

Research Frontiers

Text Mining

Rory Duthie

Centre for Argument Technology

rwduthie@dundee.ac.uk

Outline

- Quick recap
- Part 1: Ethos explanation and annotation
- Break.
- Part 2: Inter-annotator agreement

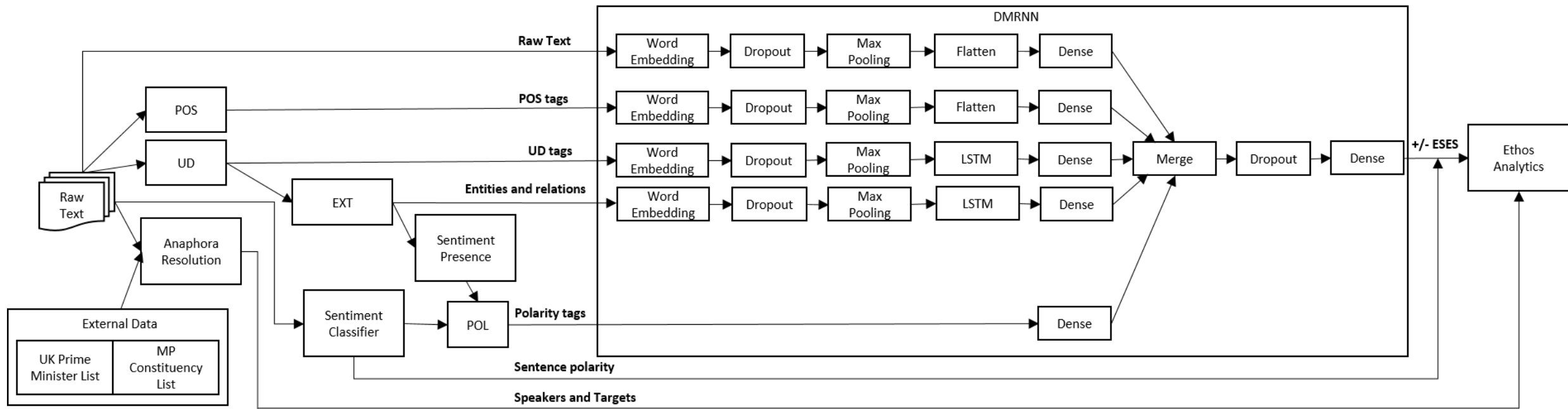
Recap

- Challenges in NLP (e.g. Irony, Slang, Context)
- Sentiment analysis: classifying a unit as Positive, Negative or Neutral.
- Lexicon and ML based
- Evaluation: F1 score (similar to what we need to do for IAA)

Ethos Mining

- Involves extracting ethos from Natural Language.
- Ethos: Character or credibility of a person/organisation.
- Ethos can be supported or attacked.

Ethos Mining Pipeline



Attacks on Ethos



Donald J. Trump ✓

@realDonaldTrump

Follow



"Congressman Schiff omitted and distorted key facts" [@FoxNews](#) So, what else is new. He is a total phony!

4:56 PM - 24 Feb 2018

19,446 Retweets 75,800 Likes



27K

19K

76K

“Congressman Schiff omitted and distorted key facts” @FoxNews So what else is new. He is a total phony!

Break it into sentences.

- “Congressman Schiff omitted and distorted key facts” @FoxNews
- So what else is new.
- He is a total phony!

“Congressman Schiff omitted and distorted key facts” @FoxNews So what else is new. **He is a total phony!**

- Anaphora resolution!
- Determine who the target of this is.
- He = Congressman Schiff.

“Congressman Schiff omitted and distorted key facts” @FoxNews So what else is new. **He is a total phony!**

- Determine source = Donald Trump
- Determine target = Schiff.
- Determine polarity = Negative.

Attack on Ethos



Donald J. Trump  @realDonaldTrump · Feb 24

"Congressman Schiff omitted and distorted key facts" @FoxNews So, what else is new. He is a total phony!



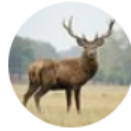
27K



19K



76K



Clever Stag

@CleverStag

Follow

Replying to @realDonaldTrump @FoxNews

Fox News... Is that why you're all ginned up on hate again today? 🤔🤔🤔🤔

You're more gullible than most of your diploables.

One thing you do have in common with Fox is both of you are seria liars.

Fox news is that why you're all ginned up on hate again today?

You're more gullible than most of your diplorables.

One thing you do have in common with Fox is both of you are serial liars.

Fox news is that why you're all ginned up on hate again today?

You're more gullible than most of your diplorables.

One thing you do have in common with Fox is both of you are serial liars.

- Watch for spelling mistakes.
- Context needed for anaphora resolution.
- You = Donald Trump

Fox news is that why you're all ginned up on hate again today?

You're more gullible than most of your diplorables.

One thing you do have in common with Fox is both of you are serial liars.

- Source = Clever Stag
- Target = Trump.
- Polarity = Negative and Negative

Supports of Ethos



Donald J. Trump  @realDonaldTrump · Feb 24

"Congressman Schiff omitted and distorted key facts" @FoxNews So, what else is new. He is a total phony!



27K



19K



76K



Katrina West

@Kwestartist

Follow

Replying to @realDonaldTrump @FoxNews

You are so right Mr. President. If only we don't have to put up with him and his ilk we could so #MAGA sooner. We patriots will continue to stand in spite of the left's constant attacks.

5:00 PM - 24 Feb 2018

You are so right Mr. President. If only we don't have to put up with him and his ilk we could do #MAGA sooner. We patriots will continue to stand in spite of the left's constants attacks.

You are so right Mr. President. If only we don't have to put up with him and his ilk we could do #MAGA sooner. We patriots will continue to stand in spite of the left's constants attacks.

- Support of Mr. President
- Attack on Schiff.
- Attack on left: too large a group to determine who it refers to.

You are so right Mr. President. If only we don't have to put up with him and his ilk we could do #MAGA sooner. We patriots will continue to stand in spite of the left's constants attacks.

- Source = Katrina West
- Target = Trump and Schiff.
- Polarity = Positive and Negative.

Ethos is Applied Sentiment Analysis

- Sentiment analysis:
 - Target = Person / Organisation
 - Source = Person / Organisation
 - Polarity = Positive or Negative

Exercise 1

- **Exercise 1 (30 minutes): Ethos Annotation**
- In groups (2 – 4 people) collect a small dataset (~50 sentences) of tweets or comments on newspapers or social media.
- Annotate each sentence with:
 - Source – who said the sentence
 - Target – who is the target of the ethotic sentence
 - Ethos – Positive (supporting someone), Negative (attacking someone), Neutral (No ethotic supports or attacks)
- Split the annotation task evenly between your group.

Corpus Format

	A	B	C	D	E
1	Annotator	Tweets	Source	Target	Polarity
2	Rory	I will be making a major statement from the @WhiteHouse upon my return to D.C. Time and date to	TRUMP		Neutral
3	Rory	Just arrived at #ASEAN50 in the Philippines for my final stop with World Leaders.	TRUMP		Neutral
4	Rory	Will lead to FAIR TRADE DEALS, unâ€¦! https://t.co/ExRBdQnJru After my tour of Asia, all Countries c	TRUMP		Neutral
5	Rory	The United States hâ€¦! https://t.co/MoMNa1MUI3 Great to see @RandPaul looking well and back o	TRUMP	PAUL	Positive
6	Rory	He will help us with TAX CUTS and REFORM!	TRUMP	PAUL	Positive
7	Rory	Excited to be heading home to see the House pass a GREAT Tax Bill with the middle class getting big	TRUMP		Neutral
8	Rory	#MakeAmericaGreatAgainðŸ†ºðŸ†º, I am proud of the Rep. House & Senate for working so	TRUMP	REPUBLICANS	Positive
9	Rory	Weâ€™re getting close!	TRUMP		Neutral
10	Rory	Now, hoâ€¦! https://t.co/bPmeYwoMhU Happy to announce, I am nominating Alex Azar to be the ne	TRUMP	AZAR	Positive
11	Rory	He will be a star for better healthcare and lower drug prices!	TRUMP	AZAR	Positive
12	Rory	Will be leaving the Philippines tomorrow after many days of constant mtgs & work in ord	TRUMP		Neutral
13	Rory	My promises are rapidly being fulfilled.	TRUMP	TRUMP	Positive
14	Rory	Great day of bilateral meetings at #ASEANSummit on trade, which we are turning around to be grea	TRUMP	QUANG	Positive
15	Rory	Heading to the #ASEANSummit 50th Anniv Galâ€¦! https://t.co/s5gIKMqt50 Just landed in the Philip	TRUMP		Neutral
16	Rory	Why would Kim Jong-un insult me by calling me "old," when I would NEVER call him &c	TRUMP	Jong-Un	Negative
17	Rory	Does the Fake News Media remember when Crooked Hillary Clinton, as Secretary of State, was beg	TRUMP	CLINTON	Negative

Break

Inter-annotator agreement

- Calculate the agreement between two annotators.
- Determines how reliable the annotation is.

Percentage Agreement

- Number of times both annotators agree.
- E.g. Annotating sentences as positive or negative.
- Annotator 1: Positive, Negative, Positive, Negative, Negative.
- Annotator 2: Negative, Negative, Negative, Negative, Negative.
- Percentage agreement = $3 / 5 = 60\%$

Cohen's Kappa

- Annotator 1: Positive, Negative, Positive, Negative, Negative.
- Annotator 2: Negative, Negative, Negative, Negative, Negative.
- Construct a confusion matrix:

		Annotator 2	
		Positive	Negative
Annotator 1	Positive	0	2
	Negative	0	3

Cohen's Kappa

		Annotator 2	
		Positive	Negative
Annotator 1	Positive	0	2
	Negative	0	3

- Observed Agreement = percentage agreement: $3/5 = 0.6$
- Expected Agreement =
(Annotator 1 positive total) x (Annotator 2 positive total) / total examples
 $(2 \times 0) / 5 = 0$
(Annotator 1 negative total) x (Annotator 2 negative total) / total examples
 $(3 \times 5) / 5 = 3$

Expected Agreement = $(0 + 3) / 5 = 0.6$

Cohen's Kappa

- $K = (\text{Observed Agreement} - \text{Expected Agreement}) / (1 - \text{Expected agreement})$
- $K = (0.6 - 0.6) / (1 - 0.6)$
- $K = 0 / 0.4 = 0$

Cohen's Kappa

- Takes into account chance agreement
- Annotator 2 annotates everything as negative
- Any agreement was completely by chance

Exercise 2

- **Exercise 2 (30 minutes): Calculate Inter-annotator agreement**
- Pair off in each group. Annotate a subset (~50%) of your partners data. Only the data annotated as positive and negative NOT neutral.
- Calculate percentage agreement and Cohen's kappa manually.
- Construct a confusion matrix for Cohen's kappa (in excel, word or anything you can send digitally).
- Calculate both percentage agreement and Cohen's kappa (in excel, word or anything you can send digitally).
- (Optional) Back in groups calculate Kappa automatically using nltk.
- **Resources:**
 - http://www.nltk.org/_modules/nltk/metrics/agreement.html

Summary

- Looked at ethos
- Created a corpus from manual annotation
- Looked at Inter-annotator agreement