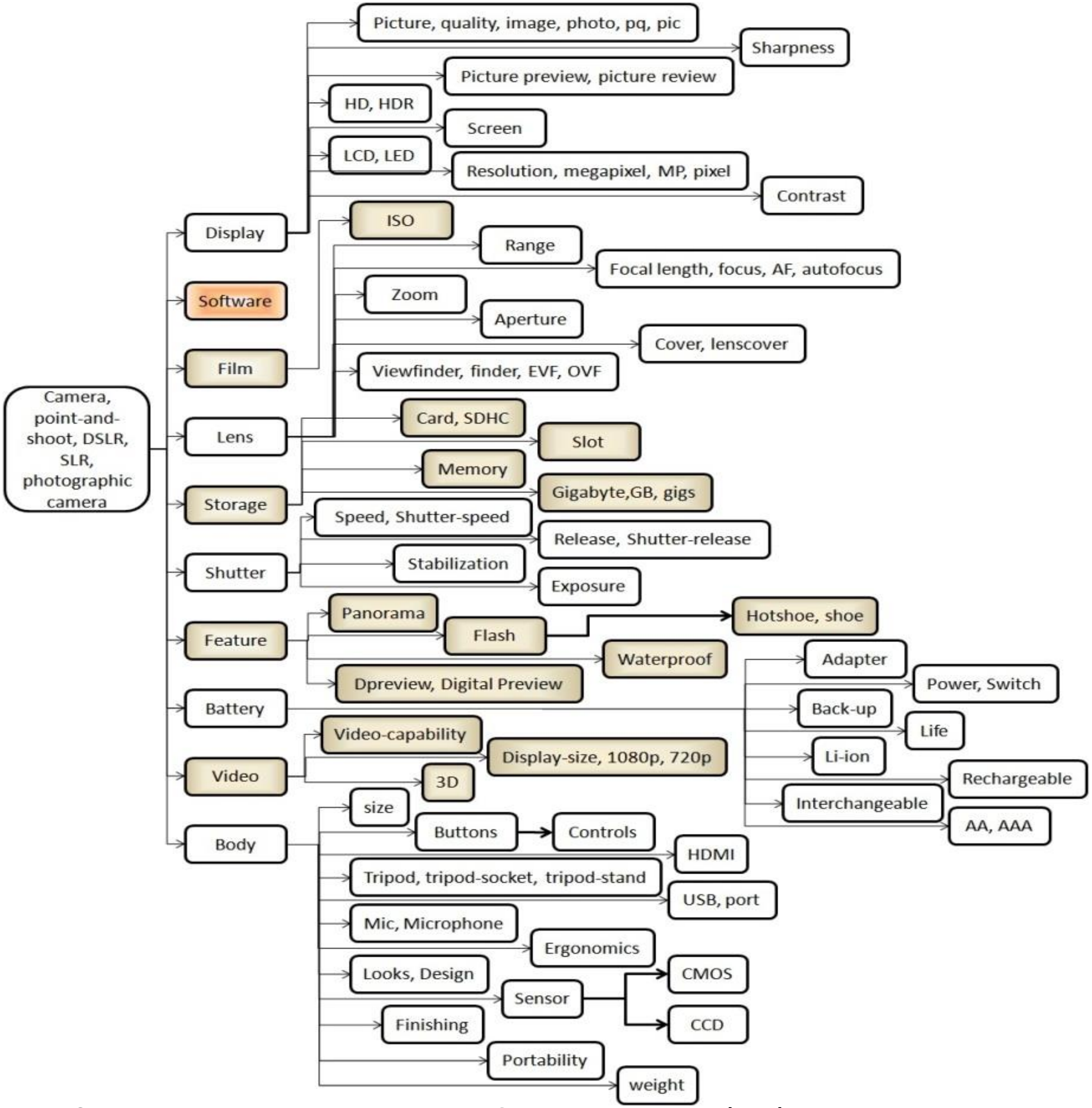
This camera has everything that you need. A Superb lens, an amazing picture quality and a long battery life. I love it.

Definition of thwarting

- **Thwarting:** Minority of a document's content determines its polarity.
- Thwarting is a rare phenomenon and thus faces data skew
- Approaches to handling data skew in other tasks
  - Tao *et al.* (2006)
  - Hido *et al.* (2008)
  - Provost *et al.* (1999)
  - Viola *et al.* (2001)

Domain Ontology

- Need for a weighting of entities related to a domain
- **Domain Ontology**: Aspects (entity parts) arranged in the form of a hierarchy
- An ontology naturally gives such weighting
  - Each level has a weight



Camera Ontology

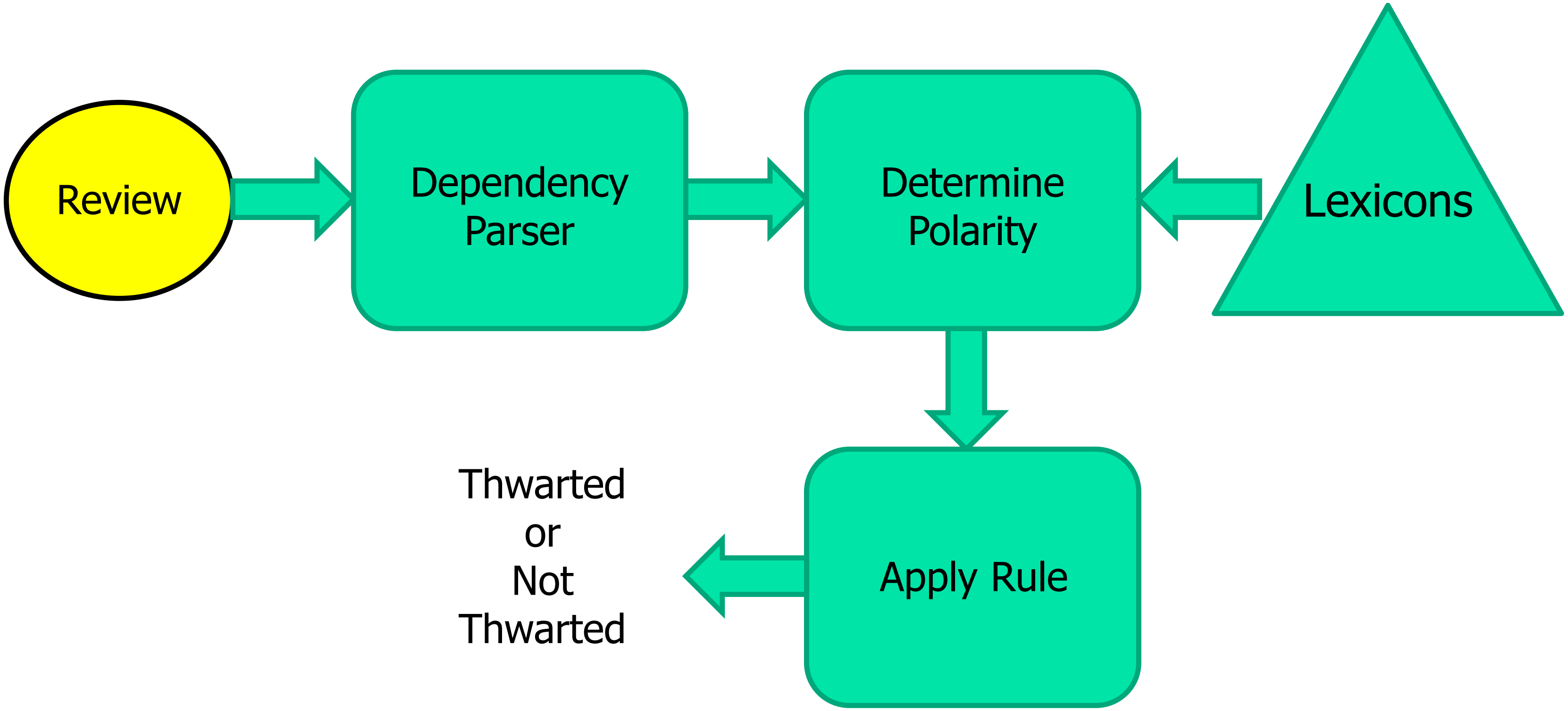
Basic idea

From the perspective of the domain ontology, the sentiment towards the overall product or towards some critical feature mentioned near the root of the ontology should be opposite to the sentiment towards features near the leaves.

An Example

"I love the sleek design. The lens is impressive. The pictures look good but, somehow this camera disappoints me. I do not recommend it."

Process flow

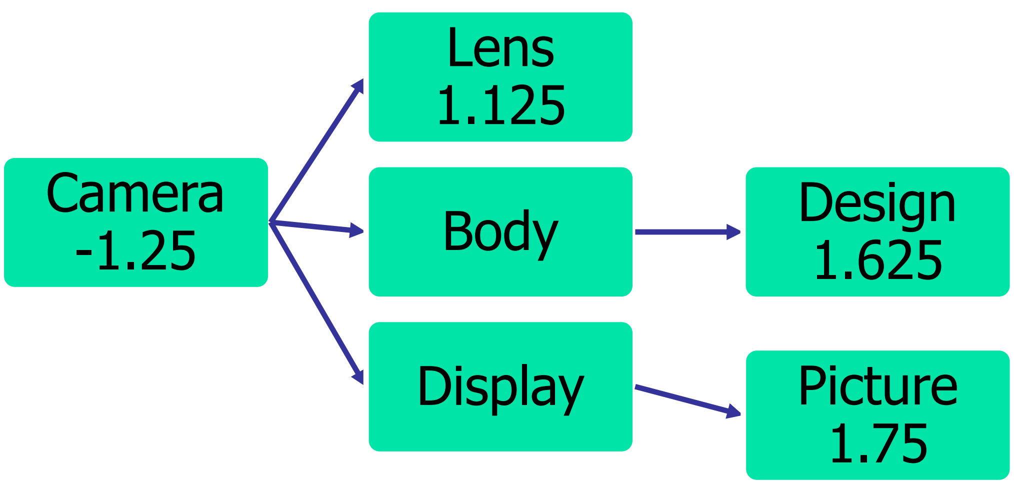


Dependency, weighting, decision

dobj(love-2, design-5)
nsubj(impressive-4, lens-2)
nsubj(look-3, pictures-2)
acomp(look-3, good-4)
nsubj(disappoints-10, camera-9)

Weights from:

SentiWordNet (Esuli et al., 2006), **Taboada** (Taboada et al., 2004), **BL lexicon** (Hu et al., 2004) and **Inquirer** (Stone et al., 1966).



Thwarted!!

AUC accuracy of the Rule based approach: **53%**

Need more principled approach to find weights

- Different Weight for nodes on the same level
  - Body and Video Capability
    - Individual tastes, not so critical
  - Lens or the Battery
    - More critical feature
- Learn Weights from corpus

ML Approach

Extract Weights

- Domain aspects: $A_1, A_2 \ \dots \ A_N$
- Weights: $W_1, W_2 \ \dots \ W_N$
- Overall polarity $P = \sum_i A_i * \ W_i$

- Minimize Hinge loss: $\max(0,1 - P.W^T.A)$

Modify weights by percolation

- Percolate polarity of child to parent
  - Complete Percolation
    - $\text{polarity}_{\text{parent}}$ = sum of polarities of children
  - Controlled Percolation

$$P_{camera} = p_{camera} + \frac{p_{lens}}{2} + \frac{p_{body}}{2} + \frac{p_{display}}{2} + \frac{p_{design}}{4} + \frac{p_{picture}}{4}$$

Representing Reviews

Extract a vector of values

$$V_1, V_2 \ \dots \ V_M$$

from each review.

Each V_i represents a weighted aspect polarity value.

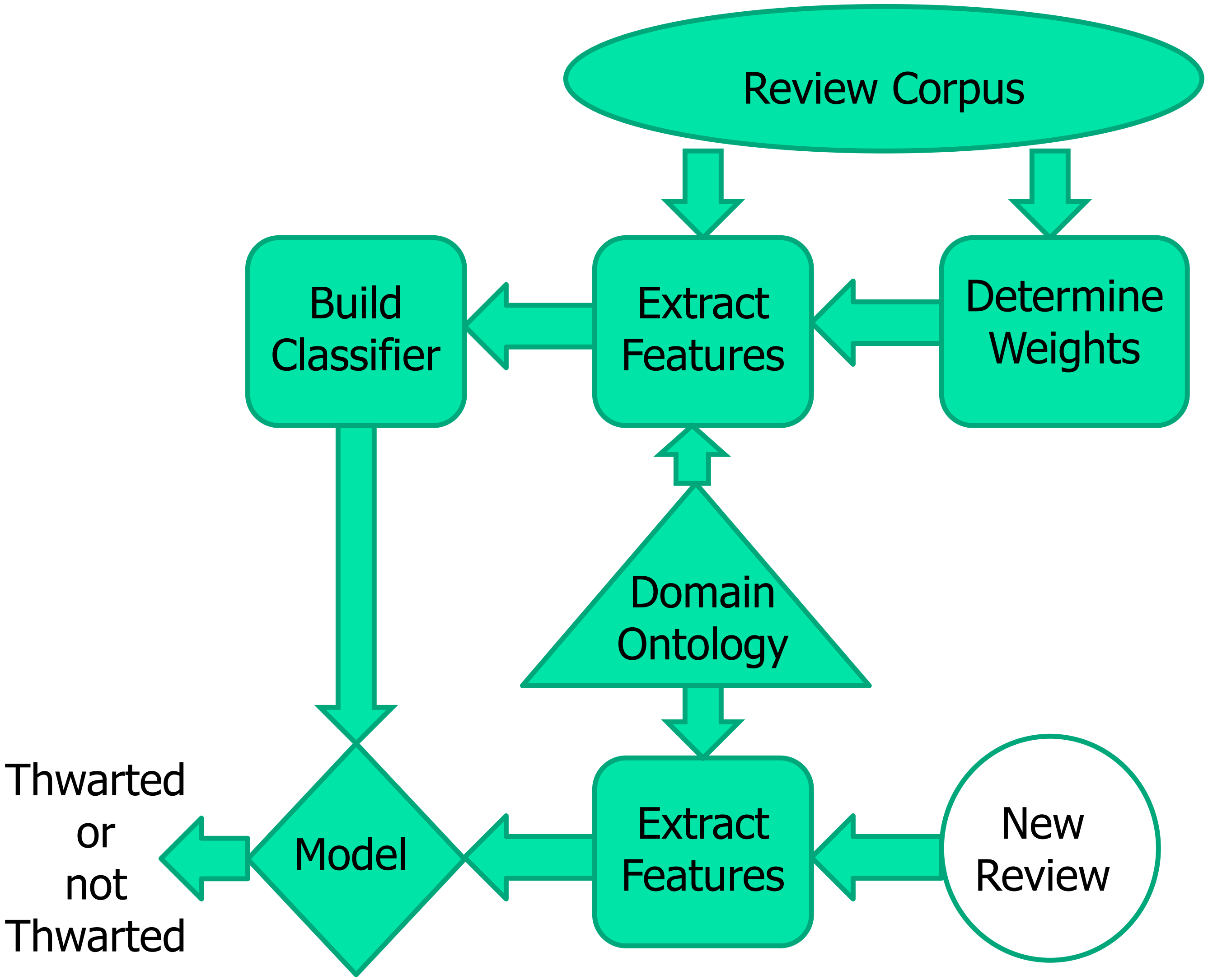
Features (1/2)

- Document polarity
- Number of flips of sign (i.e. from positive to negative and vice versa) normalized by the number of terms in the sequence
- The Maximum and the Minimum values in a sequence
- The length of the longest positive contiguous subsequence
- The length of the longest negative contiguous subsequence
- The mean of the values

Features (2/2)

- Total number of positive values in the sequence
- Total number of negative values in the sequence
- The first and the last value in the sequence
- The variance of the moving averages
- The difference in the averages of the longest positive and longest negative contiguous subsequences

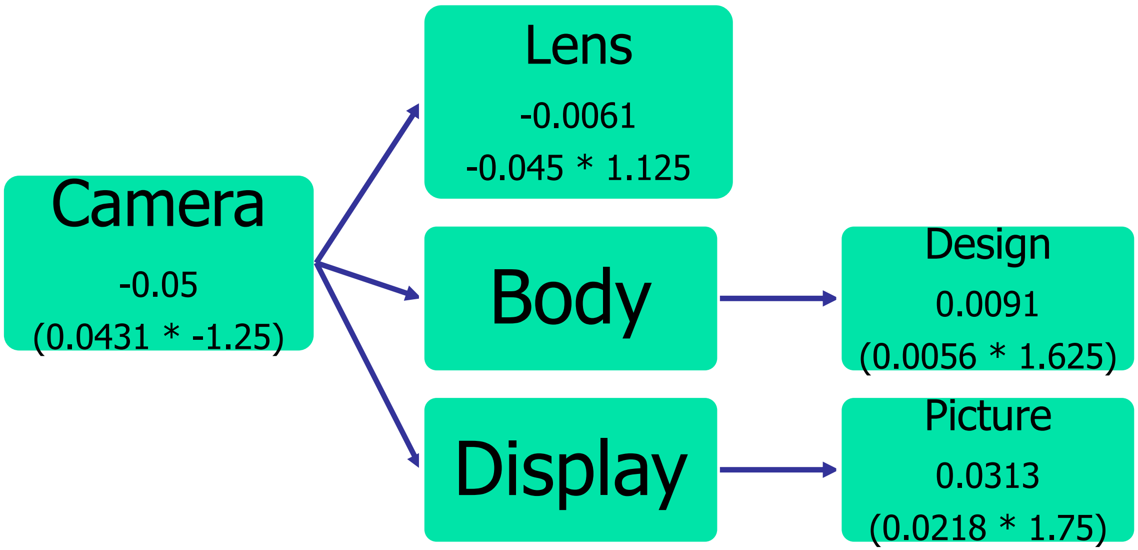
Process flow



Running example

"I love the sleek design. The lens is impressive. The pictures look good but, somehow this camera disappoints me. I do not recommend it."

“Tree” from the example



Features in the example

Feature	Value
Document Polarity	-1
Number of flips of sign	3
The Maximum value in a sequence	0.031325
The Minimum value in a sequence	-0.05
The length of the longest positive contiguous subsequence	1
The length of the longest negative contiguous subsequence	1
The mean of the values	0.003940625
Total number of positive values in the sequence	2
Total number of negative values in the sequence	2
The first value in the sequence	0.0091
The last value in the sequence	-0.05
The variance of the moving averages	0
The difference in the averages of LPCS and LNCS	0.081325

Experiments

- Setup:
  - Dataset by Malu (2012)
  - We crawled[1] an additional 1000 reviews out of which 24 reviews were Thwarted
  - Camera domain
  - 2198 reviews 60 thwarted
  - Ontology for domain specific features
  - Data is skewed so weighing of classes employed
- Inter annotator Agreement
- Classification experiments
  - 10 fold cross validation
- Ablation Test

Reviews crawled from www.epinions.com

Results: Inter annotator Agreement

- Cohen’s kappa : 0.7317
- Agreement of 70% for the thwarted class
- Agreement of 98% for the non-thwarted
- Identifying thwarting is difficult even for humans

Results: Classification - 1

	Loss Type	
Percolation Type	Linear	Hinge
No percolation	**68.9**	65.6
Controlled	66.89	62.39
Complete	67.65	63.43

Table 5.2: Results for non negative weights with prior

	Loss Type	
Percolation Type	Linear	Hinge
No percolation	**69.01**	67.42
Controlled	65.09	62.16
Complete	62.77	60.94

Table 5.3: Results for non negative weights without prior

Results: Classification - 2

	Loss Type	
Percolation Type	Linear	Hinge
No percolation	73.87	70.12
Controlled	**81.05**	77.17
Complete	63.85	60.94

Table 5.4: Results for unconstrained weights without prior

	Loss Type	
Percolation Type	Linear	Hinge
No percolation	73.99	70.56
Controlled	**78.47**	72.03
Complete	62.88	61.36

Table 5.5: Results for unconstrained weights with prior

Results: Ablation Test

Feature Removed	Loss in AUC
Document Polarity	10.01%
Number of flips of sign	2.13%
The Maximum value in a sequence	1.24%
The Minimum value in a sequence	1.0%
The length of the longest positive contiguous subsequence	1.2%
The length of the longest negative contiguous subsequence	0.9%
The mean of the values	2.0%
Total number of positive values in the sequence	1.2%
Total number of negative values in the sequence	1.0%
The first value in the sequence	0.5%
The last value in the sequence	1.1%
The variance of the moving averages	5.0%
The difference in the averages of LPCS and LNCS	3.0%

Observations and insights

- **Ontology** guides a rule based approach to thwarting detection, and also provides difference-making features for SVM based learning systems

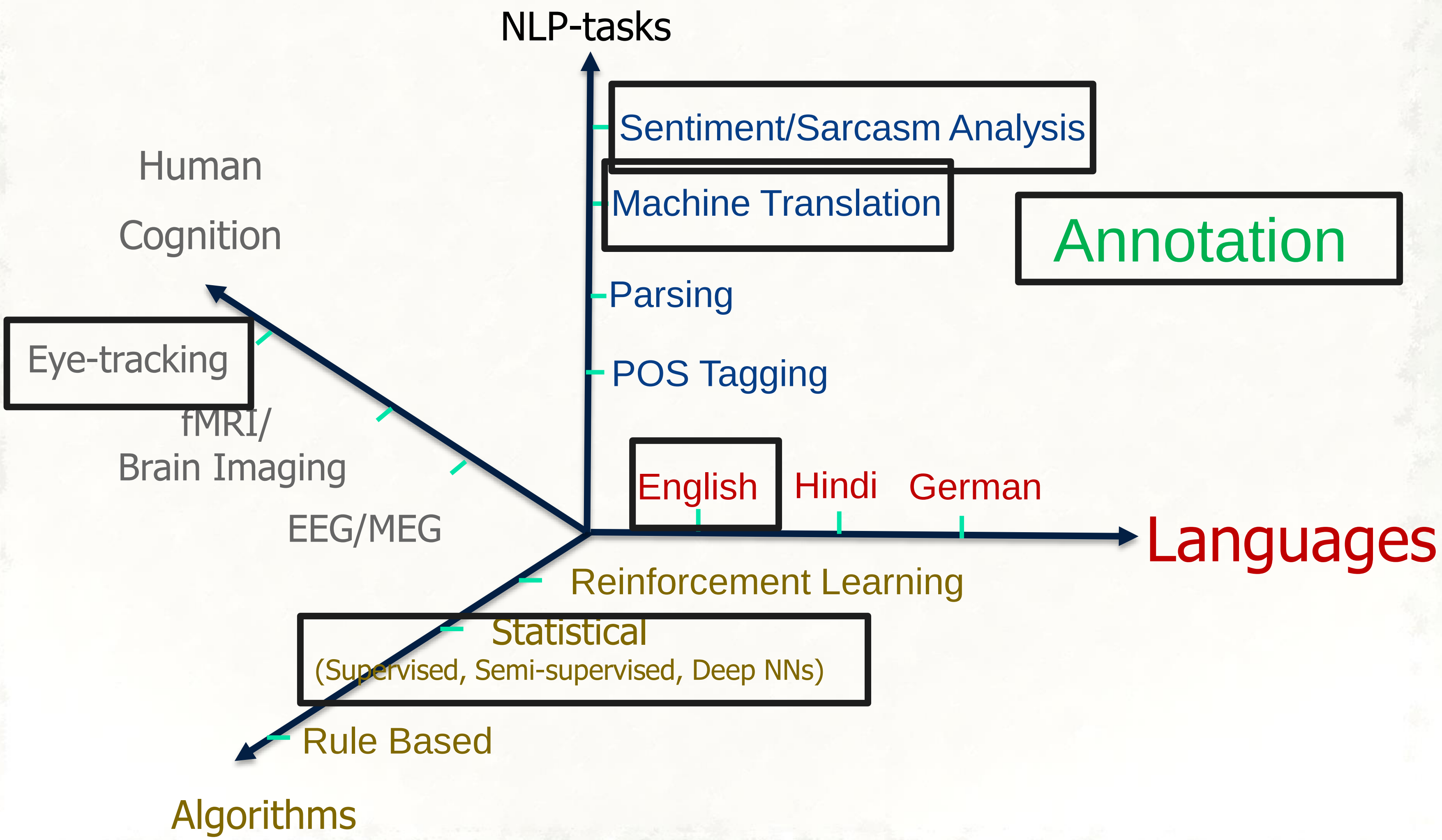
- Percolating polarities is needed

- ML scores over the rule based system by 25%

back

Enter cognition

NLP-trinity



NLP-tasks
Sentiment/Sarcasm Analysis
Machine Translation
Parsing
POS Tagging
Annotation
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

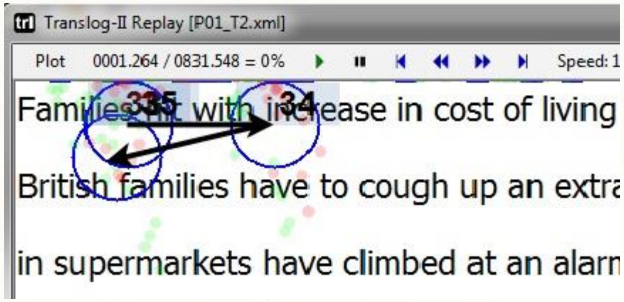
(image - sources: http://www.tobii.com/)

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

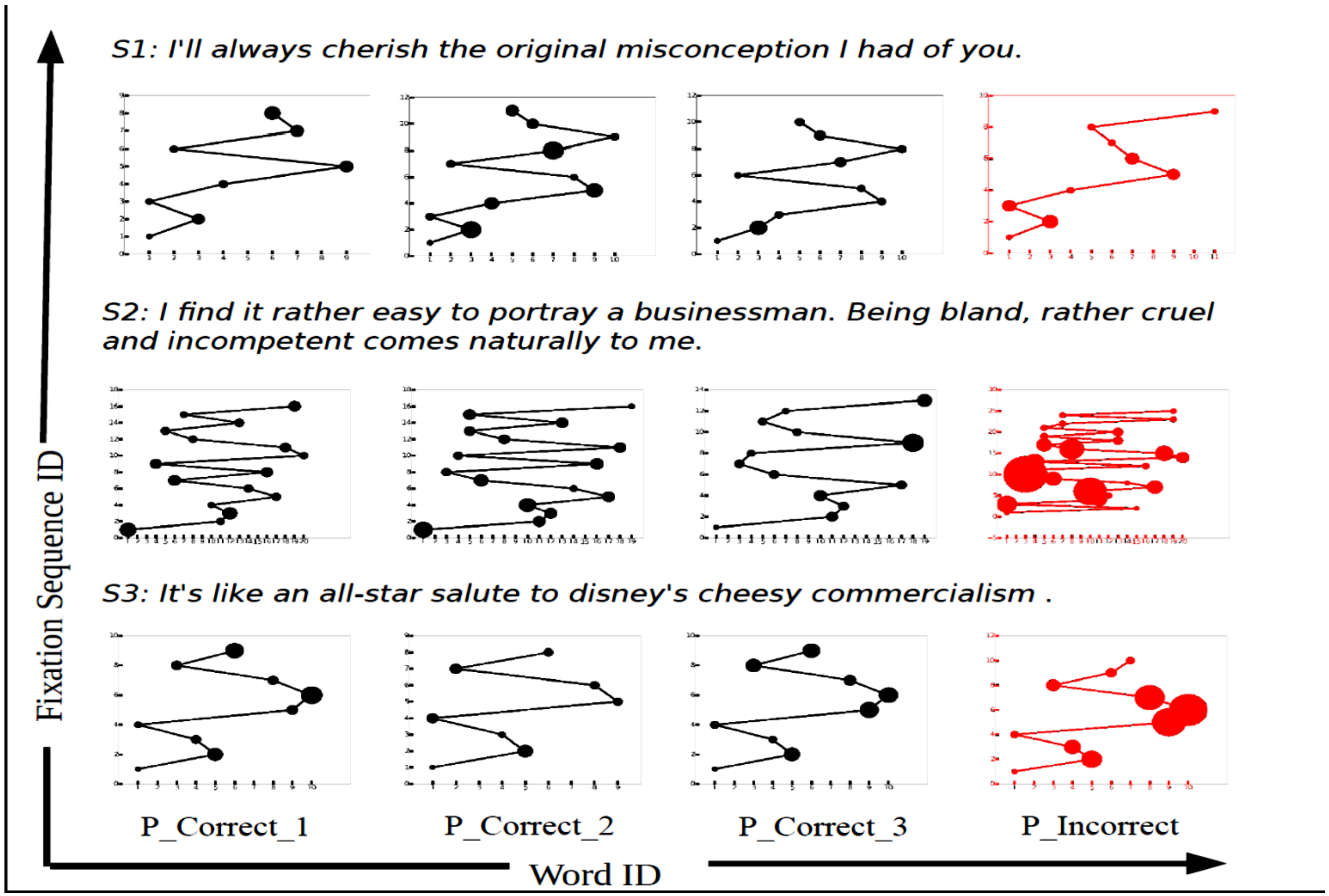
- Eye-Mind Hypothesis (Just and Carpenter, 1980)

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

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)

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**

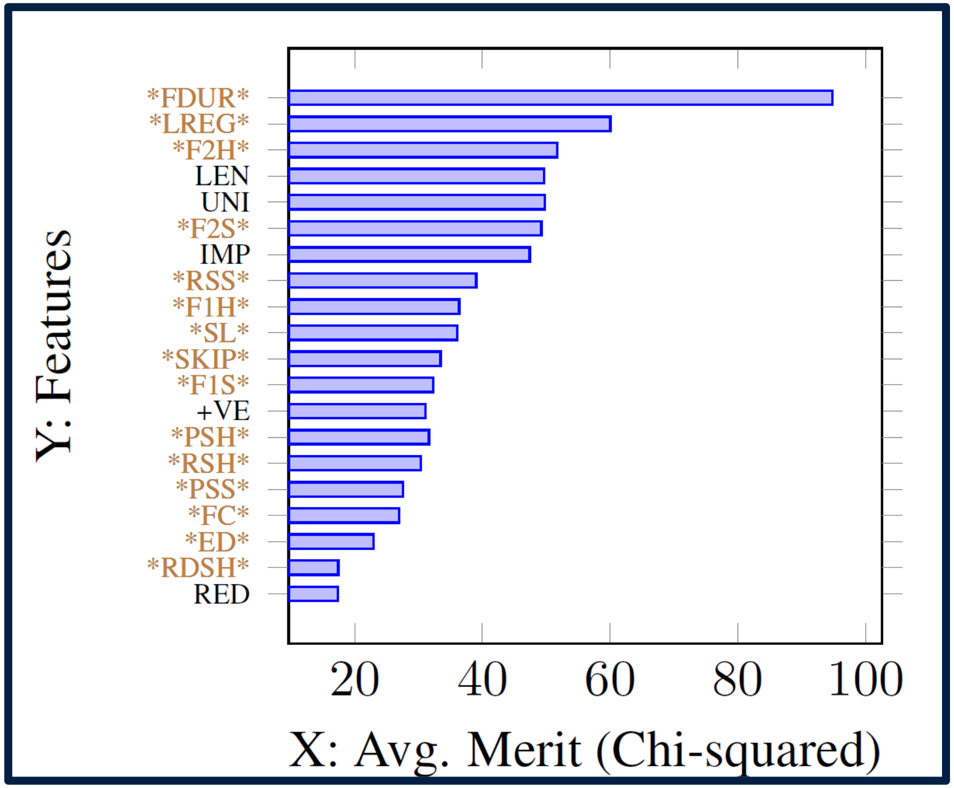
Results

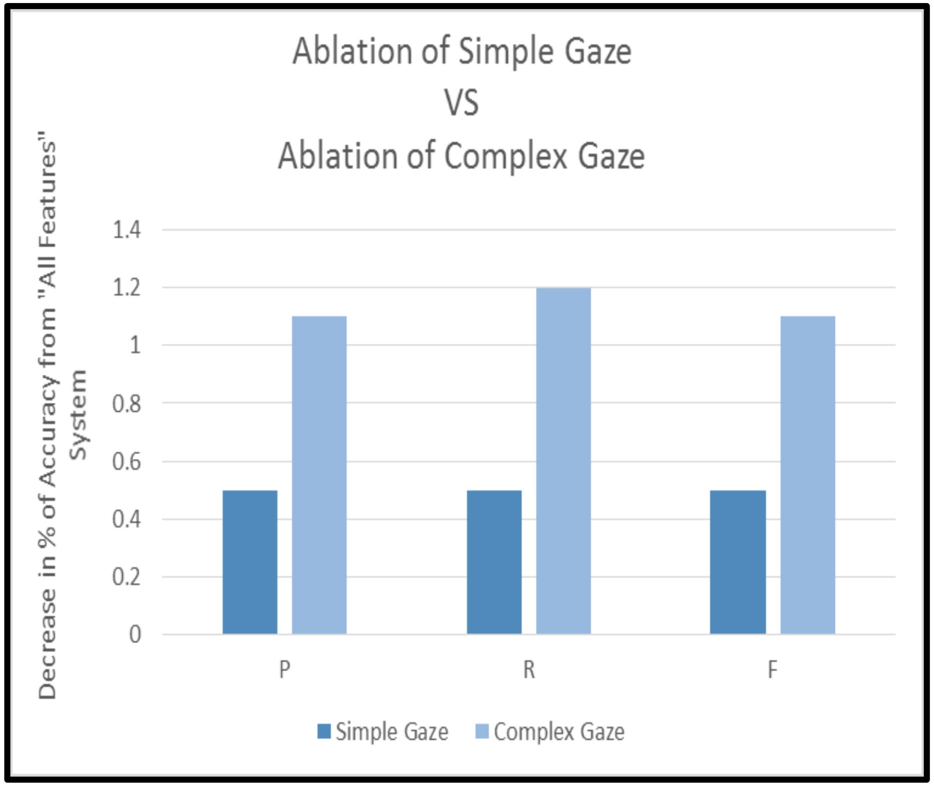
Features	P(1)	P(-1)	P(avg)	R(1)	R(-1)	R(avg)	F(1)	F(-1)	F(avg)
Multi Layered Neural Network									
Unigram	53.1	74.1	66.9	51.7	75.2	66.6	52.4	74.6	66.8
Sarcasm (Joshi et. al.)	59.2	75.4	69.7	51.7	80.6	70.4	55.2	77.9	69.9
Gaze	62.4	76.7	71.7	54	82.3	72.3	57.9	79.4	71.8
Gaze+Sarcasm	63.4	75	70.9	48	84.9	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	72.2
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

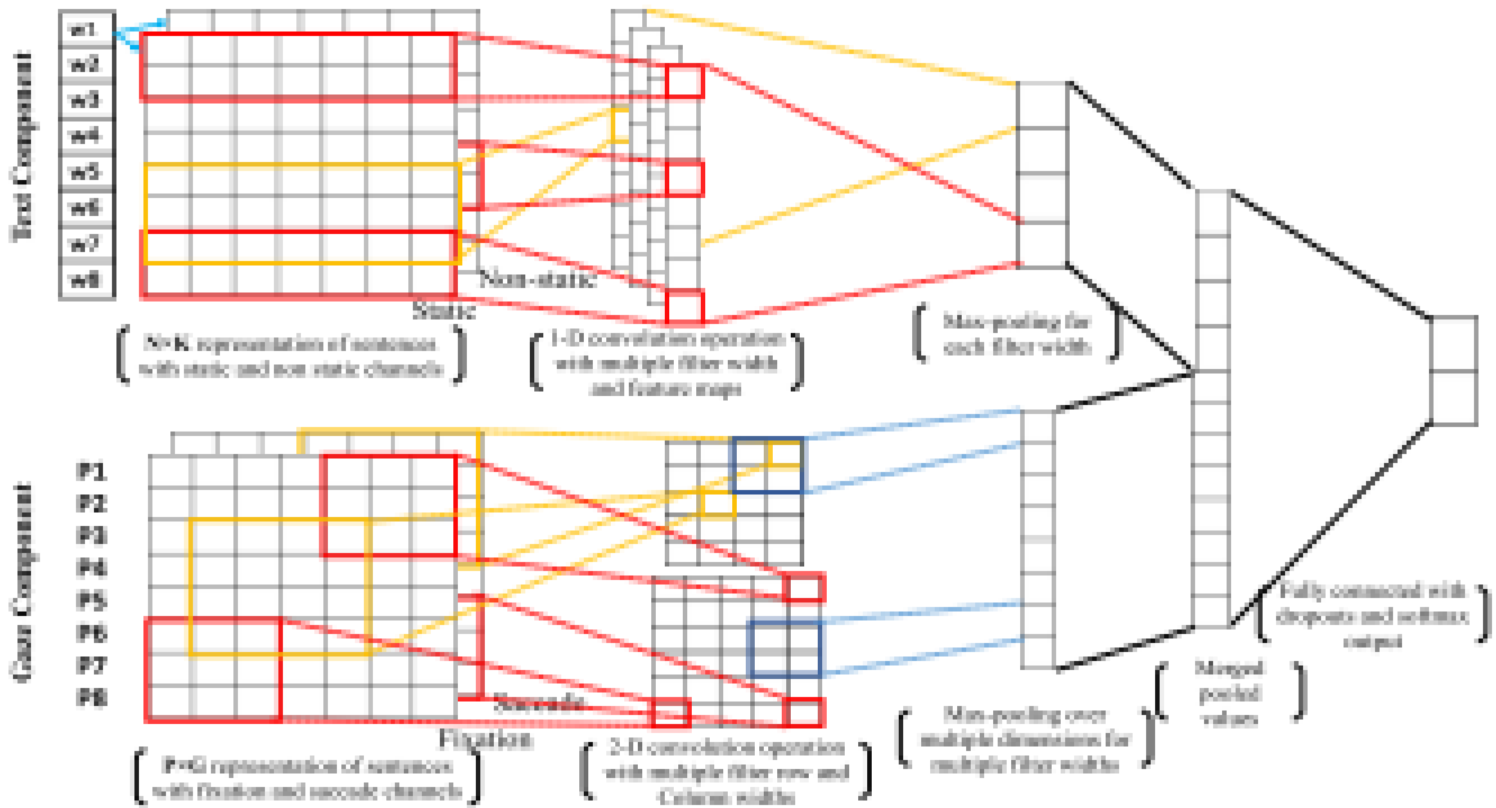
Feature Significance





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

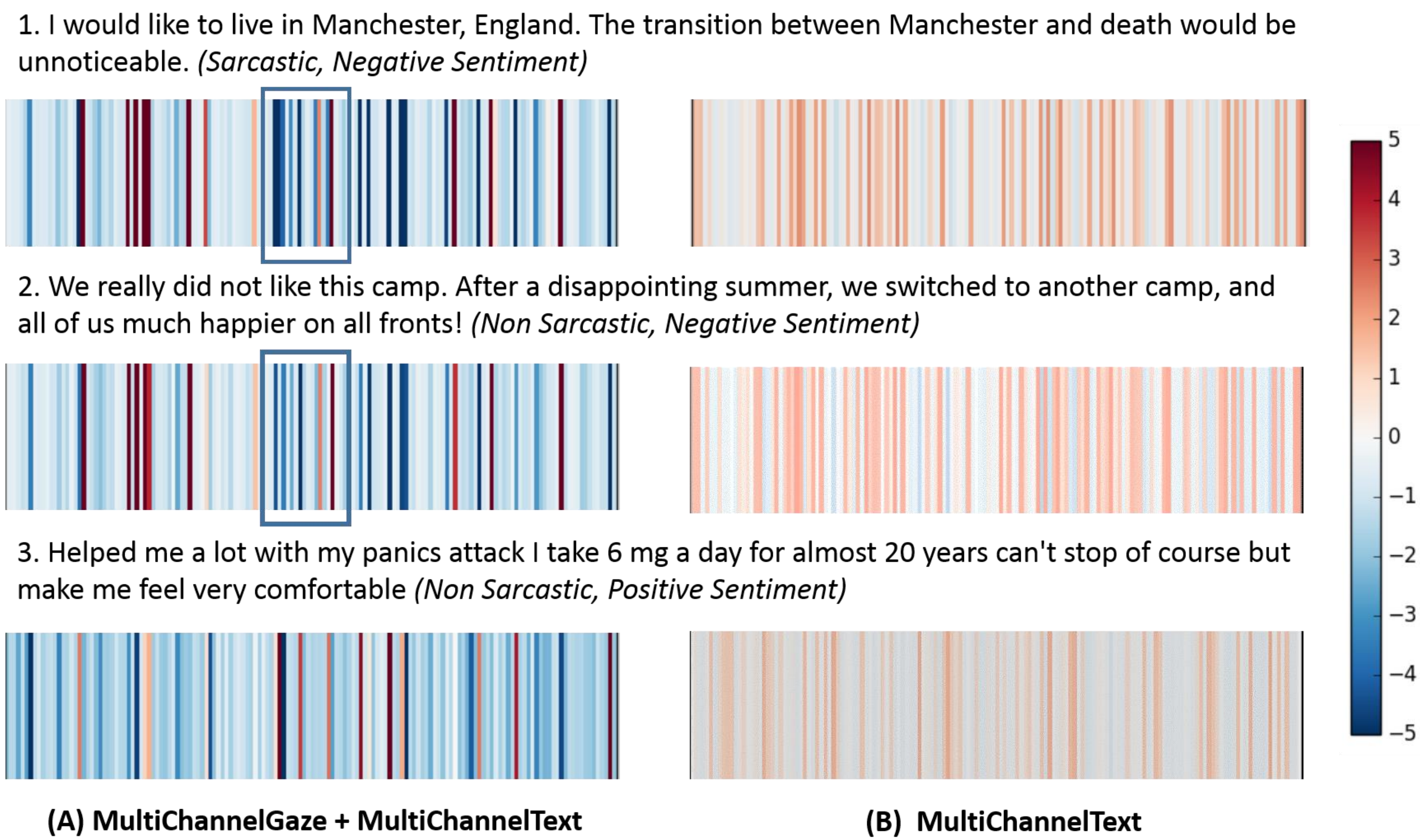
Configuration	Precision	Recall	F_Score
Gaze_Text_NB	**70.9**	71.9	71.2
Gaze_Text_SVM	**74**	**74.1**	**74**
Gaze_Text_MLP	70.9	71.9	70.9

(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)

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 data and annotation
- Success from Deep Learning

Future Work: All forms of **Incongruity**

- Humour (*A man coming back from movie notices "parking fine" on his car and thanks the policeman for appreciating his parking skill*)

- Humble bragging (*my leg aches everyday after inspecting all the 26 rooms in my small house!!*)

- Rumour and Fake News detection

- **Solution**: incongruity + additional machinery (*what?*)

Future Work: Resource building, Lab→land, Multilinguality-multimodality

- Mine the web for more training data of numerical saracasm, and build interface to collect sarcasm snippets
- Perform large scale sentiment and sarcasm detection on social media, tweet, blogs etc.
- Multi and Cross lingual sarcasm study (very culture and language dependent)
- Multimodal sentiment analysis- picture, speech and text (“haa aap to bade aadmi hai”)

Resources and Publications

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

THANK YOU