Cross-Target Stance Classification with Self-Attention Networks

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks

CSIRO Data61
July 2018

www.data61.csiro.au

ACL 2018

Stance Classification in Tweets

- Automatically identify users' positions on a pre-chosen target of interest (e.g., public issues) from text

(**Input**)
(**Input**)

Target (given): Climate Change is Real Concern
Tweet (given): We need to protect our islands and stop the destruction of coral reef.

(**Output**)

Stance label (to be predicted): *Favour*

Cross-Target Stance Classification

- Generalise user stance on **unseen** targets

Target: A mining project in Australia (**Destination**)
Tweet: Environmentalists warn the $16 billion coal facility will damage the Great Barrier Reef.
Stance: **???**

⭐ Destination target

unlabelled

Apply classifiers trained on a source target to the destination target

Target: Climate Change is Real Concern (**Source**)
Tweet: We need to protect our islands and stop the destruction of coral reef.
Stance: *Favour*

Source target

labelled

Our Approach: Basic Idea

- For targets both related to a common domain, stance generalisation is possible via domain-specific information that reflects users' major concerns

Tweet: Environmentalists warn the $16 billion coal facility will <u>damage</u> the <u>Great Barrier Reef</u>.
Target: *A mining project in Australia*
Stance: *Favour*

Destination target

Tweet: We need to protect our islands and stop the <u>destruction</u> of <u>coral reef</u>.
Target: *Climate Change is Real Concern*
Stance: *Favour*

Source target

Domain aspects: e.g., reef, destruction/damage
(Implicit) Domain: environment

Extraction of Domain Aspects

- Key properties of domain aspects
  - They tend to be mentioned by multiple users in a corpus
  - They tend to carry the core meaning of a stance-bearing tweet

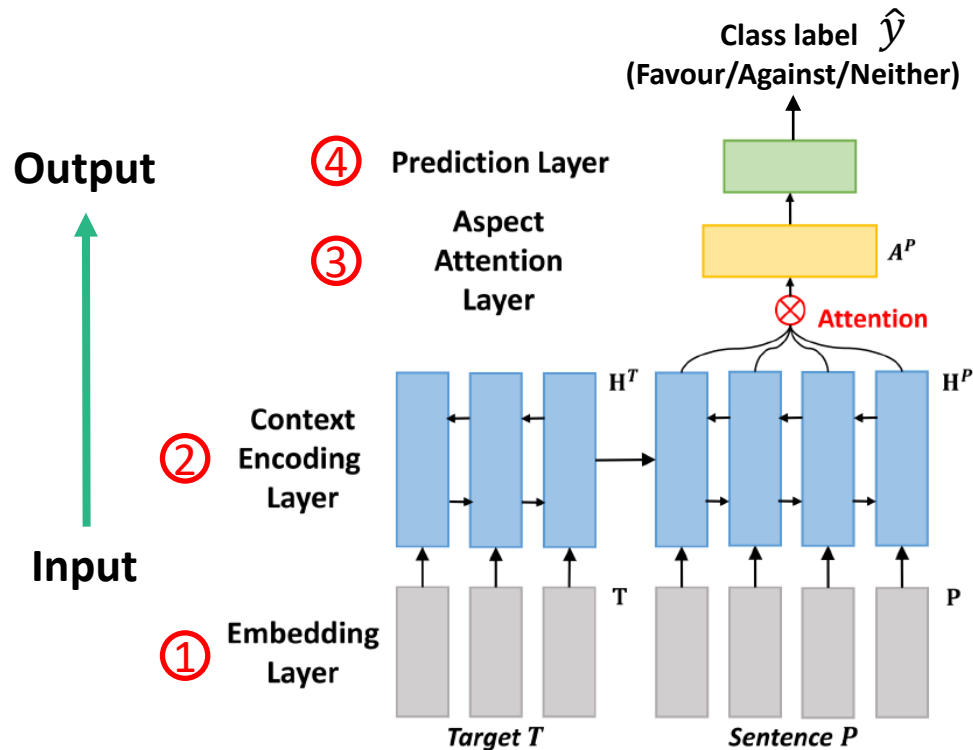 ➢ *In our project dataset, 3776 our of 41805 tweets mentioned the aspect "reef"*

  "why fund Adani #Coal Mine and destroy our Reef when there's so much sun in Queensland?"

  "And your massive polluting Carmichael mine will do its bit to kill Australia's great barrier reef?"

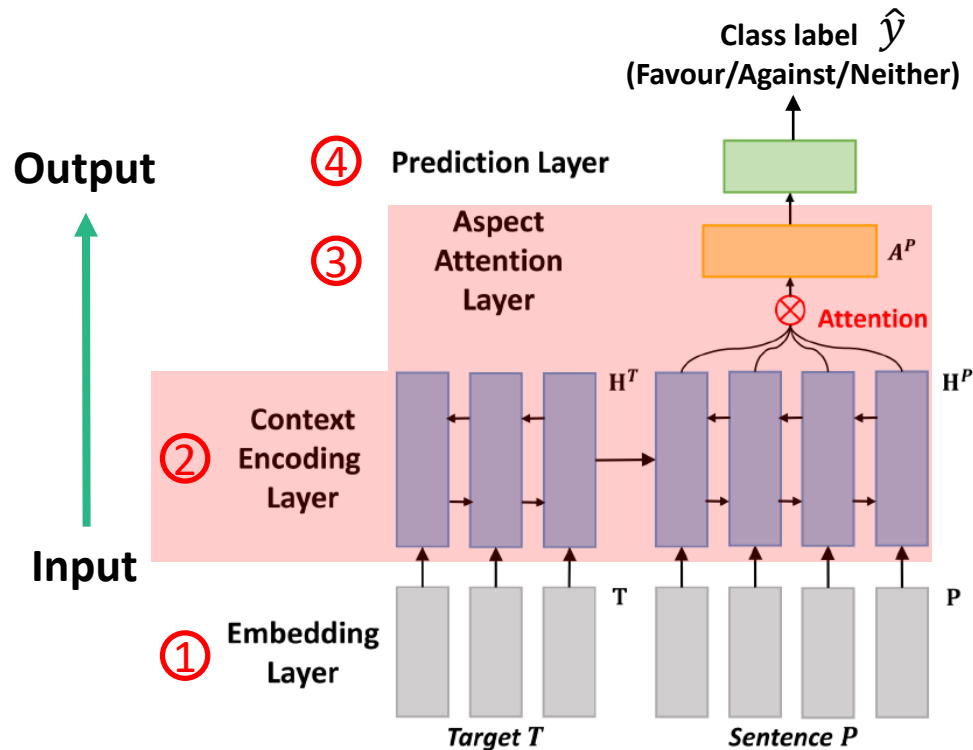 "And thousands of jobs will be lost in reef tourism when Adani goes ahead."

  "The coral reef crisis is actually a crisis of governance."

A Self-Attention Neural Model: Overview

Class label $\hat{y}$
(Favour/Against/Neither)

Output

④ Prediction Layer

③ Aspect Attention Layer A^P

⊗ Attention

Context Encoding Layer H^T H^P

②

Input

① Embedding Layer

T P

Target T *Sentence P*

A Self-Attention Neural Model: Overview

Class label $\hat{y}$
(Favour/Against/Neither)

④ Prediction Layer

③ Aspect Attention Layer

② Context Encoding Layer

① Embedding Layer

Output

Input

H^T H^P A^P

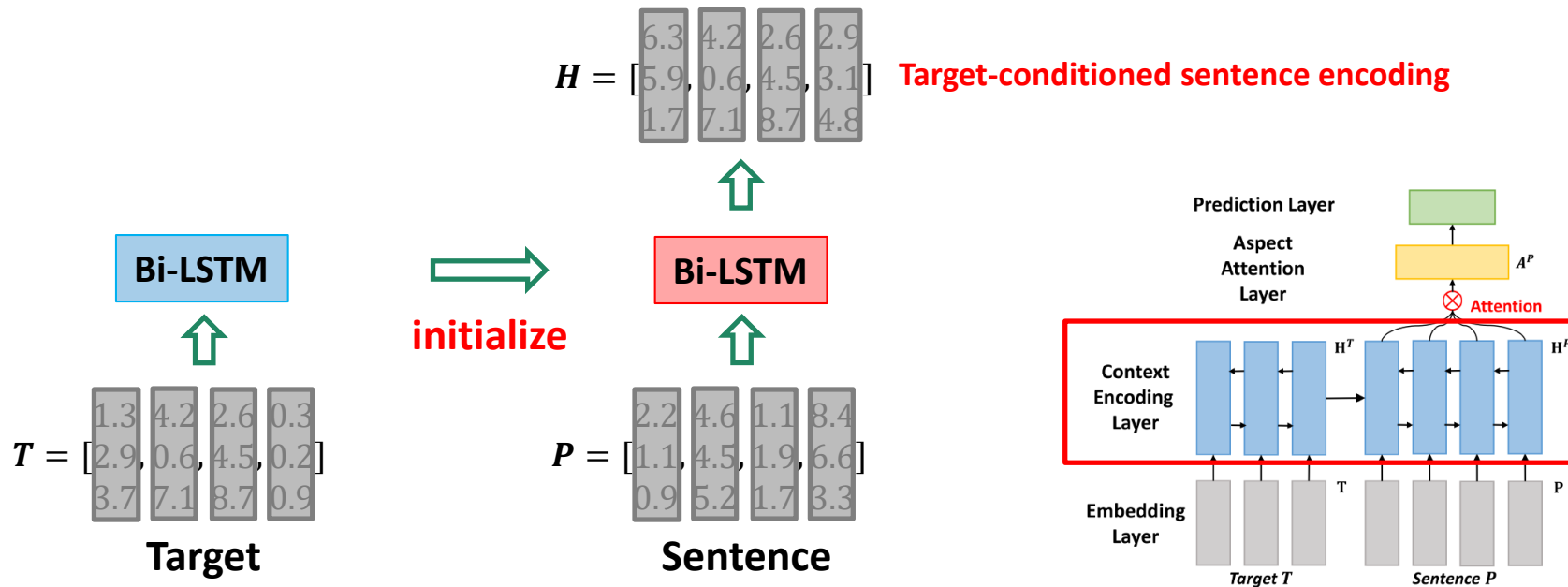Attention

T P

Target T Sentence P

Aspect-aware & target-dependent sentence encoding

Preliminary modelling
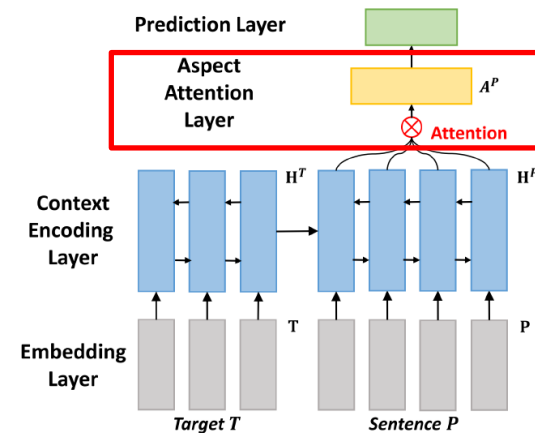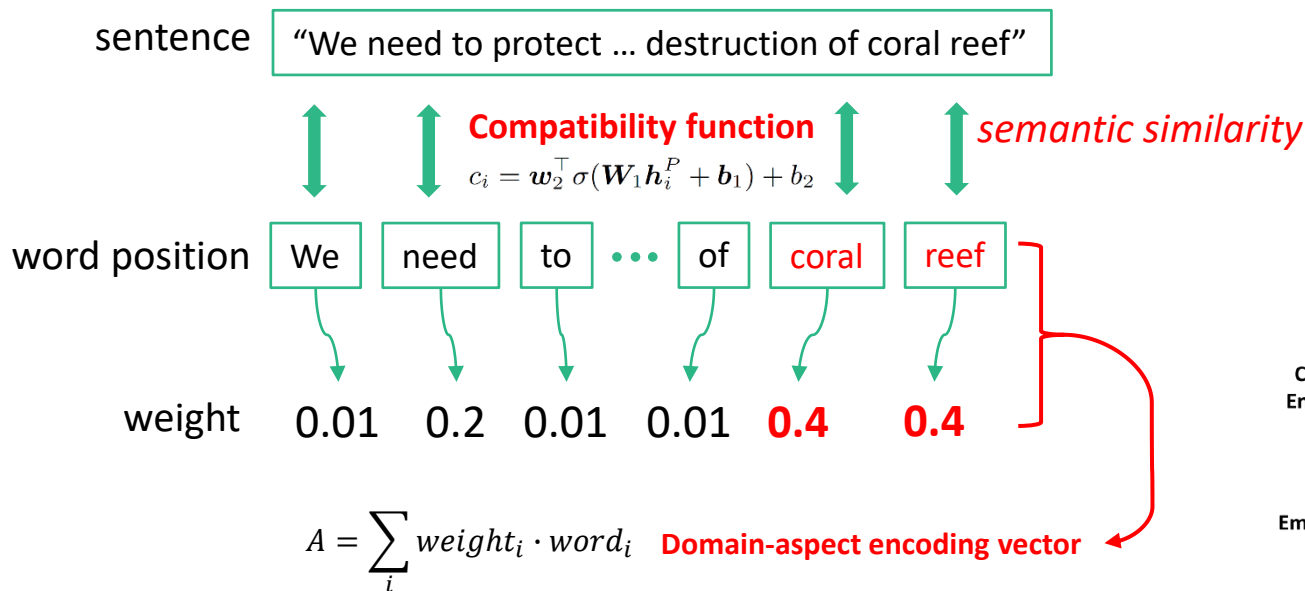- *The simplest case: source-side-only model*

Context Encoding Layer

- Conditional sentence encoding [Augenstein et al., 2016]: Learn a **target-dependent** representation for the sentence

$$H = \begin{bmatrix} 6.3 \\ 5.9 \\ 1.7 \end{bmatrix}, \begin{bmatrix} 4.2 \\ 0.6 \\ 7.1 \end{bmatrix}, \begin{bmatrix} 2.6 \\ 4.5 \\ 8.7 \end{bmatrix}, \begin{bmatrix} 2.9 \\ 3.1 \\ 4.8 \end{bmatrix}$$

Target-conditioned sentence encoding

Bi-LSTM $\Longrightarrow$ **initialize** $\Rightarrow$ **Bi-LSTM**

$$T = \begin{bmatrix} 1.3 \\ 2.9 \\ 3.7 \end{bmatrix}, \begin{bmatrix} 4.2 \\ 0.6 \\ 7.1 \end{bmatrix}, \begin{bmatrix} 2.6 \\ 4.5 \\ 8.7 \end{bmatrix}, \begin{bmatrix} 0.3 \\ 0.2 \\ 0.9 \end{bmatrix}$$

Target

$$P = \begin{bmatrix} 2.2 \\ 1.1 \\ 0.9 \end{bmatrix}, \begin{bmatrix} 4.6 \\ 4.5 \\ 5.2 \end{bmatrix}, \begin{bmatrix} 1.1 \\ 1.9 \\ 1.7 \end{bmatrix}, \begin{bmatrix} 8.4 \\ 6.6 \\ 3.3 \end{bmatrix}$$

Sentence

Prediction Layer

Aspect Attention Layer A^P

Attention

Context Encoding Layer H^T H^P

Embedding Layer

Target T T P Sentence P

Aspect Attention Layer

- Extract domain aspect words using **self-attention weighting**
- Attention weights on word positions : the importance in carrying the sentence meaning

sentence "We need to protect … destruction of coral reef"

Compatibility function   *semantic similarity*

$$c_i = \boldsymbol{w}_2^\top \sigma(\boldsymbol{W}_1 \boldsymbol{h}_i^P + \boldsymbol{b}_1) + b_2$$

word position   We   need   to $\cdots$ of   coral   reef

weight   0.01   0.2   0.01   0.01   **0.4**   **0.4**

$$A = \sum_i weight_i \cdot word_i$$ **Domain-aspect encoding vector**

Experiments

Datasets

- *SemEval 2016 Task 6*: Twitter stance detection

	Target	%Favor	%Against	%Neither	#Total
Climate Change is Concern	CC	59.4	4.6	36.0	564
Feminist Movement	FM	28.2	53.8	18.0	949
Hillary Clinton	HC	16.6	57.4	26.0	984
Legalization of Abortion	LA	17.9	58.3	23.8	933
Donald Trump	DT	20.9	42.3	36.8	707

Domains and targets

1. ***Women's Rights***: Feminist Movement **<>** Legalisation of Abortion
2. ***American Politics***: Hillary Clinton **<>** Donald Trump

Classification performance

DATA 61 | CSIRO

$\Delta \overline{F_1}$=3.0% →

Extracted domain aspects benefit cross-target task more

$\Delta \overline{F_1}$=6.6% →

Better performance on both tasks across almost all targets

FM: *Feminist Movement*
LA: *Legalization of Abortion*
HC: *Hillary Clinton*
DT: *Donald Trump*
CC: *Climate Change is Concern*

Domain	Target	BiLSTM	BiCond	CrossNet
In-target performance (F1-score)				
Women's	FM	51.4(2.6)	53.7(3.9)	55.9(4.6)†
Rights	LA	59.4(6.4)	61.9(6.3)	61.8(4.7)
American	HC	55.5(4.7)	56.0(3.1)	60.0(3.3)†‡
Politics	DT	57.9(6.0)	59.6(5.8)	60.2(5.1)†‡
Cross-target performance (F1-score)				
Women's	FM→LA	40.1(2.3)	40.3(2.2)	44.2(1.3)†‡
Rights	LA→FM	37.9(2.9)	39.2(1.5)	43.1(1.3)†‡
American	HC→DT	43.3(2.4)	44.2(3.2)	46.1(3.7)†
Politics	DT→HC	40.1(2.6)	40.8(2.2)	41.8(3.2)

(Standard deviation in parentheses)
$^\dagger (^\ddagger)$ Improvements over BiLSTM (BiCond) at $p < .05$ with paired t-test

Visualisation of attention

The heatmap of the attention weights assigned to some tweet examples

ID	Target	Tweet
1	FM→LA	Abortion has nothing to do with feminism. Its about the BABYS body, not yours. (A)
2	LA→FM	All humans, male and female, should have equal political, economic and social rights. Equality. (F)
3	HC→DT	Trumps presidential dreams r about as real as KimJonguns unicorns. (A)
4	DT→HC	Maybe a woman should be President. (F)
5	CC→AMP	[N] still will destroy the reef. It is criminal that QLD federal govts are promoting it. (A)
6	CC→AMP	[N], who are trying to change our laws, 'forgot' to tell us about their CEO's environmental disaster! (A)

Women's rights

American politics

Environments

FM: Feminist Movement
LA: Legalization of Abortion
HC: Hillary Clinton
DT: Donald Trump
CC: Climate Change is Concern
AMP: Australian mining project

A: Against
F: Favour

Words central to expressing stances
are highlighted by our model!

Visualisation of attention

The heatmap of the attention weights assigned to some tweet examples

ID	Target	Tweet
1	FM→LA	Abortion has nothing to do with feminism. Its about the BABYS body, not yours. (A)
2	LA→FM	All humans, male and female, should have equal political, economic and social rights. Equality. (F)
3	HC→DT	Trumps presidential dreams r about as real as KimJonguns unicorns. (A)
4	DT→HC	Maybe a woman should be President. (F)
5	CC→AMP	[N] still will destroy the reef. It is criminal that QLD federal govts are promoting it. (A)
6	CC→AMP	[N], who are trying to change our laws, 'forgot' to tell us about their CEO's environmental disaster! (A)

Women's rights
American politics
Environments

FM: Feminist Movement
LA: Legalization of Abortion
HC: Hillary Clinton
DT: Donald Trump
CC: Climate Change is Concern
AMP: Australian mining project

A: Against
F: Favour

Words central to expressing stances
are highlighted by our model!

Visualisation of attention

The heatmap of the attention weights assigned to some tweet examples

	ID	Target	Tweet
Women's rights	1	FM→LA	Abortion has nothing to do with feminism. Its about the BABYS body, not yours. (A)
	2	LA→FM	All humans, male and female, should have equal political, economic and social rights. Equality. (F)
American politics	3	HC→DT	Trumps presidential dreams or about as real as KimJonguns unicorns. (A)
	4	DT→HC	Maybe a woman should be President. (F)
Environments	5	CC→AMP	[N] still will destroy the reef. It is criminal that QLD federal govts are promoting it. (A)
	6	CC→AMP	[N], who are trying to change our laws, 'forgot' to tell us about their CEO's environmental disaster! (A)

FM: Feminist Movement
LA: Legalization of Abortion
HC: Hillary Clinton
DT: Donald Trump
CC: Climate Change is Concern
AMP: Australian mining project

A: Against
F: Favour

Words central to expressing stances
are highlighted by our model!

Visualisation of attention

The heatmap of the attention weights assigned to some tweet examples

ID	Target	Tweet
1	FM→LA	Abortion has nothing to do with feminism. Its about the BABYS body, not yours. (A)
2	LA→FM	All humans, male and female, should have equal political, economic and social rights. Equality. (F)
3	HC→DT	Trumps presidential dreams r about as real as KimJonguns unicorns. (A)
4	DT→HC	Maybe a woman should be President. (F)
5	CC→AMP	[N] still will destroy the reef. It is criminal that QLD federal govts are promoting it. (A)
6	CC→AMP	[N], who are trying to change our laws, 'forgot' to tell us about their CEO's environmental disaster (A)

Women's rights

American politics

Environments

FM: *Feminist Movement*
LA: *Legalization of Abortion*
HC: *Hillary Clinton*
DT: *Donald Trump*
CC: *Climate Change is Concern*
AMP: *Australian mining project*

A: Against
F: Favour

Words central to expressing stances
are highlighted by our model!

Conclusion

- A self-attention model which can attend high-level information about the domain for stance generalisation

- Domain aspect words are useful to determine the user stance

- Future directions
  - Incorporation of target divergence into our modelling.
  - Learning aspects from multiple sources (e.g., environment, community, and economics aspects for "mining projects")