NLP= Linguistics + ML?

Pushpak Bhattacharyya
IIT Patna and IIT Bombay
CICLING 2018- special event
23rd March, 2018
Hanoi

NLP= Linguistics + ML?

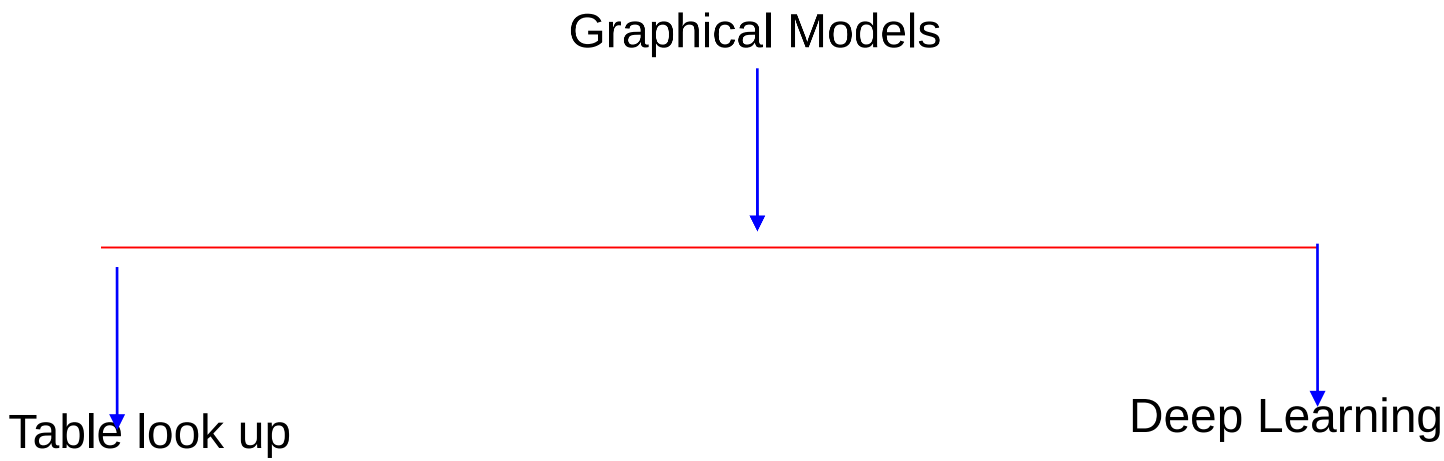
Phenomenon Study

Technique Study

No quarrel really!

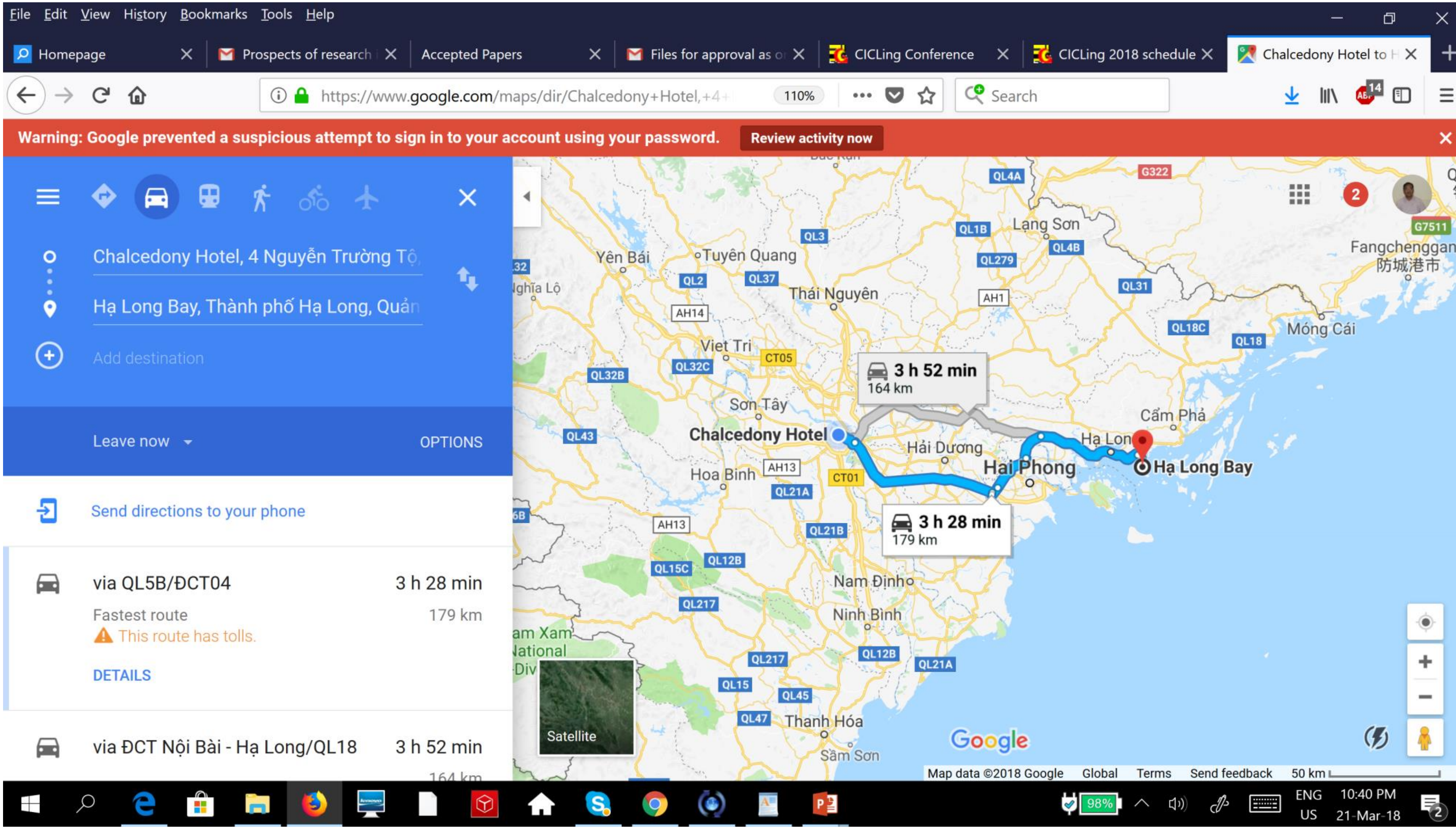
Phenomenon study leads to correct TECHNIQUE !!

ML can be very simple to very sophisticated



Perspectivising

Chalcedony Hotel to Ha Long Bay



How did Google calculate the time?

- Rule Based?
- ML Based?
- DL Based?

Dig deeper!

- Rule (ideal condition):
  - ***Time= Shortest Path/Av.-speed***
  - **simple, clean, application of speed-time-distance equation**
- Enter traffic (chaotic world, reality)
  - ***Time= Ideal Time + Traffic Time=*** $T_i + T_{tr}$

How to know T_{tr} ?

- $\boldsymbol{T_{tr} = F(Tr)}$,
  - where *Tr* models traffic

- If *F* is known, again a Rule Based Situation

- Phenomenon understood and expressed
  - Rules necessary and sufficient

- But *Tr* is chaotic, UNPREDICTABLE !!!
  - Say “unpredictable”, still want to learn? Ironical?
  - Very approximate model ☹

To model *Tr, where* $T_{tr}=F(Tr)$

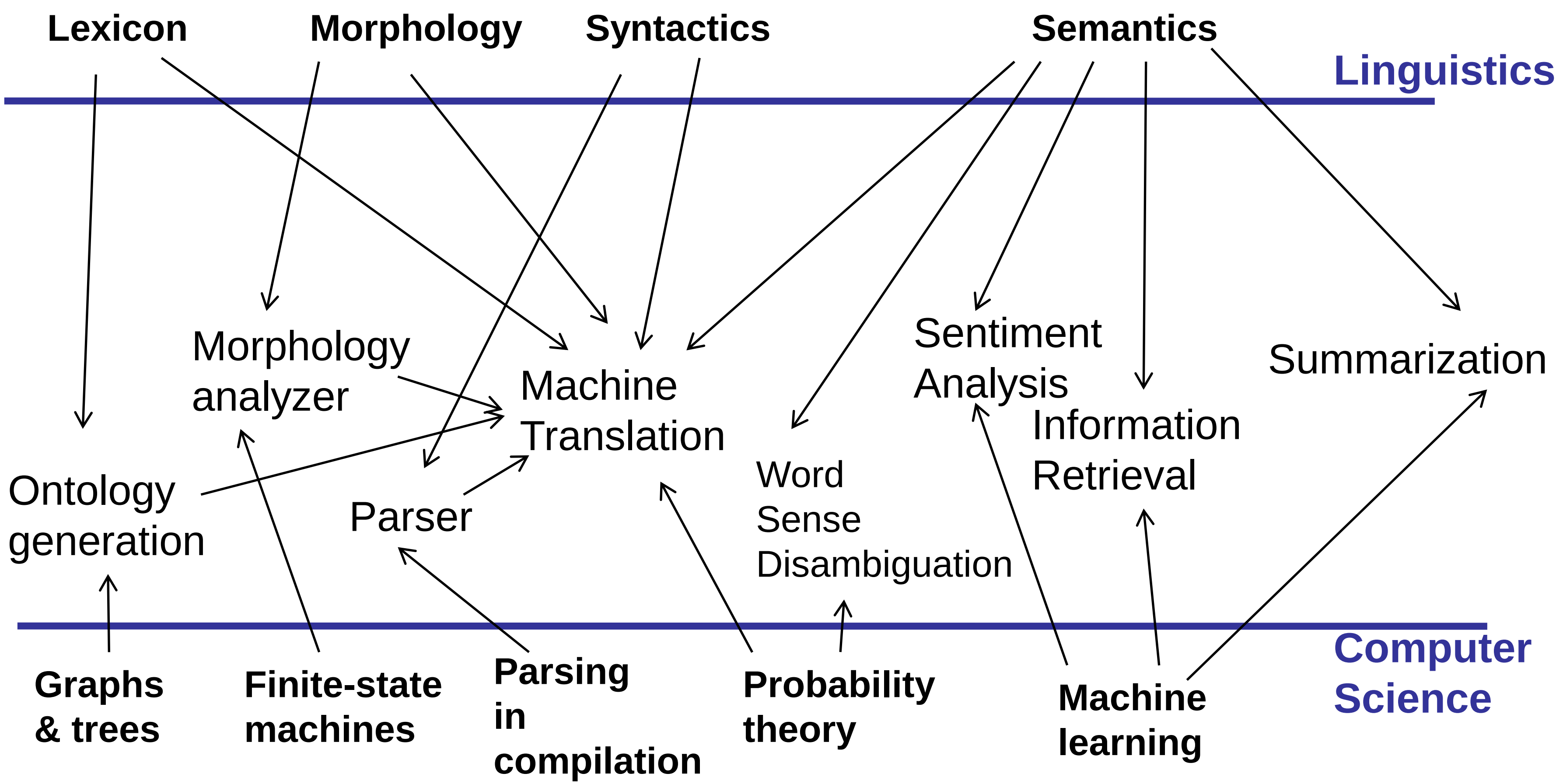
- Only handle is data
- Data → Model
- Also Known As, **MACHINE LEARNING**
- Produce a model that FITS THE DATA
- ONLY THAT MUCH !
- Also known as, *Maximum Likelihood*

Summarizing points from maps discussion

- LEARN WHEN YOU DO NOT KNOW
- Learn by observing
- Learn from data
- Learn my MLE
- This is the predominant modus operandi

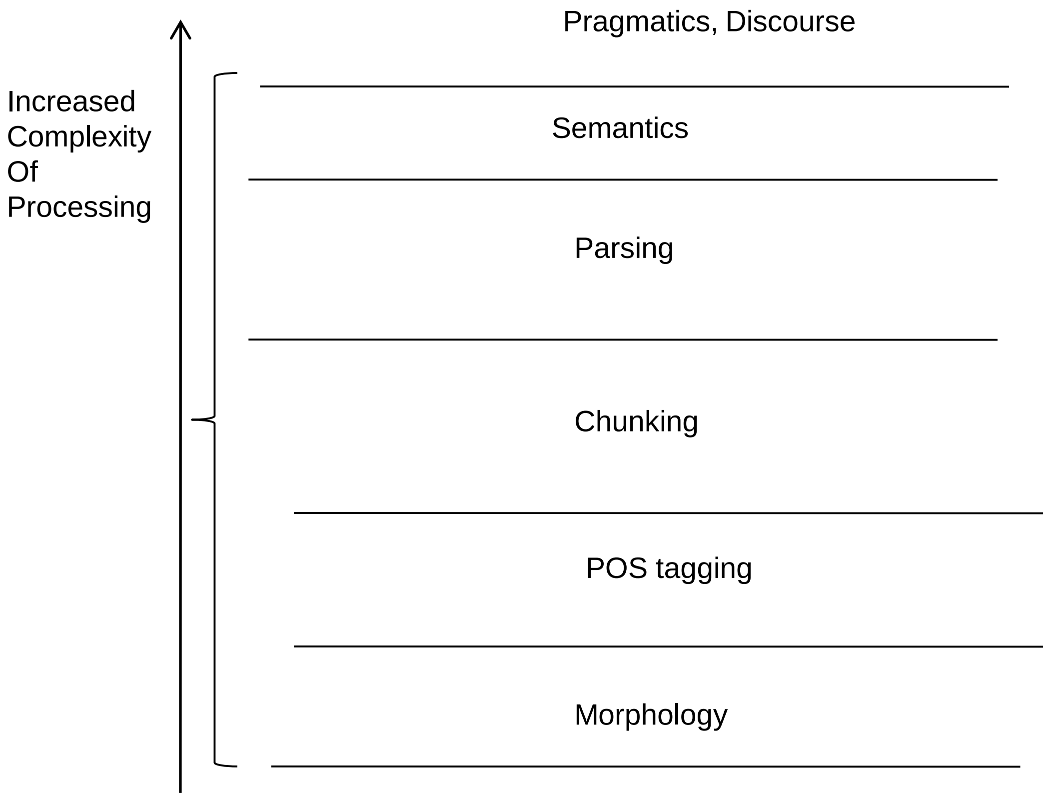
Nature of CL/NLP

NLP: At the confluence of linguistics & computer science



Linguistics is the Eye, Computation is the Body

NLP: multilayered, Multi dimensional



Each layer provides

FEATURES

To *Next upper layers*

Ambiguity: the crux of the problem (Language)

1. (ellipsis) Amsterdam airport: “Baby Changing Room”

2. (Attachment/grouping) Public demand changes (credit for the phrase: Jayant Haritsa):

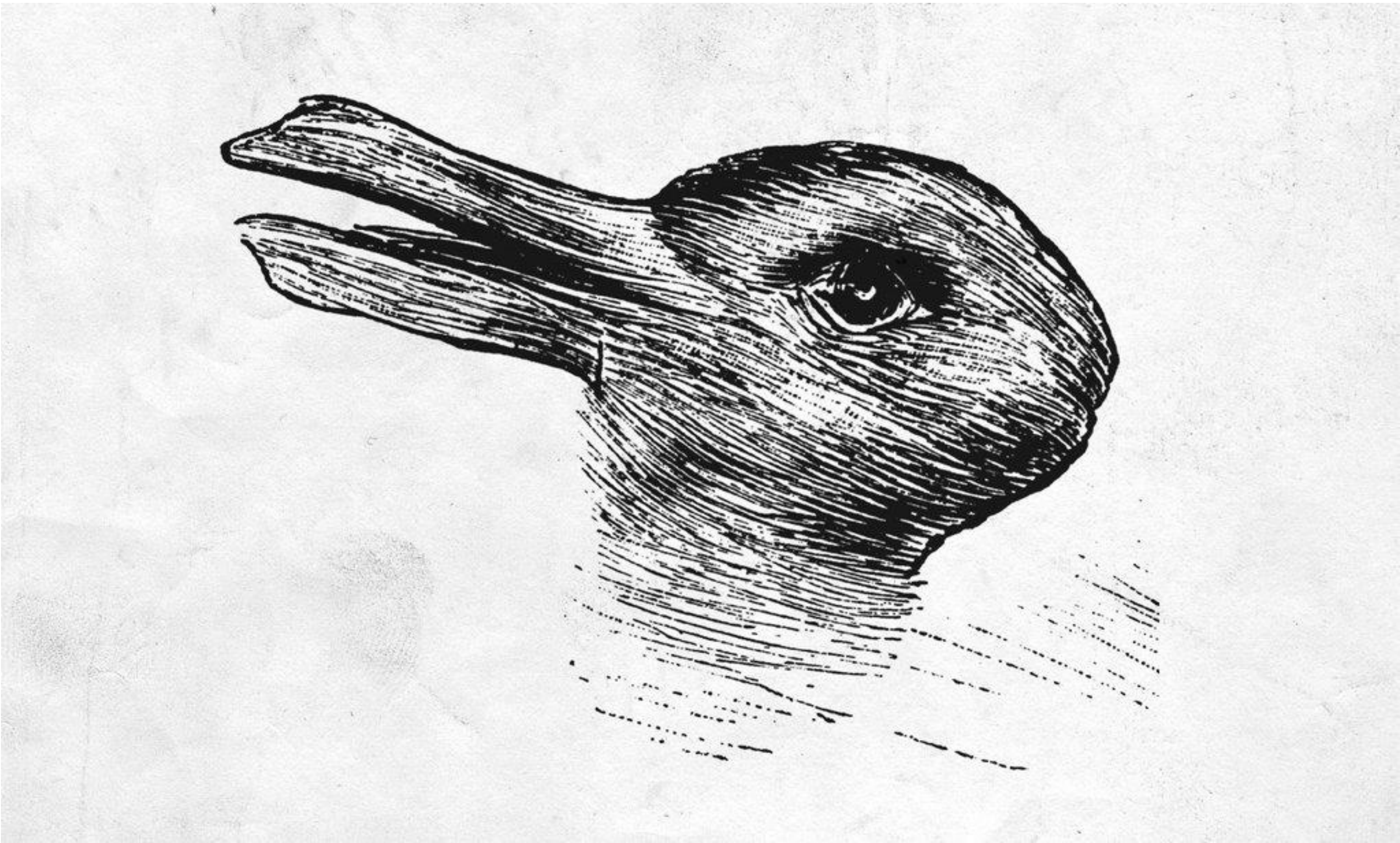
   (a) Public demand changes, but does any body listen to them?

   (b) Public demand changes, and we companies have to adapt to such changes.

   (c) Public demand changes have pushed many companies out of business

3. (Pragmatics-1) The use of shin bone is to locate furniture in a dark room

Ambiguity is prevalent in all of AI (picture)



Role of Machine Learning

- Ambiguity resolution by classification

- Multiple classes

- Choose the one with HIGHEST SCORE

- Score=Probability

New age NLP

Emphasis on FAST, ROBUST, SHALLOW processing

Impact of probability

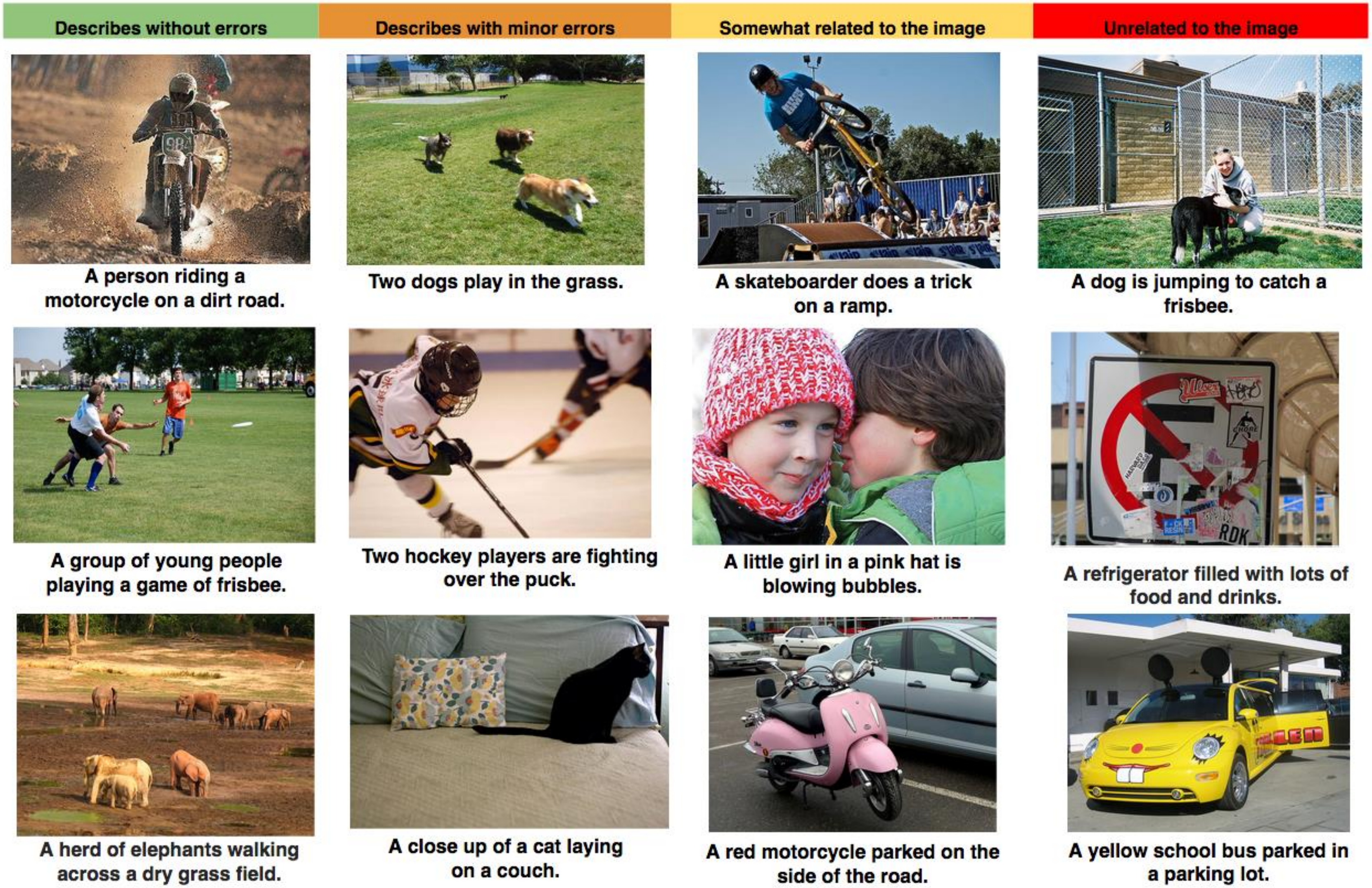
Probabilities computed in the context of corpora

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.

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

Change of point of view in ML-NLP

- “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***

NLP= Linguistics + ML

Win situation for ML!

Machine Translation

Paradigm Shift: 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$\leftarrow\rightarrow$ w), t(a$\leftarrow\rightarrow$ 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

Part of Speech Tagging

With Hidden Markov Model

NLP Layer

What a gripping movie was TITANIC!

What/WP a/DT gripping/JJmovie/NN was/VBD Dangal/NNP !/.

Parse

```
(ROOT
  (FRAG
    (SBAR
      (WHNP
        (WP What))
        (S
          (NP
            (DT a)
            (JJ gripping)
            (NN movie)
          )
          (VP
            (VBD was)
            (NP
            (NNP TITANIC)))))
          (. !)
  )
)
```

Universal dependencies

dobj(TITANIC-6, What-1)
det(movie-4, a-2)
amod(movie-4, gripping-3)
nsubj(TITANIC-6, movie-4)
cop(TITANIC-6, was-5)
root(ROOT-0, Dangal-6

Part of Speech Tagging

- POS Tagging: attaches to each word in a sentence a part of speech tag from a given set of tags called the **Tag-Set**
- Standard Tag-set : Penn Treebank (for English).

POS ambiguity instances

best ADJ ADV NP V
better ADJ ADV V DET
close ADV ADJ V N
cut V N VN VD
even ADV DET ADJ V
grant NP N V –
hit V VD VN N
lay ADJ V NP VD
left VD ADJ N VN
like CNJ V ADJ P –
near P ADV ADJ DET
open ADJ V N ADV
past N ADJ DET P
present ADJ ADV V N
read V VN VD NP
right ADJ N DET ADV
second NUM ADV DET N
set VN V VD N –
that CNJ V WH DET

Part-of-speech tag

- A word can have more than one POS tags.

Adjective

- E.g.
  1. *What a **gripping** movie was TITANIC!*

  2. *He is **gripping** it firm.*

Verb

Linguistic fundamentals

- A word can have two roles
  - Grammatical role (Dictionary POS tag)
  - Functional role (Contextual POS tag)

  - E.g. ***<u>Golf</u>* stick**

- POS tag of "*Golf*"
  - Grammatical: Noun
  - Functional: Adjective (+ al)

The "al" rule!

- If a word has different functional POS tag than its grammatical pos then add **"al"** to the functional POS tag

- E.g. ***<u>Golf</u>*** *stick*

Adjective + al

$\underbrace{\hphantom{\textbf{Adjective + al}}}_{}$

Adjectival

Noun	**+ al**	**= Nominal**
Verb	**+ al**	**= Verbal**
Adjective	**+ al**	**= Adjectival**
Adverb	**+ al**	**= Adverbial**

The “al” rule cntd.

- Examples:
  - Nominal
    - *Many don’t understand the problem of* ***hungry****.*
  - Adverbial
    - *Come* ***quick.***
  - Verbal

POS tagging as an ML problem

- Question
  - Is one instance of example enough for ML?
  - E.g. Known example of “people”

    People → Noun

    POS Ambiguity

  - But it can be verb as well

    People → Verb (to populate)
- Answer
  - We need at least as many instances as number of different labels (POS tags)-1 to make decision.

Disambiguation of POS tag

- If no ambiguity, learn a table of words and its corresponding tags.

- If ambiguity, then look for the contextual information i.e. look-back or look-ahead.

Data for “present”

1. He gifted me the/a/this/that **present_NN**.
2. They **present_VB** innovative ideas.
3. He was **present_JJ** in the class.

Rules for disambiguating “present”

- For Present_NN (look-back)
  - If present is preceded by determiner (the/a) or demonstrative (this/that), then POS tag will be noun.
  - Does this rule guarantee 100% precision and 100% recall?
  - False positive:
    - *The **present_ADJ** case is not convincing.*

Adjective preceded by “the”

  - False negative:
    - ***Present** foretells the future.*

Noun but not preceded by “the”

Rules for disambiguating "present"

- For Present_NN (look-back and look ahead)
  - If present is preceded by determiner (the/a) or demonstrative (this/that) or followed by a verb, then POS tag will be noun.
  - E.g.
    - **Present_NN** will tell the future.
    - **Present_NN** fortells the future.
  - Does this rule guarantee 100% precision and 100% recall?

Need for ML in POS tagging

- New examples break rules, so we need a robust system.

- Machine learning based POS tagging:
  - HMM (Accuracy increased by 10-20% against rule based systems)
  - Jelinek's work

Mathematics of POS Tagging

Best tag sequence

= T*

= argmax P(T|W)

= argmax P(T)P(W|T)   (by Baye's Theorem)

$P(T) = P(t_0=\text{^}\ t_1 t_2 \ldots t_{n+1}=.)$

$= P(t_0)P(t_1|t_0)P(t_2|t_1 t_0)P(t_3|t_2 t_1 t_0) \ldots P(t_n|t_{n-1} t_{n-2} \ldots t_0)P(t_{n+1}|t_n t_{n-1} \ldots t_0)$

$= P(t_0)P(t_1|t_0)P(t_2|t_1) \ldots P(t_n|t_{n-1})P(t_{n+1}|t_n)$

$= \prod_{i=0}^{N+1} P(t_i|t_{i-1})$ Bigram Assumption

Argmax computation

$P(W|T) = P(w_0|t_0\text{-}t_{n+1})P(w_1|w_0t_0\text{-}t_{n+1})P(w_2|w_1w_0t_0\text{-}t_{n+1}) \ldots$
$P(w_n|w_0\text{-}w_{n-1}t_0\text{-}t_{n+1})P(w_{n+1}|w_0\text{-}w_nt_0\text{-}t_{n+1})$

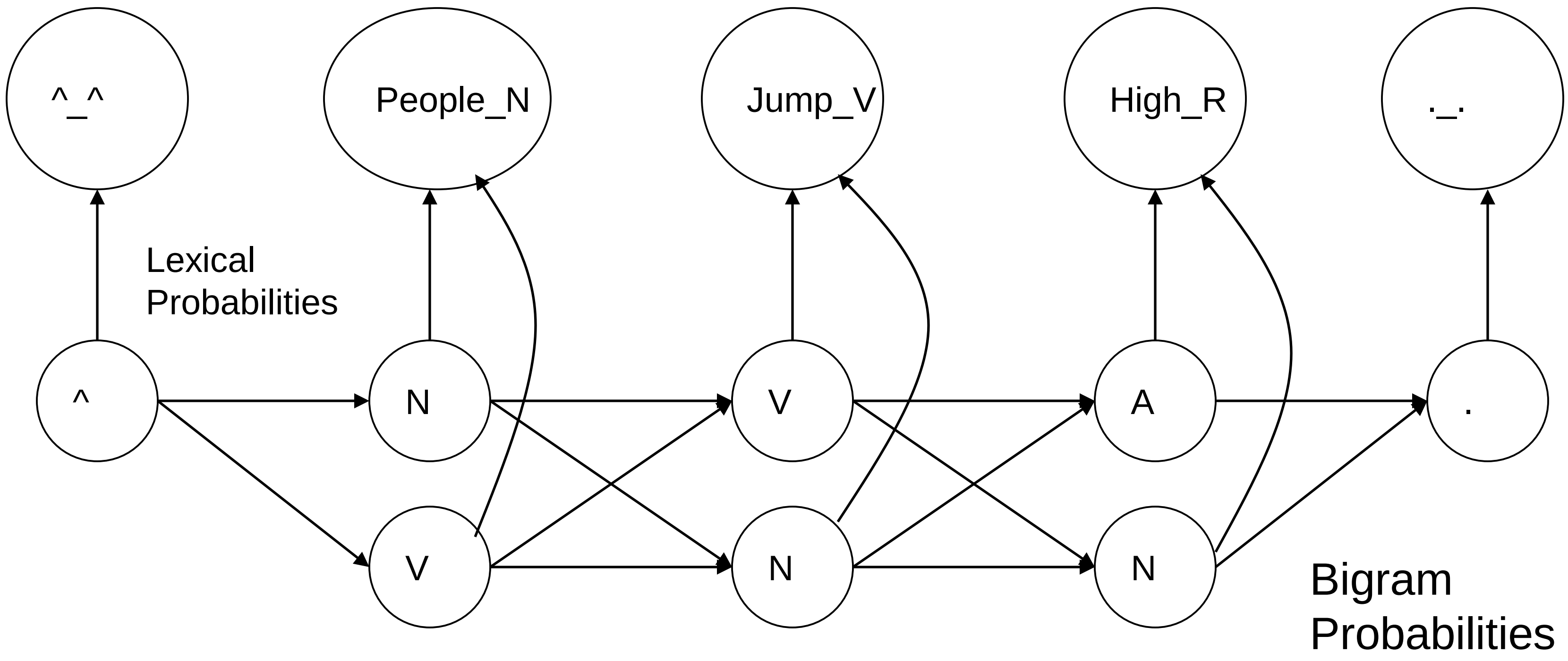
Assumption: A word is determined completely by its tag. This is inspired by speech recognition

$$= \prod_{i=0}^{n+1} P(w_o|t_o)P(w_1|t_1) \ldots P(w_{n+1}|t_{n+1})$$

$$= \prod_{i=1}^{n+1} P(w_i|t_i)$$

$$= \prod_{i=1}^{n+1} P(w_i|t_i) \quad \text{(Lexical Probability Assumption)}$$

Generative Model



This model is called Generative model.
Here words are observed from tags as states.
This is similar to HMM.

Typical POS tag steps

- Implementation of Viterbi – Unigram, Bigram.
- Five Fold Evaluation.
- Per POS Accuracy.
- Confusion Matrix.

NLP= Linguistics + ML

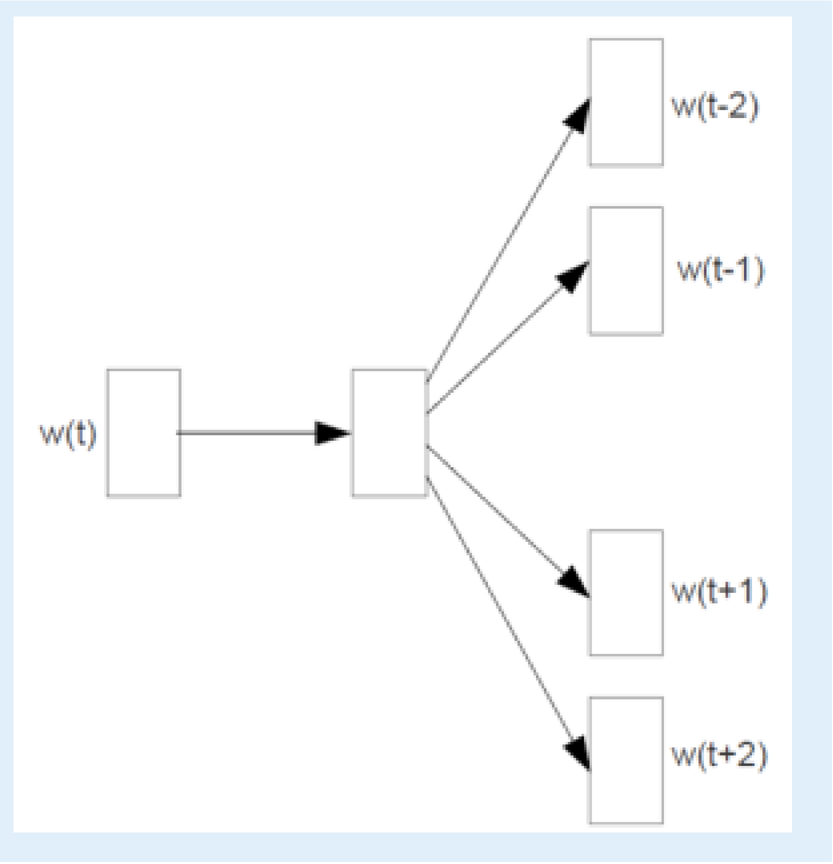
Win situation for linguistics (look closely) !

Word embedding

“Linguistics is the eye”: Harris Distributional Hypothesis

- Words with similar distributional properties have similar meanings. (Harris 1970)
- 1950s: Firth- “A word is known by the company its keeps”
- 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= Linguistics + ML

Win for ML!

Numerical Sarcasm

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!

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)

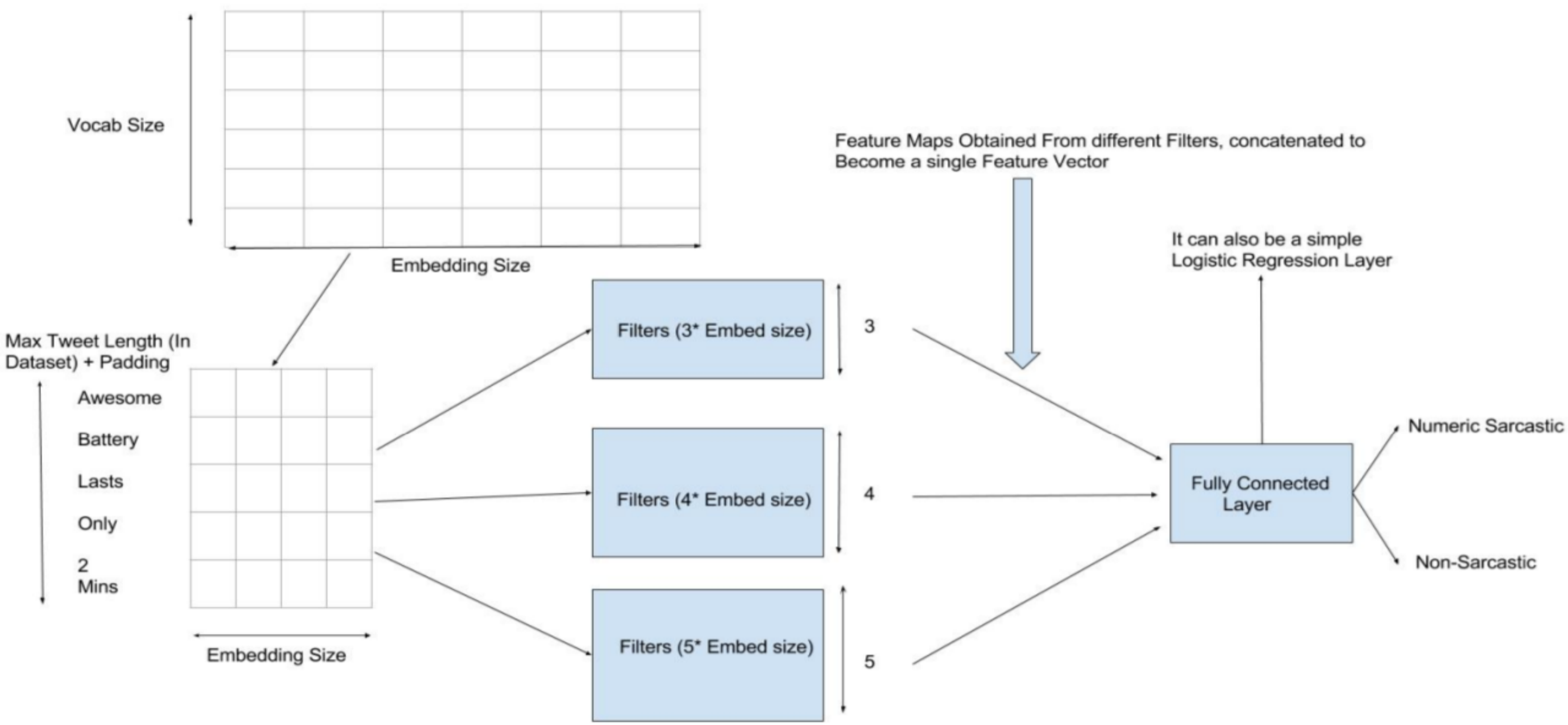
Approaches	Precision			Recall			F-score		
	P(1)	**P(0)**	**P(avg)**	**R(1)**	**R(0)**	**R(avg)**	**F(1)**	**F(0)**	**F(avg)**
Past Approaches									
Buschmeier et.al.	0.19	0.98	0.84	0.99	0.07	0.24	0.32	0.13	0.16
Liebrecht et.al.	0.19	1.00	0.85	1.00	0.07	0.24	0.32	0.13	0.17
Gonzalez et.al.	0.19	0.96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

NLP=Linguistics+ML

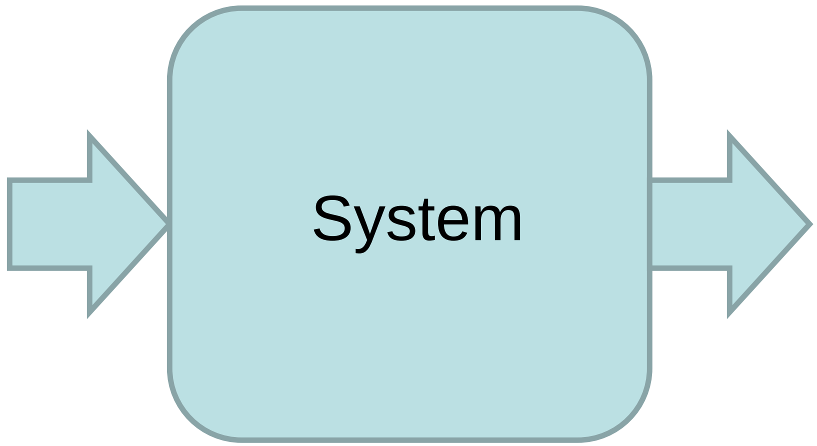
Linguistics (ontology engineering) in the background

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

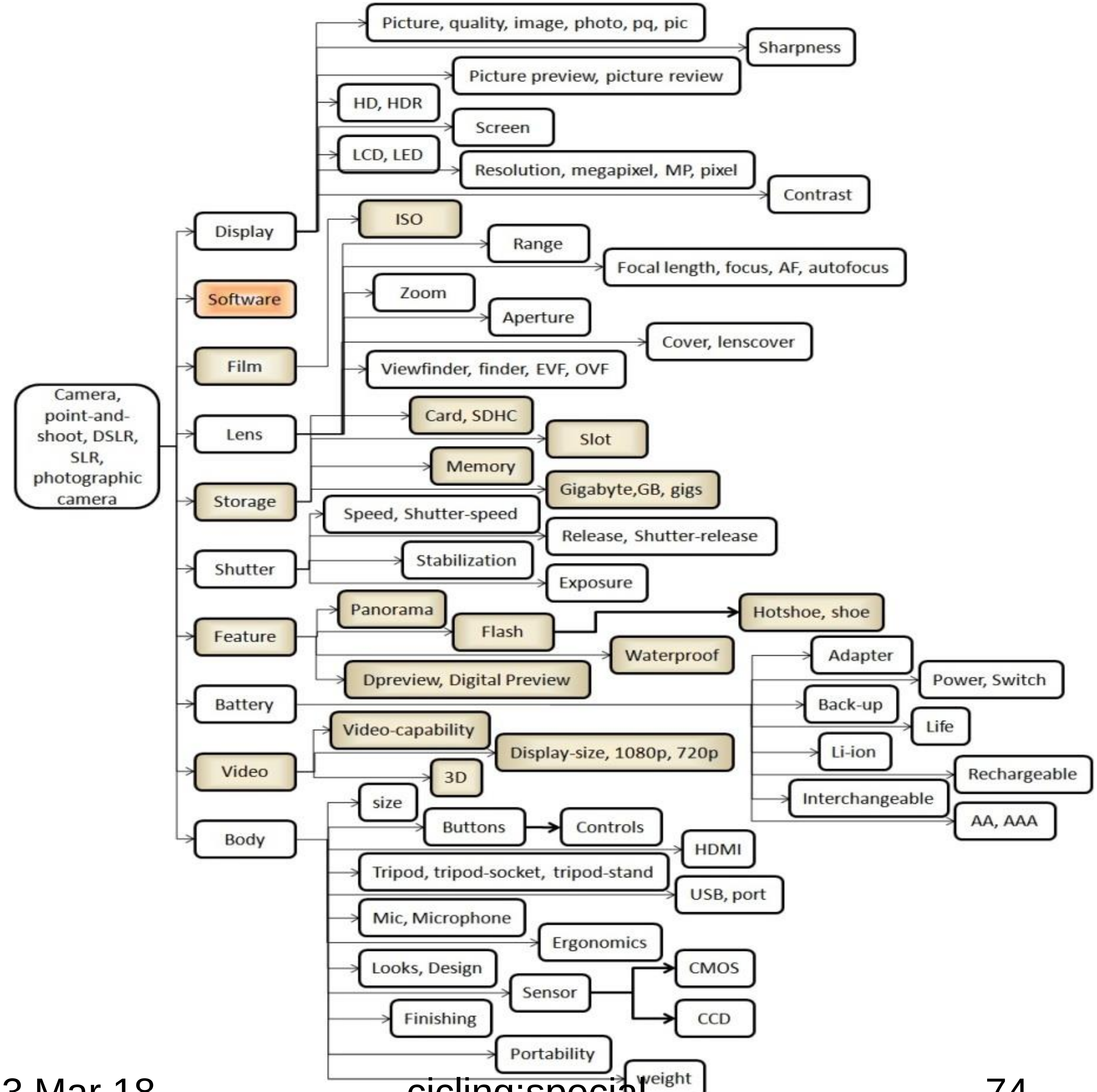
This camera has everything that you need. A Superb lens, an amazing picture quality and a long battery life. I love it.

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

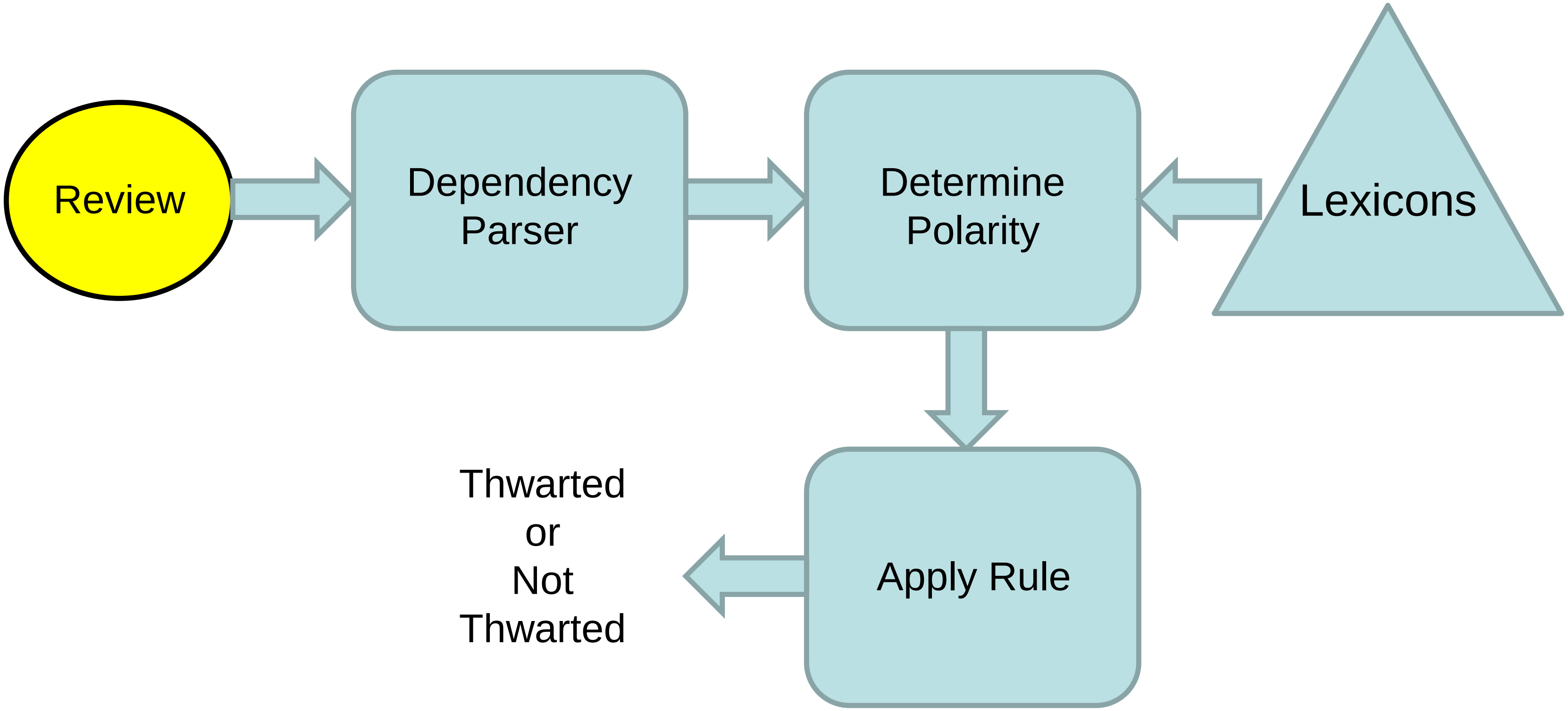
Basic idea

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

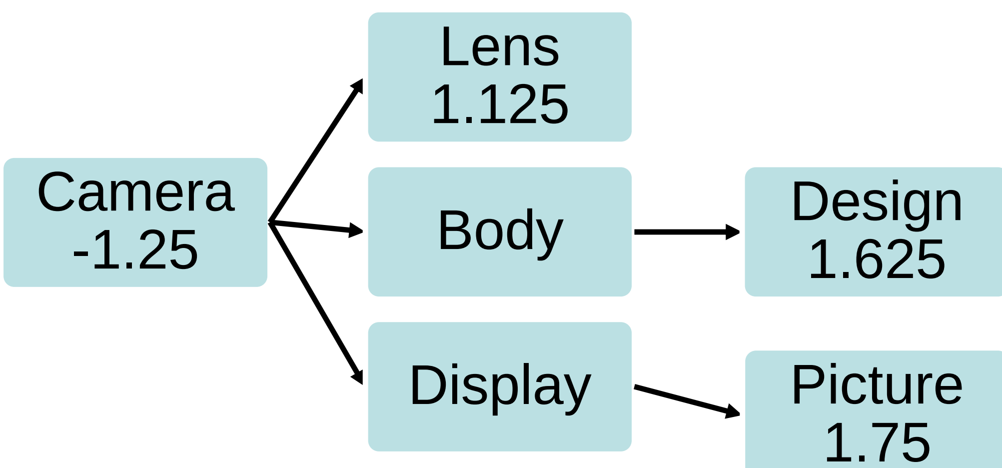


Dependency, weighting, decision

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

Extract Weights

- Domain aspects: A_1, A_2 ... A_N
- Weights: W_1, W_2 ... 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

$$V_1, V_2 \ \ldots \ V_M$$

from each review.

Each V_i represents a weighted aspect polarity value.

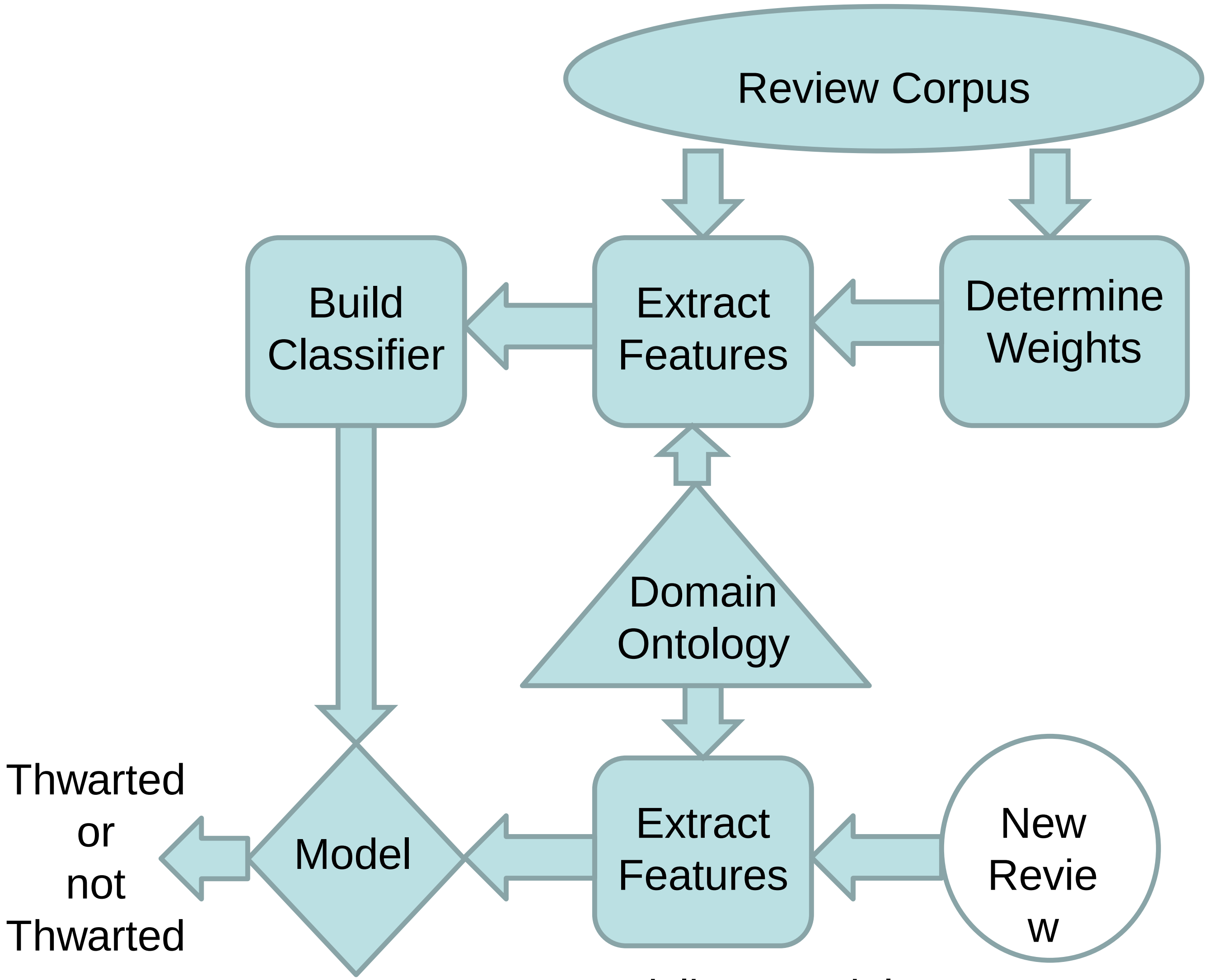
Features (1/2)

- 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

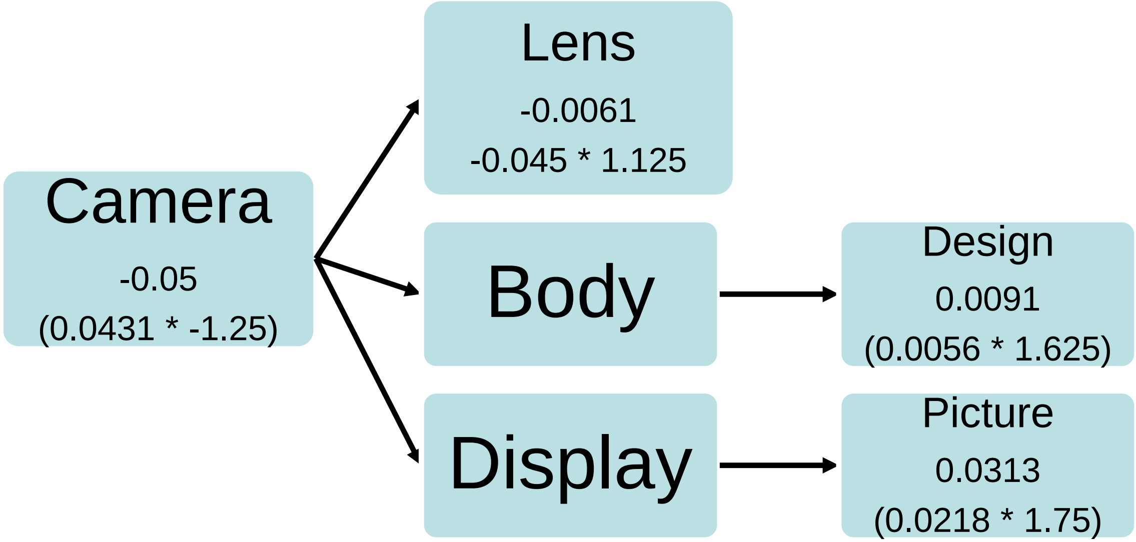
Process flow



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

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%

Summary (1/2)

- Examined the relative weight of linguistics and ML in NLP=Linguistics + ML
  - Word embedding
  - SMT
  - Numerical Sarcasm
  - Thwarting
- When phenomenon is UNDERSTOOD and EXPRESSED, use RULES
- Else use DATA
  - Essentially Maximum Likelyhood

Summary (1/2)

- When phenomenon is UNDERSTOOD and EXPRESSED
  - use RULES
  - Else use DATA
- Essentially Maximum Likelihood
- Model to fit the observations and HOPE that it works for unseen situation

New world order

- Trinity of **<*data, technique, idea*>**
- First two available a-plenty; Opening up the playing field for those with ideas
- Playing field is not a level one. There are Have’s and Have-Not’s in the world of data, as also the situation of rich getting richer
- Cycle of ***data-application-more_data*** is a steeply ascending gradient
- **Googles and Facebooks and Ubers** have and will have access to unimaginable volume of data. They will try to either outsmart entities with ideas or acquire them.

Thank you