Bringing replication and reproduction together with generalisability in NLP: Three reproduction studies for Target Dependent Sentiment Analysis

Andrew Moore and Paul Rayson

August 21, 2018

School of Computing and Communications, Lancaster University, Lancaster, UK

Document Sentiment Example

'Rude service, medicore food...there are tons of restaurants in NY...stay away from this one' (Pontiki et al., 2015)

Negative

Aspect Based Sentiment Analysis (ABSA) Example

Text
'Rude service, medicore food...there are tons of restaurants in NY...stay away from this one' (Pontiki et al., 2015)

Aspects

1. SERVICE#GENERAL – Negative
2. FOOD#QUALITY – Neutral
3. RESTAURANT#GENERAL – Negative

Target Dependent Sentiment Analysis (TDSA) Example

Text
'Rude service, medicore food...there are tons of restaurants in NY...stay away from this one' (Pontiki et al., 2015)

Targets

1. service – Negative
2. food – Neutral

Generalisability?

1. Domain – Restaurant, Laptop
2. Type – Social Media, Reviews
3. Medium – Written, Spoken
4. Data Set Size
5. Data Set Characteristics – number of targets in a sentence.

Generalisability within TDSA

	Datasets						
Methods	1	2	3	4	5	6	7
Mitchell et al. (2013)			✓				
Kiritchenko et al. (2014)				✓			
Dong et al. (2014)	✓						
Vo et al. (2015)	✓	✓	✓				
Zhang et al. (2015)			✓				
Zhang et al. (2016)	✓	✓	✓				
Tang et al. (2016b)	✓			✓			
Tang et al. (2016a)				✓			
Wang et al. (2016)				✓			
Chen et al. (2017)	✓			✓	✓		
Liu et al. (2017)	✓	✓	✓				
Wang et al. (2017)	✓					✓	
Marrese-Taylor et al. (2017)				✓			✓

1=Dong et al. (2014), 2=Wilson (2008), 3=Mitchell et al. (2013), 4=Pontiki et al. (2014), 5=Chen et al. (2017), 6=Wang et al. (2017), 7=Marrese-Taylor et al. (2017)

Table 1: Methods and Datasets

■ Not Applicable

Generalisability within TDSA

	Datasets						
Methods	1	2	3	4	5	6	7
Mitchell et al. (2013)			✓				
Kiritchenko et al. (2014)				✓			
Dong et al. (2014)	✓						
Vo et al. (2015)	✓	✓	✓				
Zhang et al. (2015)			✓				
Zhang et al. (2016)	✓	✓	✓				
Tang et al. (2016b)	✓			✓			
Tang et al. (2016a)				✓			
Wang et al. (2016)				✓			
Chen et al. (2017)	✓			✓	✓		
Liu et al. (2017)	✓	✓	✓				
Wang et al. (2017)	✓					✓	
Marrese-Taylor et al. (2017)				✓			✓

1=Dong et al. (2014), 2=Wilson (2008), 3=Mitchell et al. (2013), 4=Pontiki et al. (2014), 5=Chen et al. (2017), 6=Wang et al. (2017), 7=Marrese-Taylor et al. (2017)

Table 2: Methods and Datasets

Social Media Reviews News Not Applicable

Generalisability within TDSA

	Datasets						
Methods	1	2	3	4	5	6	7
Mitchell et al. (2013)			✓				
Kiritchenko et al. (2014)				✓			
Dong et al. (2014)	✓						
Vo et al. (2015)	✓	✓	✓				
Zhang et al. (2015)			✓				
Zhang et al. (2016)	✓	✓	✓				
Tang et al. (2016b)	✓			✓			
Tang et al. (2016a)				✓			
Wang et al. (2016)				✓			
Chen et al. (2017)	✓			✓	✓		
Liu et al. (2017)	✓	✓	✓				
Wang et al. (2017)	✓					✓	
Marrese-Taylor et al. (2017)				✓			✓

1=Dong et al. (2014), 2=Wilson (2008), 3=Mitchell et al. (2013), 4=Pontiki et al. (2014), 5=Chen et al. (2017), 6=Wang et al. (2017), 7=Marrese-Taylor et al. (2017)

Table 3: Methods and Datasets

Social Media Reviews News Not Applicable

Generalisability within TDSA

	Datasets						
Methods	1	2	3	4	5	6	7
Mitchell et al. (2013)			✓				
Kiritchenko et al. (2014)				✓			
Dong et al. (2014)	✓						
Vo et al. (2015)	✓	✓	✓				
Zhang et al. (2015)			✓				
Zhang et al. (2016)	✓	✓	✓				
Tang et al. (2016b)	✓			✓			
Tang et al. (2016a)				✓			
Wang et al. (2016)				✓			
Chen et al. (2017)	✓			✓	✓		
Liu et al. (2017)	✓	✓	✓				
Wang et al. (2017)	✓					✓	
Marrese-Taylor et al. (2017)				✓			✓

1=Dong et al. (2014), 2=Wilson (2008), 3=Mitchell et al. (2013), 4=Pontiki et al. (2014), 5=Chen et al. (2017), 6=Wang et al. (2017), 7=Marrese-Taylor et al. (2017)

Table 4: Methods and Datasets

Social Media Reviews News Not Applicable

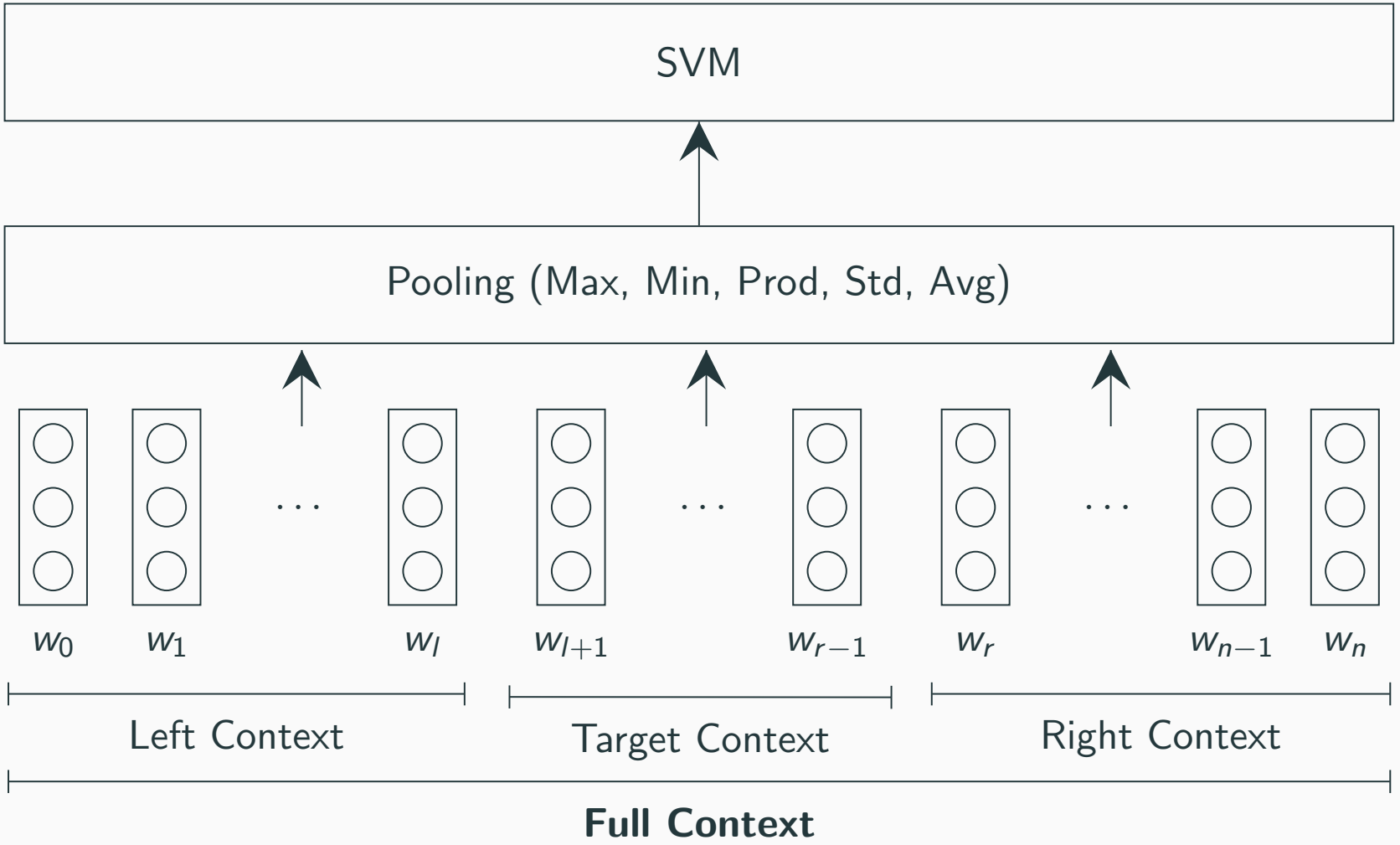
Why Reproduce?

Authors	Code with paper
Wang et al. (2017)	Yes
Tang et al. (2016b)	Unreliable
Vo et al. (2015)	No

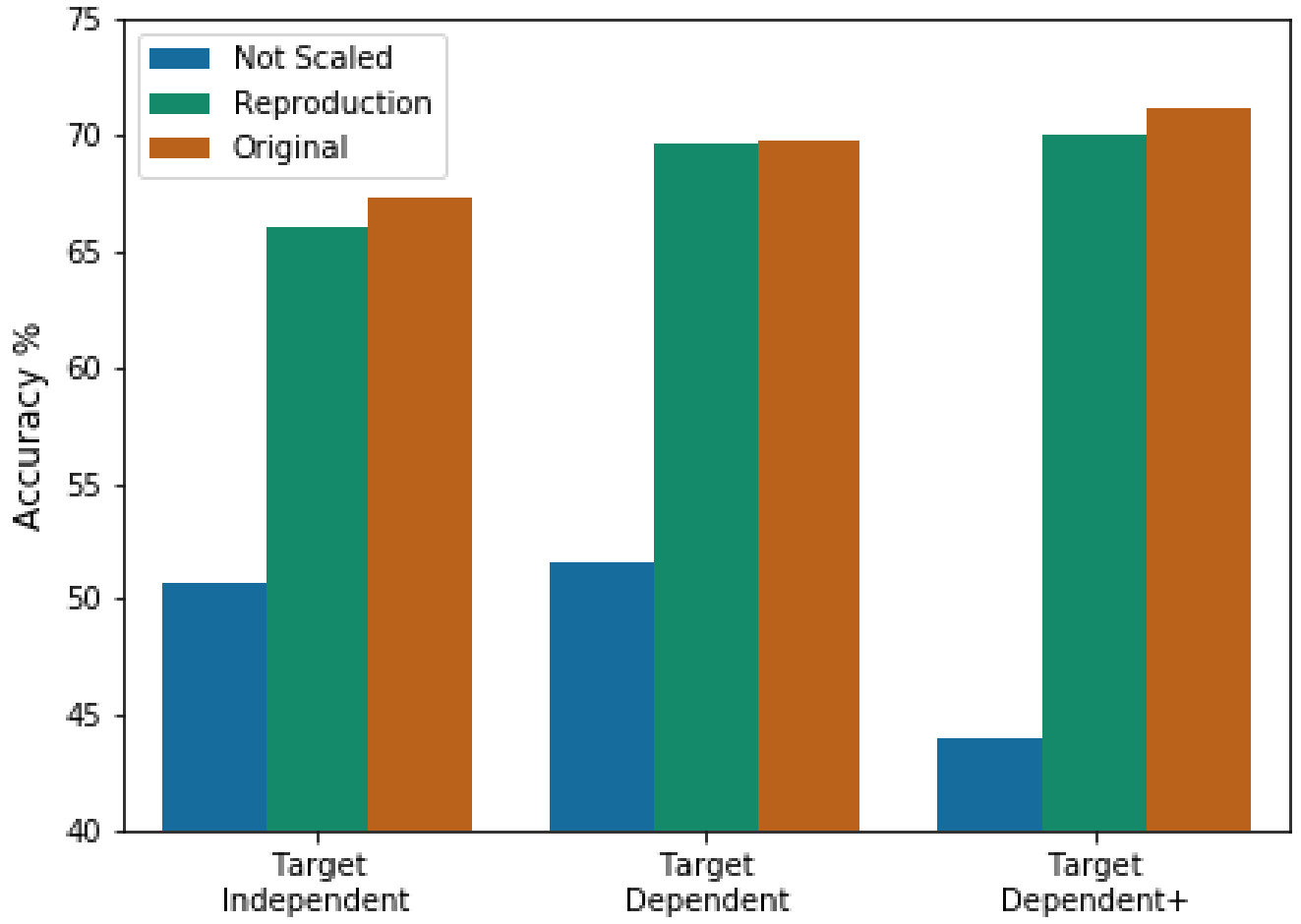
Authors	Restaurant	Laptop
Tang et al. (2016b)	75.63	68.13
Chen et al. (2017)	78.00	71.83
Tay et al. (2017)	69.73	62.38

■ Original ■ Re-used the same code ■ Re-implemented

Vo et al. (2015) Method

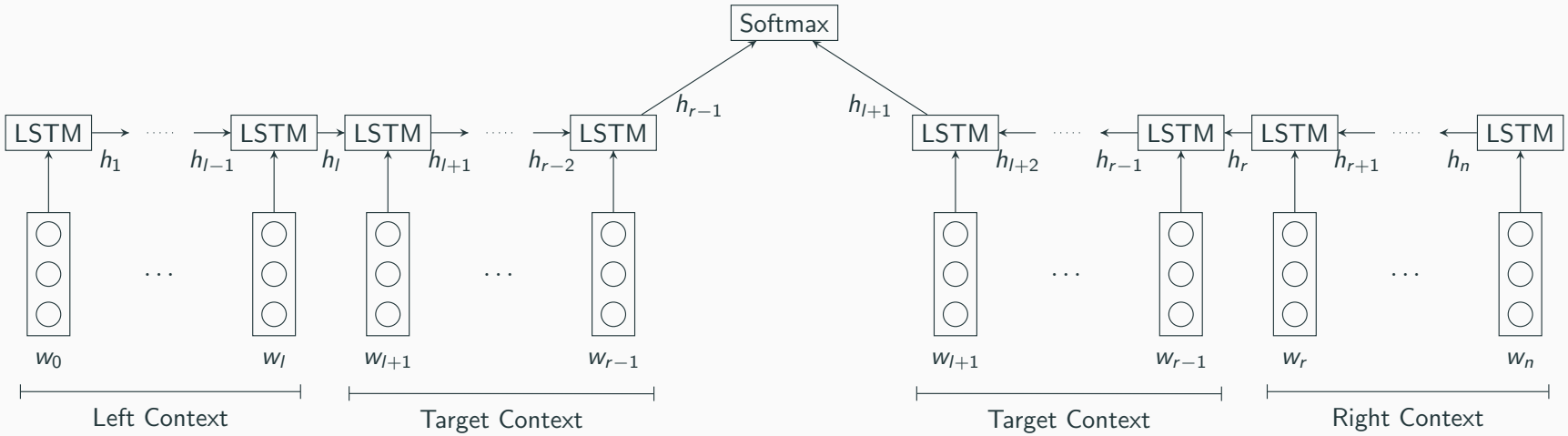


Vo et al. (2015) Reproduction Result

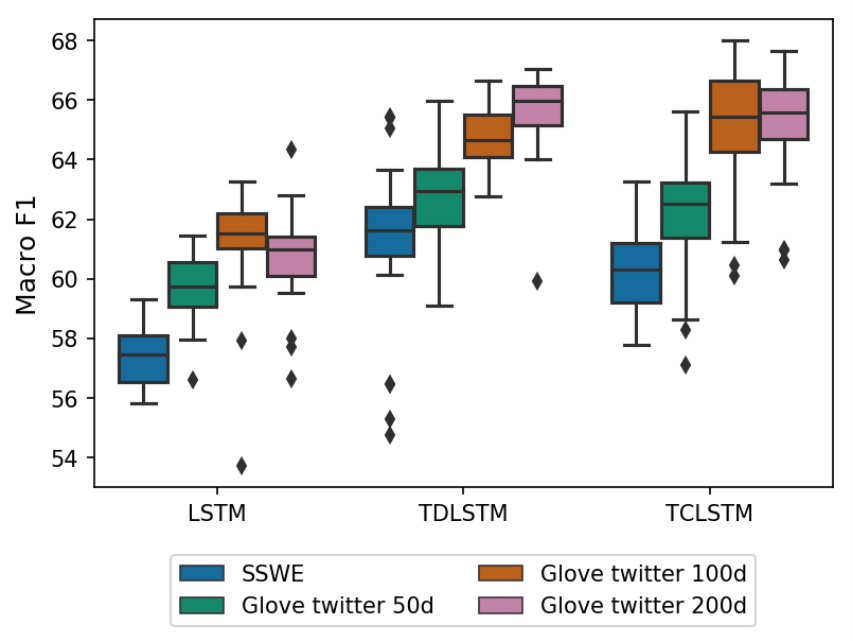


Scaling features is important - 15-25% difference

Tang et al. (2016b) Method



Tang et al. (2016b) Reproduction Result



	Macro F1		
Methods	O	R (Max)	R (Mean)
LSTM	64.70	64.34	60.69
TDLSTM	69.00	67.04	65.63
TCLSTM	69.50	67.66	65.23
O=Original, R=Reproduction			

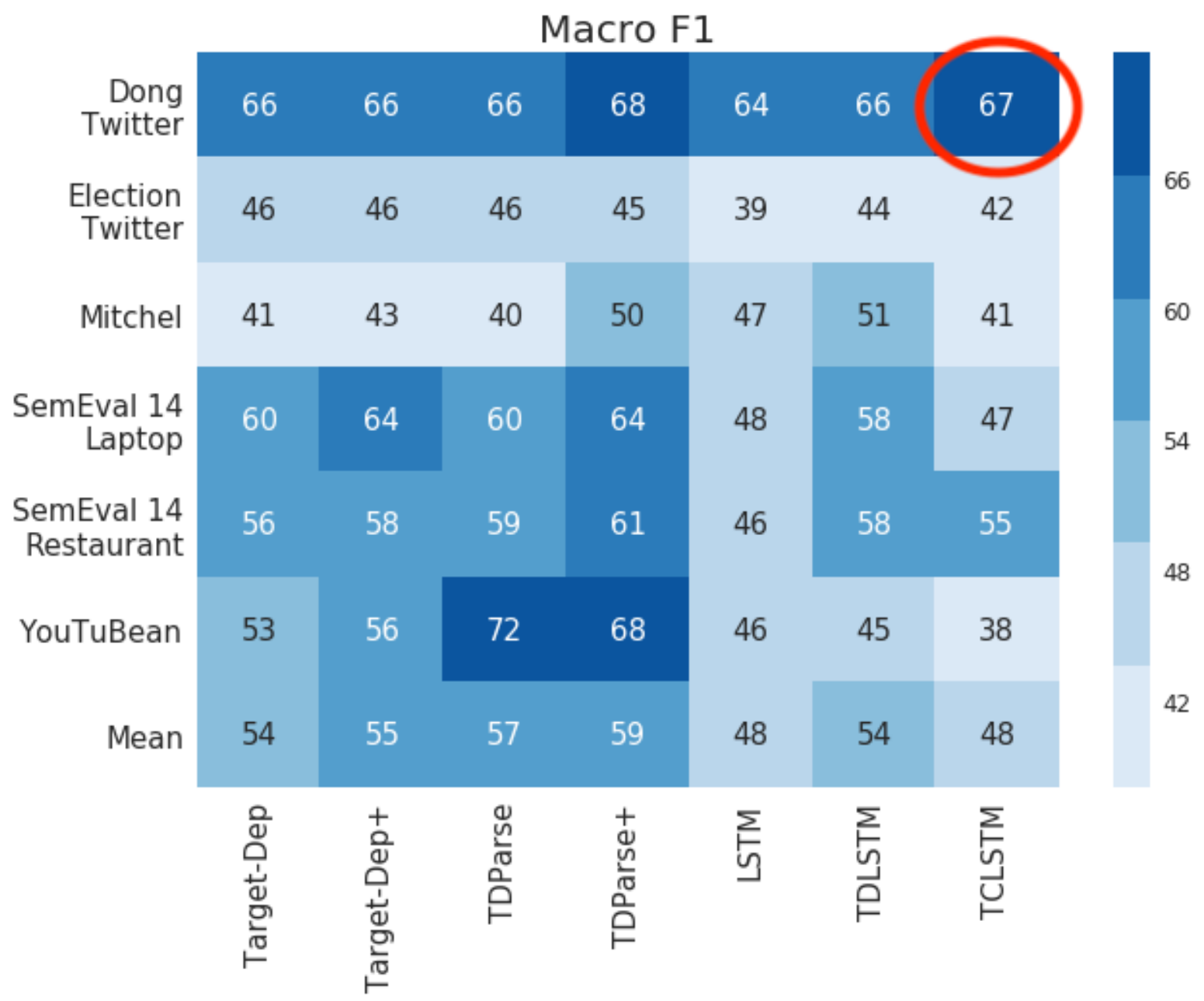
Repeating experiments with different seed values is important. (Reimers et al., 2017)

Mass Evaluation Datasets

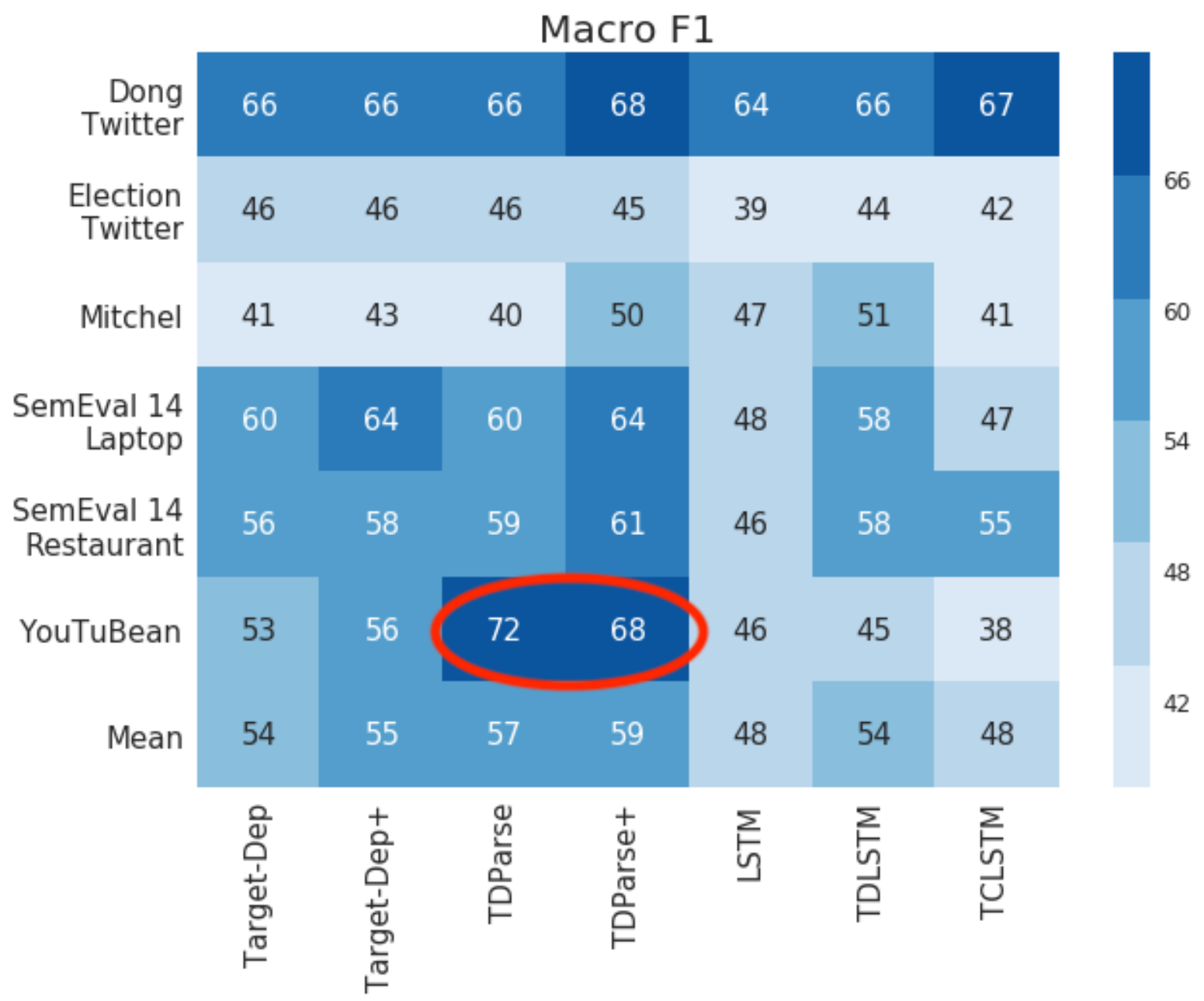
Dataset	Domain	Type	Size	Medium	ATS
SemEval 14 L	L	RE	2951	W	1.58
SemEval 14 R	R	RE	4722	W	1.83
Mitchel	G	S	3288	W	1.22
Dong Twitter	G	S	6940	W	1.00
Election Twitter	P	S	11899	W	2.94
YouTuBean	MP	RE/S	798	SP	2.07

L=Laptop, R=Restaurant, G=General, P=Politics, MP=Mobile Phones, RE=Review, S=Social Media, W=Written, SP=Spoken, ATS=Average Targets per Sentence

Mass Evaluation



Mass Evaluation



Contributions

1. **Generalisability**: First to report results across across three different dataset properties: 1. Domain, 2. Type, 3. Medium.

2. **Reproduction**: Open source TDSA framework with three different models.

Code, documentation, Jupyter notebook examples, and model zoo:
`https://github.com/apmoore1/Bella`

a.moore@lancaster.ac.uk
@apmoore94 and @perayson

References i


Chen, Peng et al. (2017). "Recurrent Attention Network on Memory for Aspect Sentiment Analysis". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, pp. 463–472. URL: http://aclweb.org/anthology/D17-1048.


Dong, Li et al. (2014). "Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, pp. 49–54. URL: http://aclanthology.coli.uni-saarland.de/pdf/P/P14/P14-2009.pdf.

References ii


Kiritchenko, Svetlana et al. (2014). "NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland: Association for Computational Linguistics, pp. 437–442. URL: http://aclanthology.coli.uni-saarland.de/pdf/S/S14/S14-2076.pdf.


Liu, Jiangming and Yue Zhang (2017). "Attention Modeling for Targeted Sentiment". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Valencia, Spain: Association for Computational Linguistics, pp. 572–577. URL: http://aclanthology.coli.uni-saarland.de/pdf/E/E17/E17-2091.pdf.

References iii

- Marrese-Taylor, Edison, Jorge Balazs, and Yutaka Matsuo (2017). *Mining fine-grained opinions on closed captions of YouTube videos with an attention-RNN*. Copenhagen, Denmark. URL: http://aclweb.org/anthology/W17-5213.
- Mitchell, Margaret et al. (2013). "Open Domain Targeted Sentiment". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 1643–1654. URL: http://www.aclweb.org/anthology/D13-1171.
- Pontiki, Maria et al. (2014). "SemEval-2014 Task 4: Aspect Based Sentiment Analysis". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland: Association for Computational Linguistics, pp. 27–35. URL: http://aclanthology.coli.uni-saarland.de/pdf/S/S14/S14-2004.pdf.

References iv

- Pontiki, Maria et al. (2015). "SemEval-2015 Task 12: Aspect Based Sentiment Analysis". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, pp. 486–495. URL: http://www.aclweb.org/anthology/S15-2082.
- Reimers, Nils and Iryna Gurevych (2017). "Reporting Score Distributions Makes a Difference: Performance Study of LSTM-networks for Sequence Tagging". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 338–348.
- Tang, Duyu, Bing Qin, and Ting Liu (2016a). "Aspect Level Sentiment Classification with Deep Memory Network". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, pp. 214–224. URL: http://aclanthology.coli.uni-saarland.de/pdf/D/D16/D16-1021.pdf.

References v

- Tang, Duyu et al. (2016b). "Effective LSTMs for Target-Dependent Sentiment Classification". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka, Japan: The COLING 2016 Organizing Committee, pp. 3298–3307. URL: http://aclanthology.coli.uni-saarland.de/pdf/C/C16/C16-1311.pdf.
- Tay, Yi, Anh Tuan Luu, and Siu Cheung Hui (2017). "Learning to Attend via Word-Aspect Associative Fusion for Aspect-based Sentiment Analysis". In: *arXiv preprint arXiv:1712.05403*.
- Vo, Duy-Tin and Yue Zhang (2015). "Target-Dependent Twitter Sentiment Classification with Rich Automatic Features.". In: *IJCAI*, pp. 1347–1353.

References vi


Wang, Bo et al. (2017). "TDParse: Multi-target-specific sentiment recognition on Twitter". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, pp. 483–493. URL: http://aclanthology.coli.uni-saarland.de/pdf/E/E17/E17-1046.pdf.


Wang, Yequan et al. (2016). "Attention-based LSTM for Aspect-level Sentiment Classification". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, pp. 606–615. URL: http://www.aclweb.org/anthology/D16-1058.


Wilson, Theresa Ann (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. University of Pittsburgh.


Zhang, Meishan, Yue Zhang, and Duy Tin Vo (2015). "Neural Networks for Open Domain Targeted Sentiment". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, pp. 612–621. URL: http://www.aclweb.org/anthology/D15-1073.


– (2016). "Gated Neural Networks for Targeted Sentiment Analysis.". In: *AAAI*, pp. 3087–3093.