CASCADE

ContextuAl SarCAsm DEtection in Online Discussion Forums

Devamanyu Hazarika [1], Soujanya Poria [2], Sruthi Gorantla [3],
Erik Cambria [2], Roger Zimmermann [1], Rada Mihalcea [4]

[1] National University of Singapore
[2] Nanayang Technological University, Singapore
[3] Indian Institute of Science, Bangalore
[4] University of Michigan, Ann Arbor

COLING 2018
August 20-26, 2018, Santa Fe, New Mexico, USA

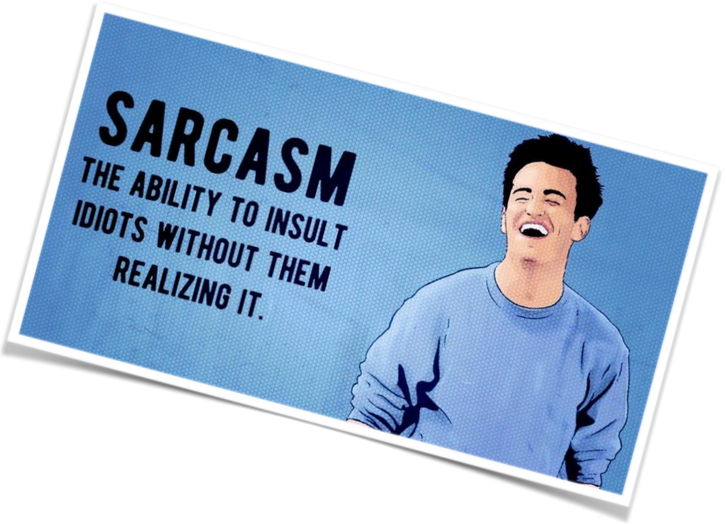
Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

What is Sarcasm ?



sarcasm

noun [U] · UK /ˈsɑː.kæz.ᵊm/ US /ˈsɑːr.kæz.ᵊm/

the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticize something in a humorous way:

*"You have been working hard," he said with **heavy** sarcasm, as he looked at the empty page.*

- dictionary.cambridge.com

photo source: https://karmayogadaily.com/2017/06/25/6-reasons-sarcasm-superpower/

Types of Sarcasm ?

Explicit

- Depends on lexical and pragmatic cues.
  - Major Indicators : Interjections, Punctuations, Sentimental Shifts, etc.

Don't bother me. I am living happily ever after.

- We call this : Content-based sarcasm.
- Major focus in previous work.

Implicit

- Presumption of commonsense and background knowledge.
  - *I ll happily send you off to Mars*
    - Absurdity evident through common sense.
  - *I'm sure Hillary would have done that.*
    - Temporal info about occured events required.
- We call this : Context-based sarcasm.
- Gaining focus in recent research.

Types of Sarcasm ?

Our Aim

Create a hybrid model that leverages algorithms for both types of sarcasm.

Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

Content-based Models

Plethora of work in this domain

Paper Title	Author	Aspects considered
"Yeah Right": Sarcasm Recognition for Spoken Dialogue Systems	Tepperman et al. (2006)	Prosodic, spectral cues.
Clues for detecting irony in user-generated contents: oh...!! it's "so easy" ;-)	Carvalho et al. (2009)	Linguistic features: interjections, gestural cues, etc.
Semi-supervised recognition of sarcastic sentences in Twitter and Amazon	David et al. (2010)	Syntactic patterns
Identifying sarcasm in Twitter: a closer look	Roberto González-Ibáñez et al. (2011)	Role of emotions
Sarcasm as Contrast between a Positive Sentiment and Negative Situation	Ripoff et al. (2013)	Sentimental contrasts

Context-based Models

Usage of context has increased in recent years.

- Carvalho et al. (2009)
  - Text highly plagued by grammatical inaccuracies
  - Contain highly temporal and contextual information
- Wallace et al. (2014)
  - Traditional classifiers fail in cases where humans require context too.

Carvalho, Paula, et al. "Clues for detecting irony in user-generated contents: oh...!! it's so easy;-.", 2009

Wallace, Byron C., Laura Kertz, and Eugene Charniak. "Humans require context to infer ironic intent (so computers probably do, too).", 2014

Context-based Models

Sarcasm is online platforms

- Exploit historical posts by users (Rajadesingan et al. 2015; Zhang et al. 2016)
- Contrasting sentimental histories for users (Khattri et al. 2015)
- Forum-based modelling:
  - Wallace et al. (2015) : sentiments and noun-phrases within a forum to gather context

Rajadesingan, Ashwin, Reza Zafarani, and Huan Liu. "Sarcasm detection on twitter: A behavioral modeling approach.", 2015

Zhang, Meishan, Yue Zhang, and Guohong Fu. "Tweet sarcasm detection using deep neural network." 2016

Khattri, Anupam, et al. "Your sentiment precedes you: Using an author's historical tweets to predict sarcasm." 2015

Wallace, Byron C., and Eugene Charniak. "Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment." 2015

Context-based Models

Prime Inspirations

- User profiling in Reddit (Amir et al. 2016)
  - Learning user embedding that capture homophily.
  - We use stylometric and personality features (explained later).

- Role of emotions, sentiment and personality (Poria et al. 2016)
  - We incorporate personality features in our user-profiling process.

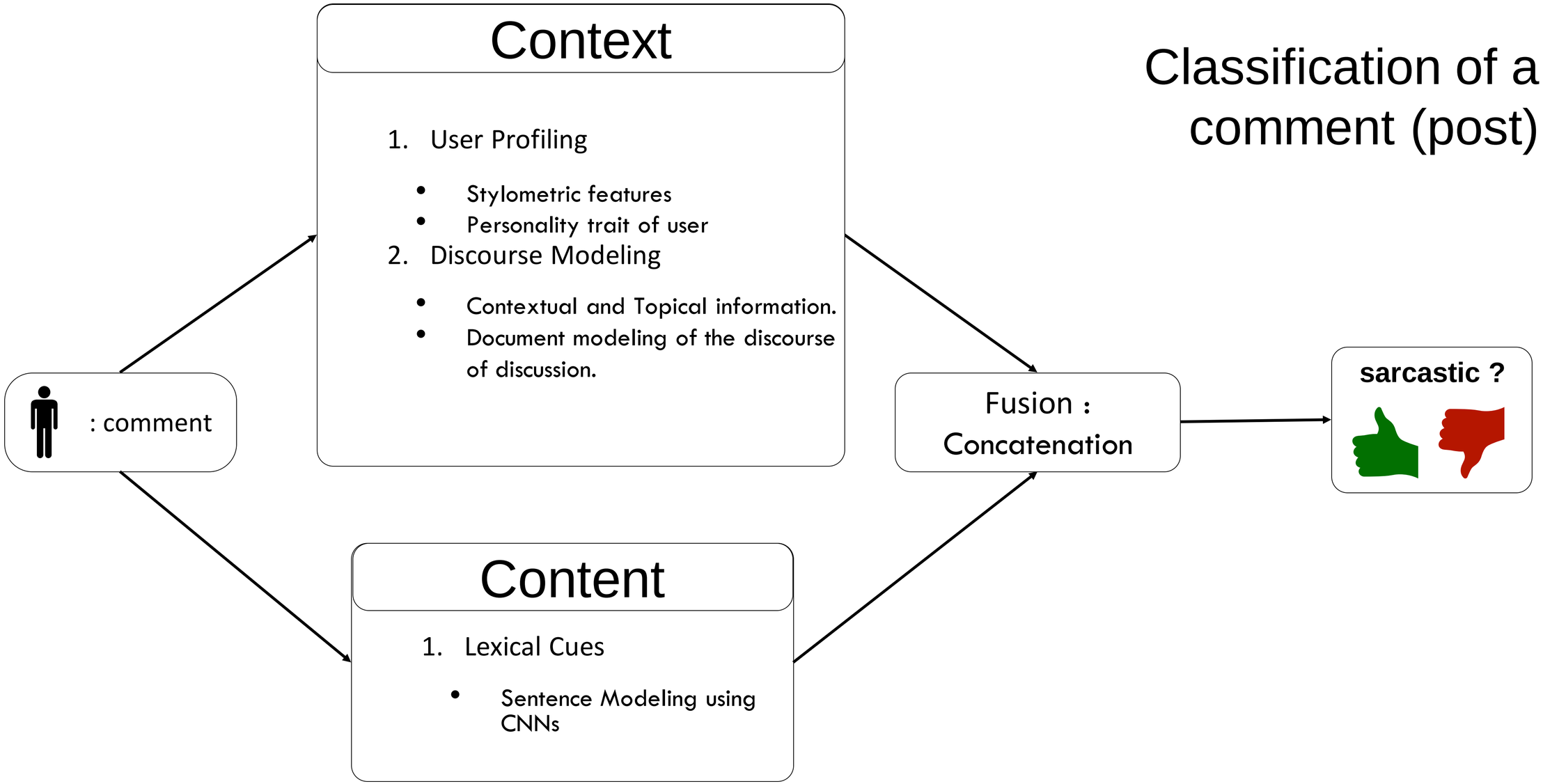
Amir, Silvio, et al. "Modelling context with user embeddings for sarcasm detection in social media." 2016

Poria, Soujanya, et al. "A deeper look into sarcastic tweets using deep convolutional neural networks." 2016

Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

Overview


Context
1. User Profiling
• Stylometric features
• Personality trait of user
2. Discourse Modeling
• Contextual and Topical information.
• Document modeling of the discourse of discussion.
: comment
Content
1. Lexical Cues
• Sentence Modeling using CNNs
Fusion : Concatenation
sarcastic ?
Classification of a comment (post)

Overview

Context

1. User Profiling
   - Stylometric features
   - Personality trait of user
2. Discourse Modeling
   - Contextual and Topical information.
   - Document modeling of the discourse of discussion.

: comment

Content

1. Lexical Cues
   - Sentence Modeling using CNNs

Fusion : Concatenation

sarcastic ?

Classification of a comment (post)

Modeling context

User Profiling

- Motivation
  - Users tend to be sarcastical / non-sarcastical across forums.
  - Utilize author's behavioral features as contextual information.

- We generate user-embeddings for each user based on two user-traits:
  - Stylometric Features
  - Personality Features

* Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." 2014.

Modeling context

User Profiling : Stylometric Features

- People possess their own idiolect and authorship style.
  - **Stylometric features** to incorporate their unique styles.

- Method
  - Each user u_i`s posts accumulated and modeled as a document d_i.
  - Embedding generated using ParagraphVector* algorithm.

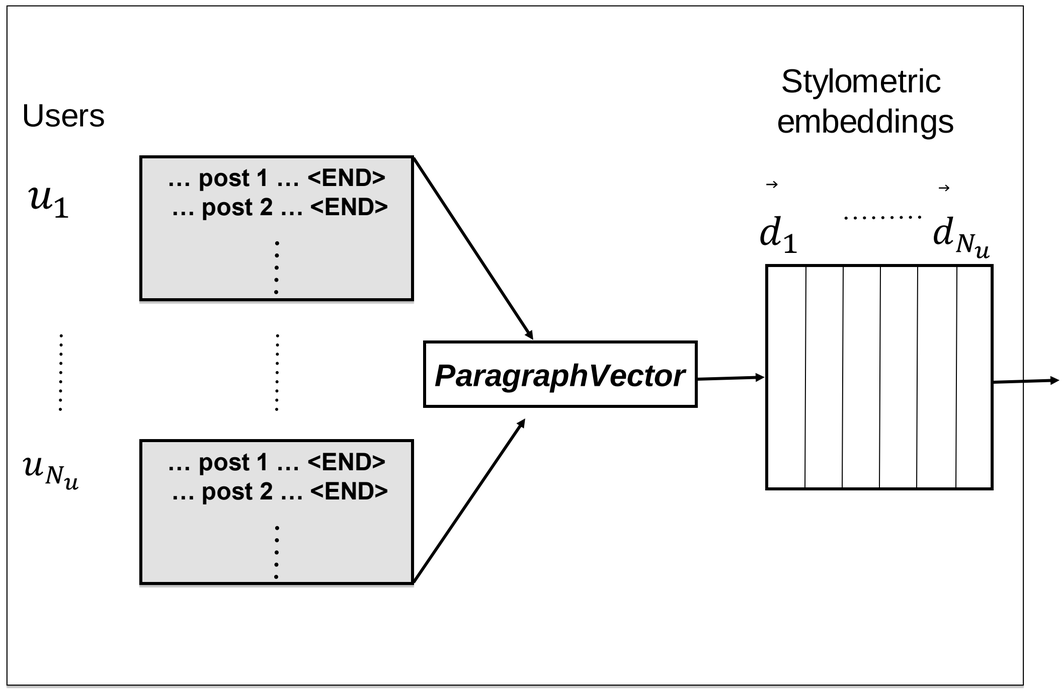
$$\vec{d_i} = ParagraphVector(d_i)$$

* Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." 2014.

Modeling context

User Profiling : Stylometric Features

Stylometric Features



User Profiling : Personality Features

- Personality traits correlate to behavioural patterns.
  - Generate **personality features** for each user.

- Method
  - Train a CNN model to predict personality traits in a benchmark dataset [1].
  - Extract expected personality features for each user based on their historical posts.

$$\vec{p_i} = \mathbb{E}_{j \in [v_i]}[\vec{p_{u_i}^j}] = \frac{1}{v_i} \sum_{j=1}^{v_i} \vec{p_{u_i}^j}$$

v_i - No. of posts by i^{th} user.

$\vec{p_{u_i}^j}$ - Personality feature from pre-trained CNN for j^{th} comment by user.

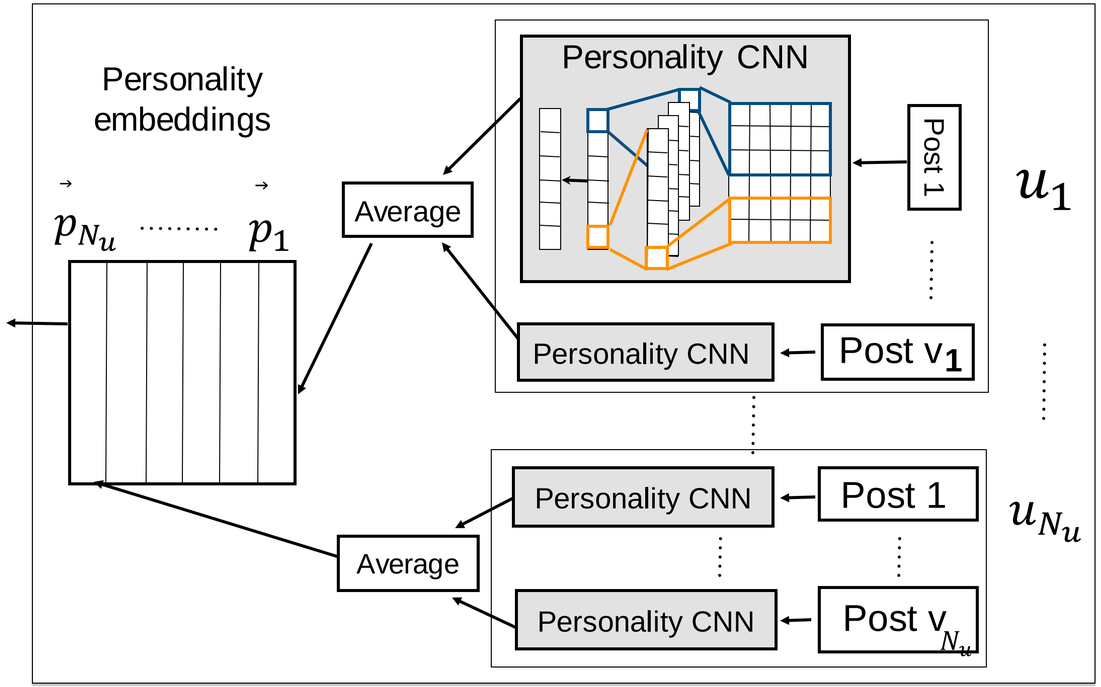
$\vec{p_i}$ - Expected personality trait of user.

[1] Poria, Soujanya, et al. "A deeper look into sarcastic tweets using deep convolutional neural networks." 2016

Modeling context

User Profiling : Personality Features

Personality Features



Modeling context

User Profiling : Multi-view Fusion using CCA

- User Embeddings are generated by fusing Stylometric and Personality features.
  - Canonical Correlation Analysis for multi-view fusion.
  - CCA captures maximal information (correlation) between views.[1]

- Let, stylometric embedding matrix be $D \in \mathbb{R}^{d_s \times N_u}$ and personality embedding matrix be $P \in \mathbb{R}^{d_p \times N_u}$
- User-Embedding of i[th] user is:

$$\vec{u_i} = (\vec{d_i})^T A_1 + (\vec{p_i})^T A_2$$

← **Fusion**

where, correlation between $W = D^T A_1$ **and** $Z = P^T A_2$ is maximized by assigning

$A_1 = R_{11}^{-\frac{1}{2}} A$, $A_2 = R_{22}^{-\frac{1}{2}} B$, for SVD factors A,B : $R_{11}^{-\frac{1}{2}} R_{12} R_{22}^{-\frac{1}{2}} = A \Lambda B^{\top}$

Here, R_{11}, R_{22} are correlation and R_{12} is cross-correlation metrix for D and P.

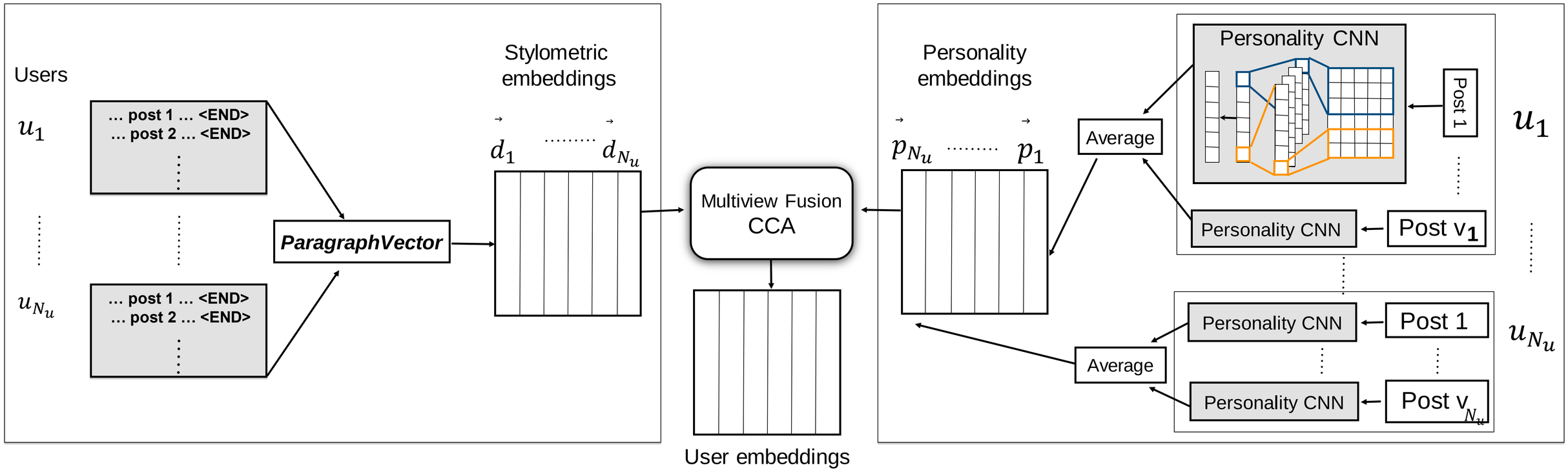
[1] Benton, Adrian, Raman Arora, and Mark Dredze. "Learning multiview embeddings of twitter users." 2016

Modeling context

User Profiling

Stylometric Features

Personality Features

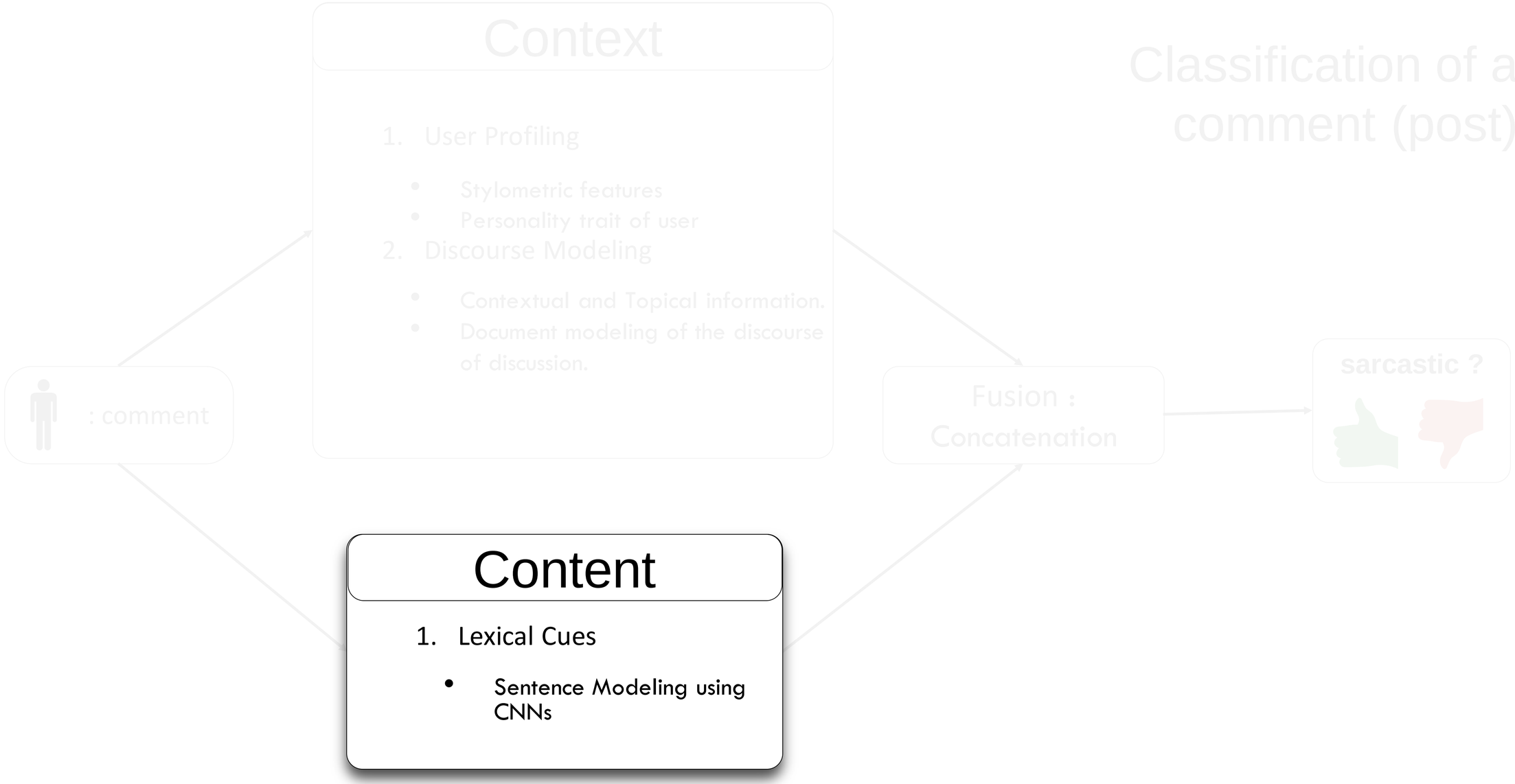


Discourse Modeling

- Motivation
  - Posts in a discussion thread have contextual dependencies.
  - Certain forum-topics correlate to sarcastic inclination of the posts.
    - e.g. Politics more prone to sarcasm than discussion on natural disasters.

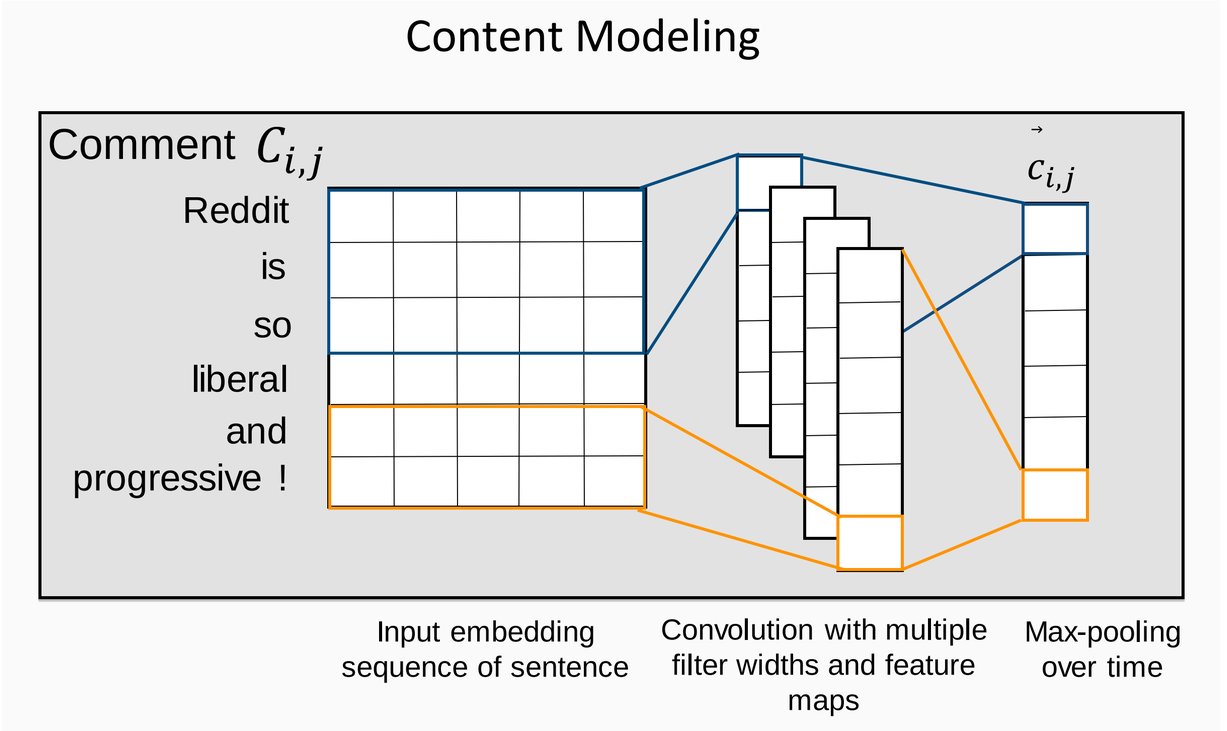
- Method
  - Similar to stylometric features, ParagraphVector model is applied on accumulated posts from each forum.

Overview

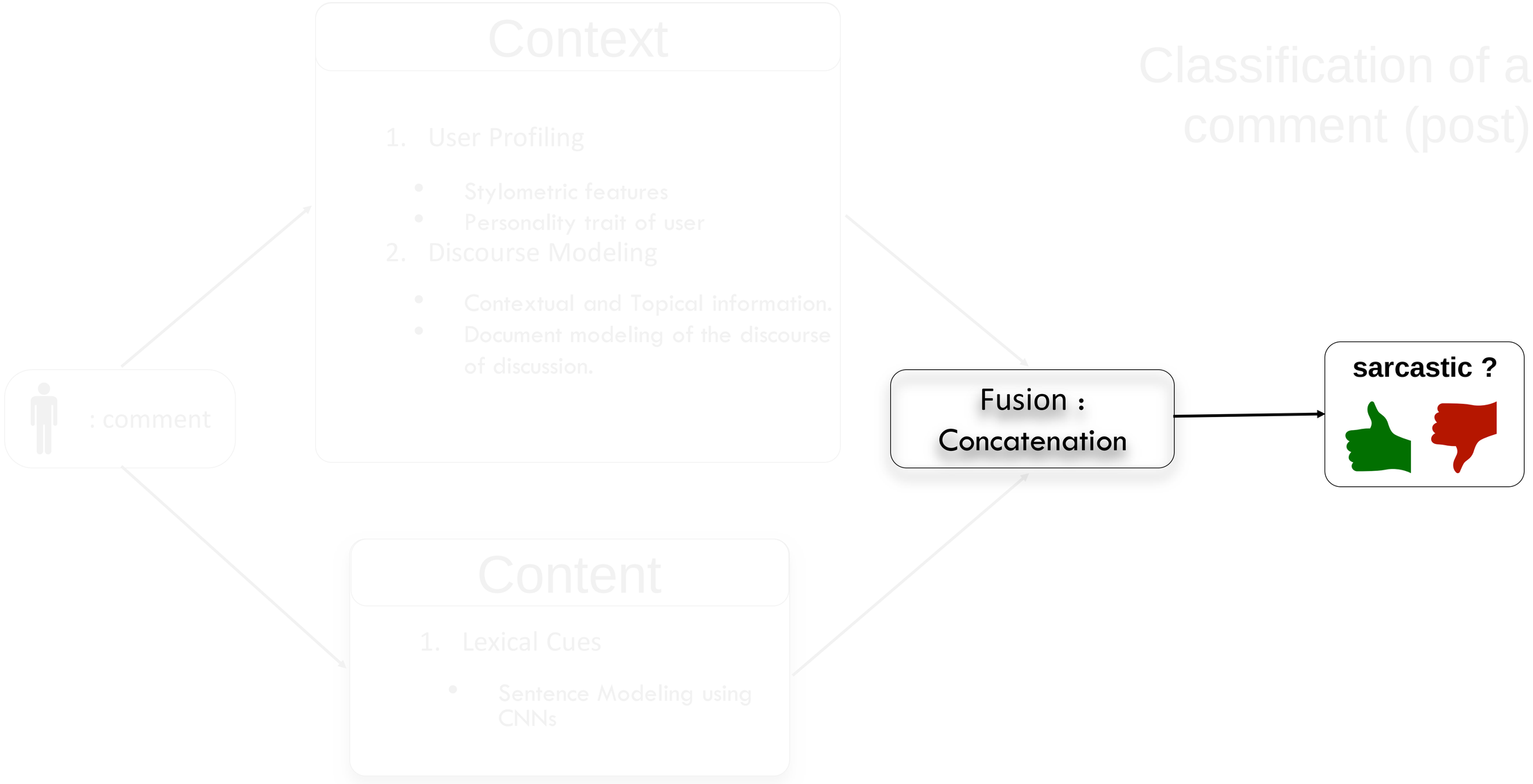

Context
1. User Profiling
• Stylometric features
• Personality trait of user
2. Discourse Modeling
• Contextual and Topical information.
• Document modeling of the discourse of discussion.
Classification of a comment (post)
: comment
Fusion : Concatenation
sarcastic ?
Content
1. Lexical Cues
• Sentence Modeling using CNNs

Content Modeling

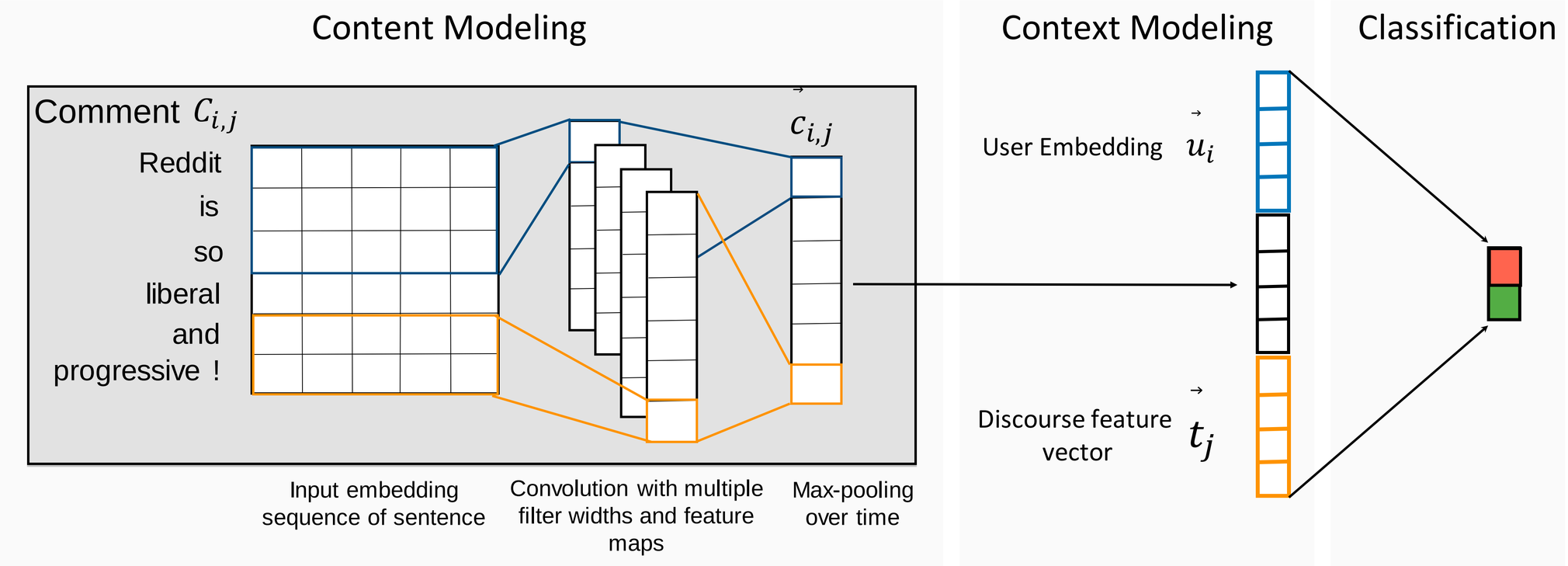
- Sentential representations are extracted from the target comment by using a CNN.
- Captures lexical cues present in sentences that help detect sarcasm.



Overview


Context
1. User Profiling
Stylometric features
Personality trait of user
2. Discourse Modeling
Contextual and Topical information.
Document modeling of the discourse of discussion.
Classification of a comment (post)
: comment
Content
1. Lexical Cues
Sentence Modeling using CNNs
Fusion : Concatenation
sarcastic ?

Overall Architecture



Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

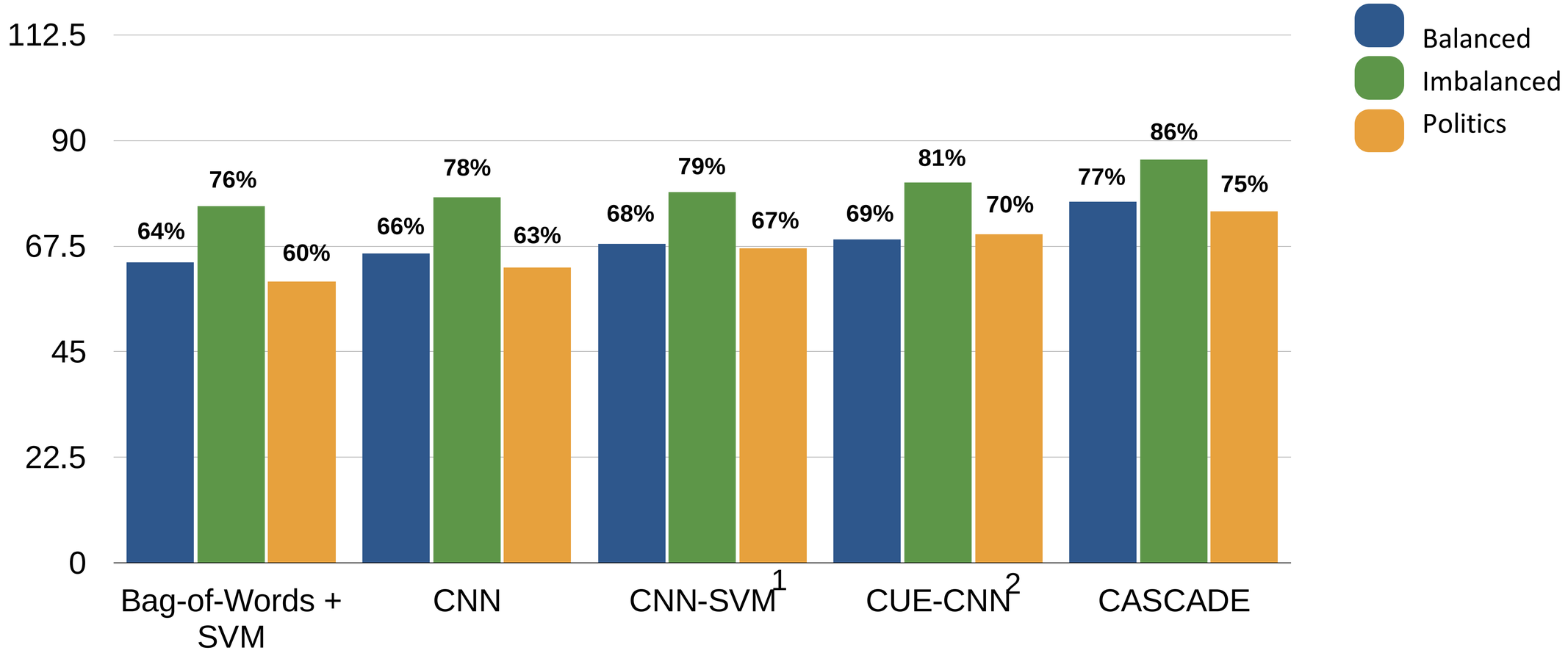
Dataset

- SARC[1]
  - We experiment on the SARC dataset.
  - **Self-Annotated:** Reddit users use the “/s” tag to self-annotate comments as sarcastic
  - **Large-Scale:** We consider approximately 220,000 comments
  - **Versions:**
    - **Balanced**
    - **Imbalanced** :: 20:80 sarcastic/non-sarc split.
    - Single Topic Subset: **Politics**

source - reddit.com

[1] Khodak, Mikhail, Nikunj Saunshi, and Kiran Vodrahalli. "A large self-annotated corpus for sarcasm."

Baseline Comparison



[1] Poria, Soujanya, et al. "A deeper look into sarcastic tweets using deep convolutional neural networks." 2016

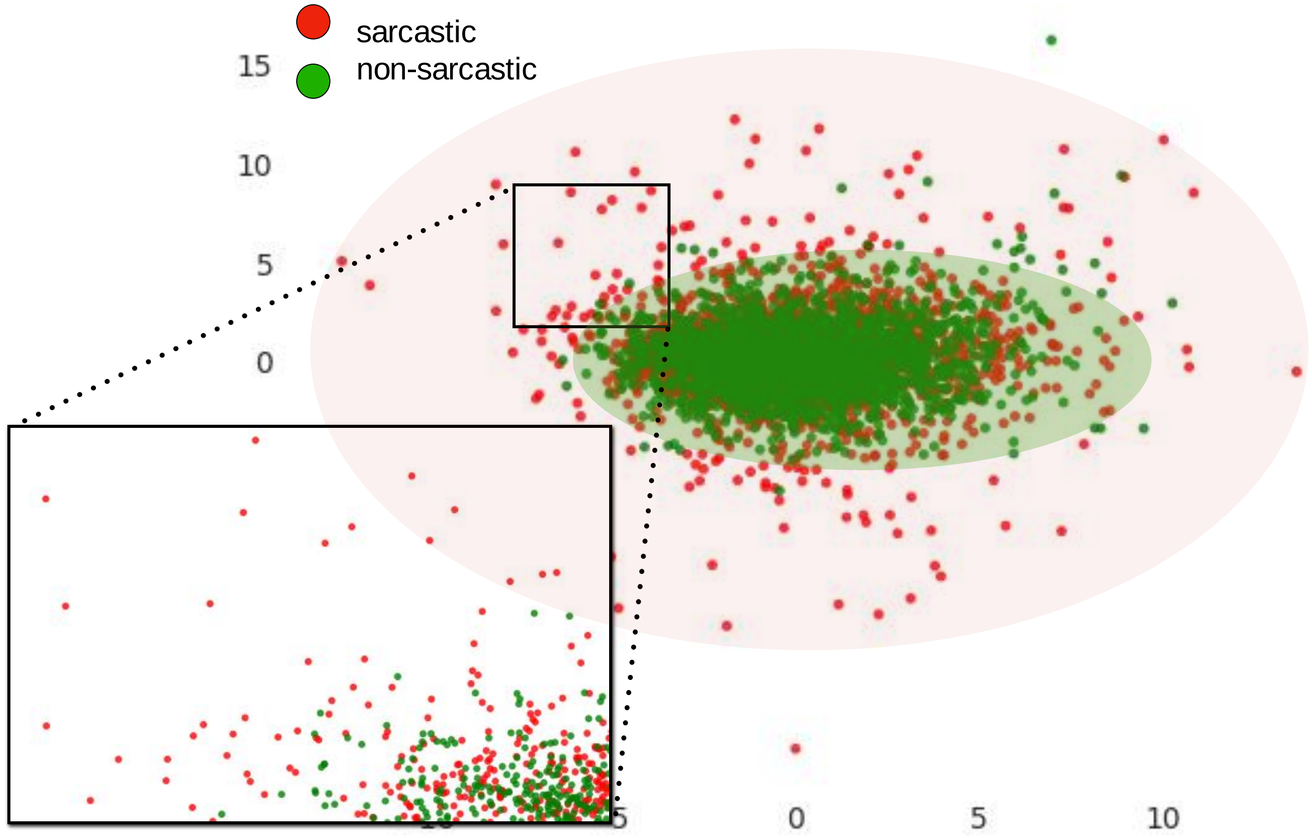
[2] Amir, Silvio, et al. "Modelling context with user embeddings for sarcasm detection in social media." 2016

Ablation Study

	CASCADE			Main				Pol	
	user		dis-	balanced		imbalanced			
	cca	concat.	course	*Acc.*	*F1*	*Acc.*	*F1*	*Acc.*	*F1*
1.	-	-	-	0.65	0.66	0.69	0.78	0.62	0.63
2.	-	-	✓	0.66	0.66	0.68	0.78	0.63	0.66
3.	-	✓	-	0.66	0.66	0.69	0.79	0.62	0.61
4.	-	✓	✓	0.65	0.67	0.71	0.85	0.63	0.66
5.	✓	-	-	0.77	0.76	**0.80**	**0.86**	0.70	0.70
6.	✓	-	✓	**0.78**	**0.77**	0.79	**0.86**	**0.74**	**0.75**

- Only CNN performs worst (need for contextual features).
- Contextual features, especially user embeddings, provide major boost to performance.
- CCA performs better fusion than simpler counterparts such as Concatenation.

User Embedding Analysis



Learnt user-embeddings; t-SNE plot.

- Users with more sarcastic comments are marked red and rest as green.
- Distribution of both user types show a greater variance for sarcastic users. This provides a large non-overlapping space where users are distinctively sarcastic.

Case Studies

CASCADE is able to correctly classify sentences that are implicitly sarcastic, i.e., needs background contextual knowledge.

		Predicted Label	True Label		Predicted Label	True Label
Target Comment	Whew, I feel much better now!			The part where Obama signed it.		
Required Contextual Comment	So all of the US presidents are terrorists for the last 5 years now.			What part of this would be unconstitutional?		

Outline

- Introduction
- Related Work
- CASCADE
- Experimentation
- Conclusion

Conclusion

- We introduced **CASCADE**, a Contextual Sarcasm Detector, which leverages both content and contextual information for the automated classification sarcasm.

- For contextual details, we perform **user profiling** along with **discourse modeling** from comments in discussion threads.

- State-of-the-art performance on a large-scale Reddit corpus: SARC

- User-embeddings and discourse features play important role in improving classification performance.

Thank You!

Find our project on Github:
https://github.com/SenticNet/CASCADE--Contextual-Sarcasm-Detection