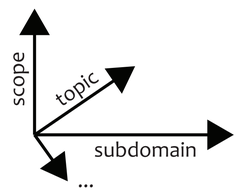
On the impact of cross-domain edge detection in biomedical event extraction

Alan Ramponi[1,2,3], Barbara Plank[2], Rosario Lombardo[3]

1 Department of Information Engineering and Computer Science, University of Trento, Italy
2 Department of Computer Science, IT University of Copenhagen, Denmark
3 Fondazione The Microsoft Research – University of Trento Centre for Computational and Systems Biology, Italy

Motivation



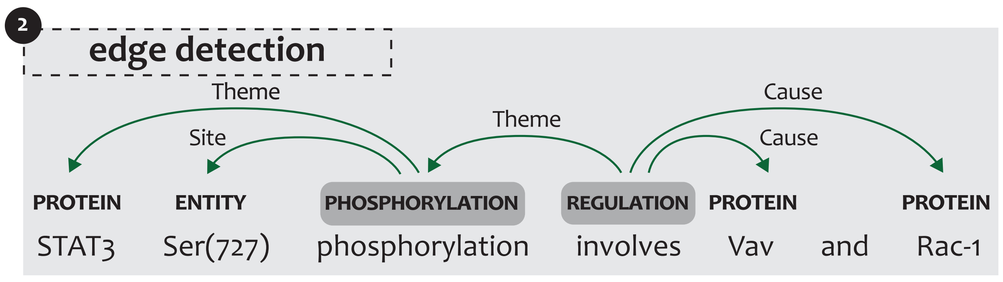
- Biomedical texts cover a diverse set of **language aspects** (≈ *domains*) [1]
- Event extraction systems for **real-world** applications
  - From in-domain to **out-of-domain evaluation**
  - Means for **evaluation of single stages** of the task

The (sub)-task of event extraction

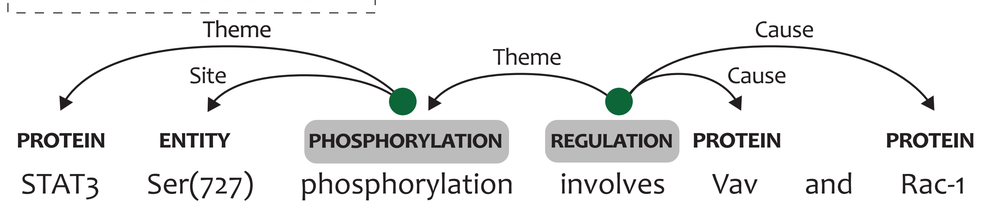
1 trigger detection

PROTEIN	ENTITY	PHOSPHORYLATION	REGULATION	PROTEIN		PROTEIN
STAT3	Ser(727)	phosphorylation	involves	Vav	and	Rac-1

2 edge detection

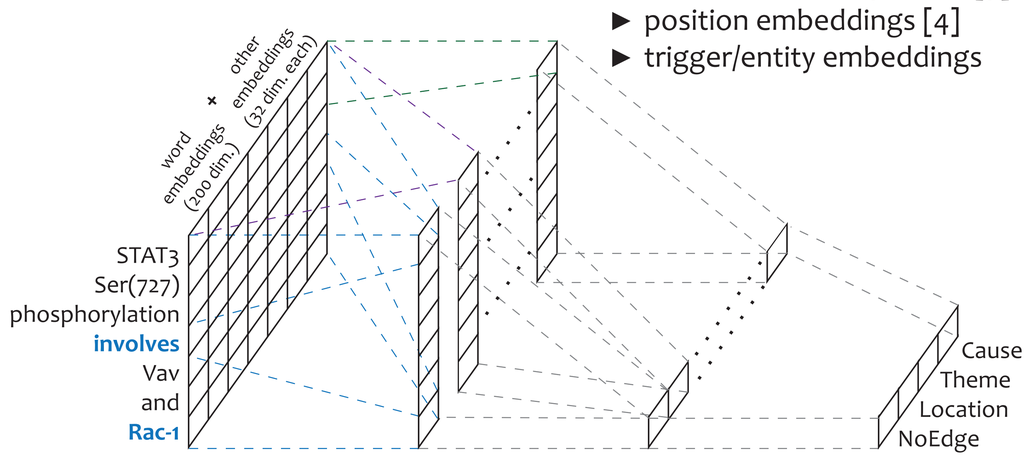


3 event construction



Methods

- biomed word embeddings [2]
- biomed POS & dep. trees [3]
- position embeddings [4]
- trigger/entity embeddings



1 encoding — 2 conv + max pooling — 3 fully connected

Tuned under 5-fold cross-validation:
- **Stratified**: account for class imbalance
- **Group**: avoid sentences to fall into different splits

Corpora

Gold standard data and *some* fuzzy language aspects

	scope	subdomain	corpus topic
GE11	full-texts	molecular biology	transcription factors in human blood cells
ID11	full-texts	infectious diseases	two-component regulatory systems
EPI11	abstracts	epigenetics	epigenetic change and modifications
PC13	abstracts	molecular biology	pathways in BioModels and PantherDB
MLEE	abstracts	physiology	angiogenesis (development of blood vessels)

Experiments and outlook

Cross-domain performance of the **in-domain baseline** for edge detection (**F1$_M$**: macro F1, **F1$_m$**: micro F1)

target →	GE11		ID11		EPI11		PC13		MLEE		*Average drop*	
source ↓	$F\,1_M$	$F\,1_m$	$F\,1_M$	$F\,1_m$	$F\,1_M$	$F\,1_m$	$F\,1_M$	$F\,1_m$	$F\,1_M$	$F\,1_m$	$F\,1_M$	$F\,1_m$
GE11	**79.84**	**88.21**	75.74	86.88	76.81	83.89	52.90	84.56	49.61	85.47	-16.08	-3.01
ID11	60.66	79.97	**69.49**	**89.52**	51.87	71.04	50.01	81.10	50.12	83.35	-16.33	-10.66
EPI11	61.71	71.39	55.45	74.12	**79.62**	**86.15**	46.31	72.45	37.91	64.37	-29.28	-15.57
PC13	56.07	83.44	54.33	85.95	55.10	73.17	**77.28**	**87.12**	55.11	85.78	-22.18	-5.04
MLEE	56.21	81.67	56.08	87.34	54.54	71.7	61.94	85.09	**76.23**	**90.37**	-19.04	-8.92
Average											**-20.58**	**-8.64**

Results highlight the **need for addressing the domain variance in biomedical texts**. We plan to:
- Use **domain adaptation** (DA) methods to improve the performance of event extraction systems *into the wild*
- Release **data for edge detection** to the community to encourage research efforts in DA for single stages

References

[1] Plank, B. (2016). "What to do about non-standard (or non-canonical) language in NLP." KONVENS '16.
[2] Pyysalo, S., et al. (2013). "Distributional semantics resources for biomedical text processing." LBM '13.
[3] Neumann, M., et al. (2019). "ScispaCy: fast and robust models for biomedical NLP." arXiv:1902.07669.
[4] Zeng, D., et al. (2014). "Relation classification via convolutional deep neural network." COLING '14.

EurNLP 2019, October 11th, 2019
@Facebook London, 1 Rathbone Square
Fitzrovia, London W1T 1FB, United Kingdom