

Supplementary Material:

Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters

Evgeny Kim and Roman Klinger
 Institut für Maschinelle Sprachverarbeitung
 University of Stuttgart
 Pfaffenwaldring 5b, 70569 Stuttgart, Germany
 {evgeny.kim, roman.klinger}@ims.uni-stuttgart.de

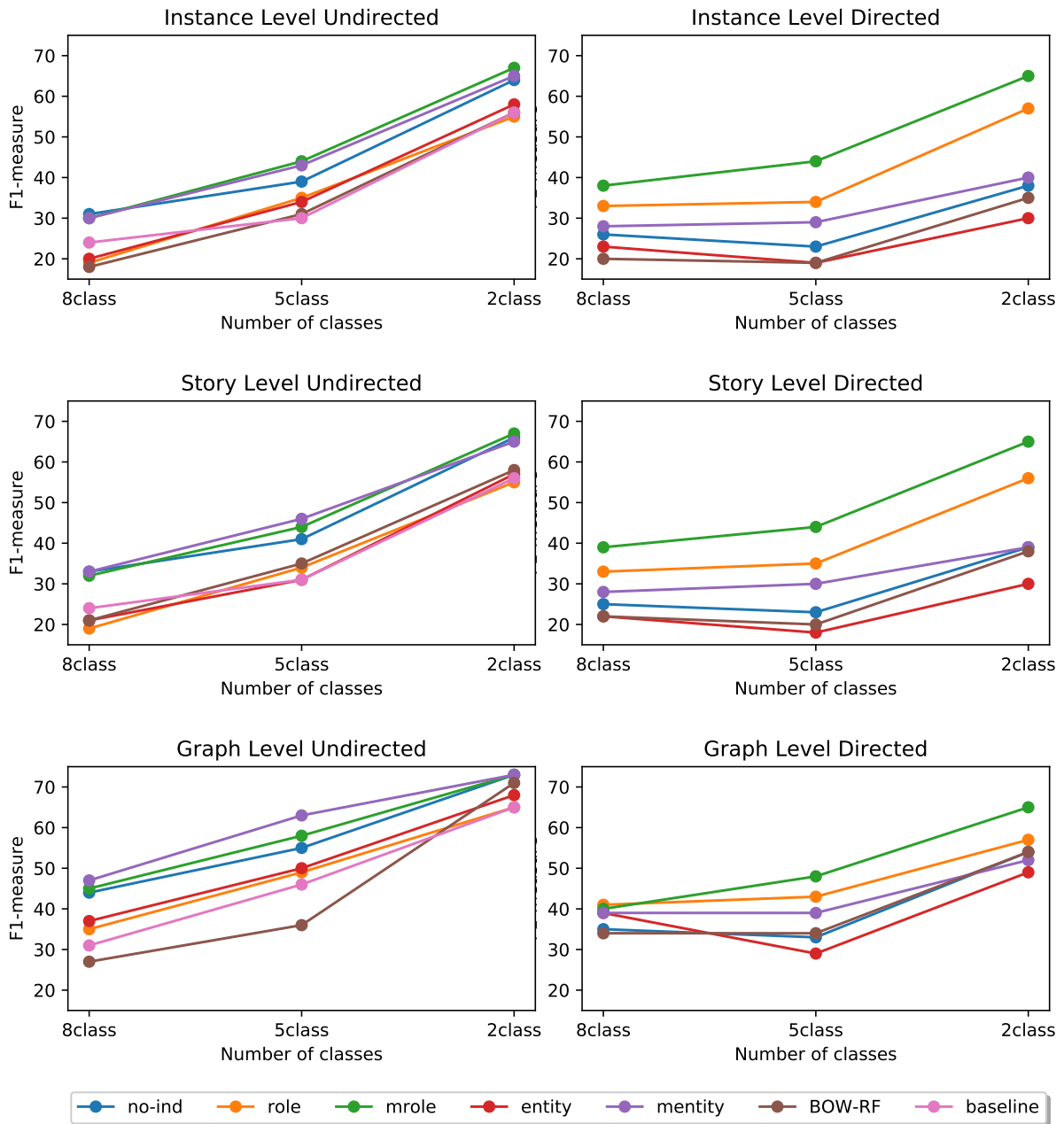
Complete Result Table

Table 1 contains the complete results with precision, recall and F_1 .

		Undirected									Directed								
		8 Class			5 Class			2 Class			8 Class			5 Class			2 Class		
		P	R	F ₁	P	R	F ₁	P	R	F ₁	P	R	F ₁	P	R	F ₁	P	R	F ₁
Dev Instance level	Baseline	19	31	24	25	38	30	39	100	56									
	BOW-RF	18	18	18	31	31	31	56	56	56	20	20	20	19	19	19	35	35	35
	GRU+NoInd.	31	31	31	39	39	39	64	64	64	26	26	26	23	23	23	37	37	37
	GRU+Role	19	19	19	35	35	35	55	55	55	33	33	33	34	34	34	57	57	57
	GRU+MaskRole	30	30	30	44	44	44	67	67	67	38	38	38	44	44	44	65	65	65
	GRU+Entity	20	20	20	34	34	34	58	58	58	23	23	23	19	19	19	30	30	30
	GRU+MaskEntity	30	30	30	43	43	43	65	65	65	28	28	28	29	29	29	40	40	40
Dev Story level	Baseline	20	32	24	27	39	31	40	100	56									
	BOW-RF	20	24	21	33	36	35	58	59	58	21	25	22	19	23	20	37	39	38
	GRU+NoInd.	33	33	33	41	41	41	66	66	66	25	25	25	23	23	23	38	38	38
	GRU+Role	19	19	19	34	34	34	55	55	55	33	33	33	35	35	35	56	56	56
	GRU+MaskRole	32	32	32	44	44	44	67	67	67	39	39	39	44	44	44	65	65	65
	GRU+Entity	21	21	21	31	31	31	57	57	57	22	22	22	18	18	18	30	30	30
	GRU+MaskEntity	33	33	33	46	46	46	65	65	65	28	28	28	30	30	30	39	39	39
Dev Graph-level	Baseline	36	38	31	50	41	46	88	52	65									
	BOW-RF	68	17	27	72	35	36	70	72	71	72	23	34	79	23	34	54	54	54
	GRU+NoInd.	44	44	44	55	55	55	73	73	73	35	35	35	33	33	33	54	54	54
	GRU+Role	35	35	35	49	49	49	65	65	65	41	41	41	43	43	43	57	57	57
	GRU+MaskRole	45	45	45	58	58	58	73	73	73	40	40	40	48	48	48	65	65	65
	GRU+Entity	37	37	37	50	50	50	68	68	68	39	39	39	29	29	29	49	49	49
	GRU+MaskEntity	47	47	47	63	63	63	73	73	73	39	39	39	39	39	39	52	52	52
Test	GRU+MaskRole Inst.	30	30	30	44	44	44	64	64	64	38	38	38	43	43	43	65	65	65
	GRU+MaskRole Story	33	33	33	45	45	45	65	65	65	39	39	39	43	43	43	66	66	66
	GRU+MaskRole Graph	45	45	45	59	59	59	71	71	71	42	42	42	49	49	49	66	66	66

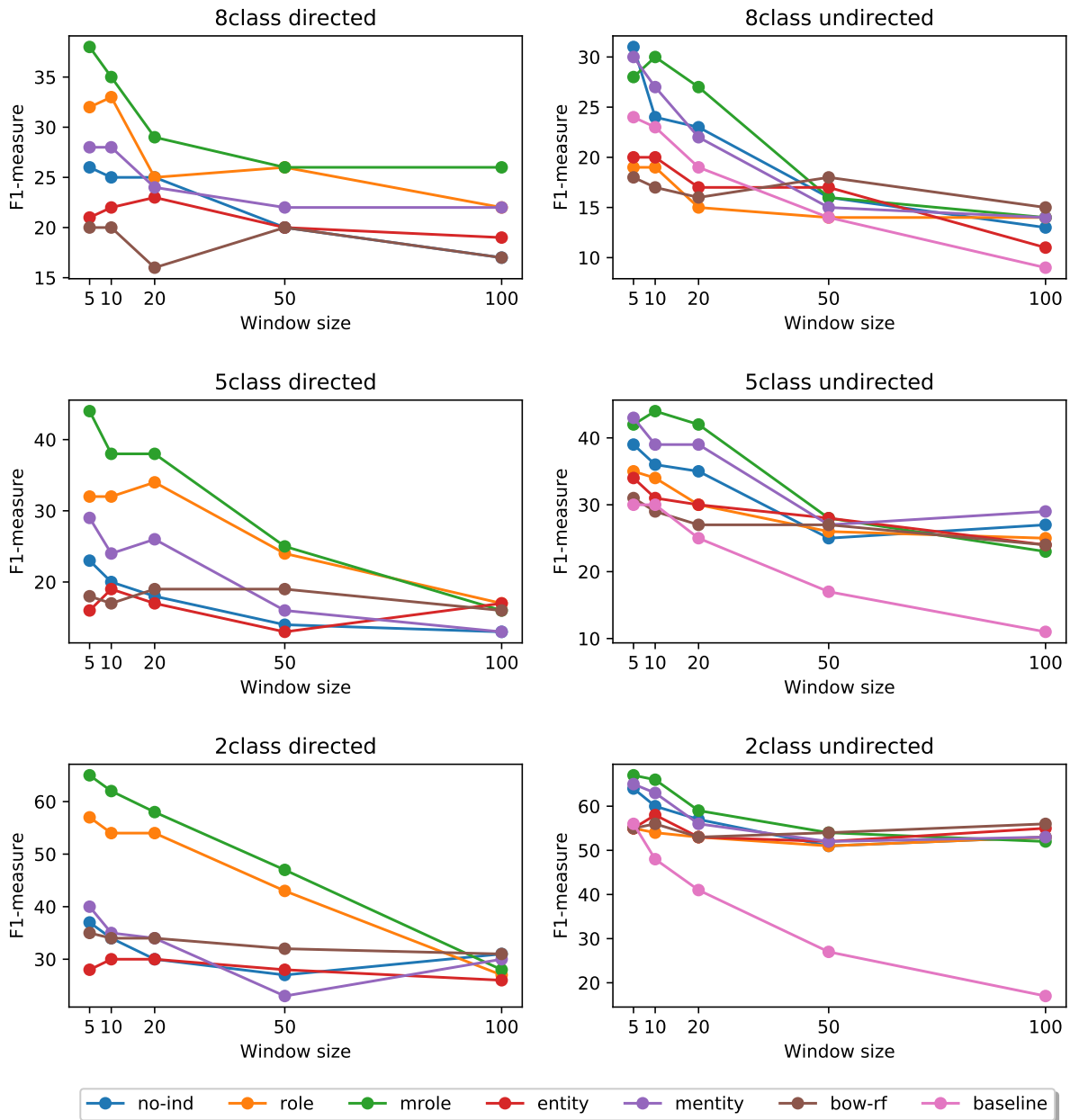
Table 1: Cross-validated results for different models in percentages of F_1 score. Inst. level: aggregated over all instances in the dataset. Story level: averaged performance on all stories. Graph-level: averaged performance on graph level on all stories. Test results are reported for the best indicator type. *GRU+NoInd.*: Alice is angry with Bob. *GRU+Role*: $\langle \text{exp} \rangle \text{Alice} \langle / \text{exp} \rangle \dots \langle \text{target} \rangle \text{Bob} \langle / \text{target} \rangle$. *GRU+MaskRole*: $\langle \text{exp} \rangle \dots \langle \text{target} \rangle$. *GRU+Entity*: $\langle \text{ent} \rangle \text{Alice} \langle / \text{ent} \rangle \dots \langle \text{char} \rangle \text{Bob} \langle / \text{char} \rangle$. *GRU+MaskEntity*: $\langle \text{ent} \rangle \dots \langle / \text{ent} \rangle$.

Results as Plots



The plots show the performance of our models with different number of modeled classes. One may observe that all models perform better in a 2-class scenario (directed and undirected). However, the differences between the models in a 2-class setting are marginal, especially in the undirected scenario. This may suggest that character relations are more nuanced than binary. It also suggests that directionality is an important aspect for the task of relation classification. In the directed classification scenario, the differences between different models are more pronounced, as compared to the undirected scenario.

Window Size Experiments on Instance-level



The plots depict the performance of all models evaluated on the instance-level for one example run. We tuned the window size parameter on a development set using a set of window sizes of 5, 10, 20, 50, and 100 tokens around character mentions. As one may see, the window size of 5 tokens is the best in the majority of cases. The GRU+Entity model shows an exception as it achieves the highest performance with 20 tokens in the 8-class directed scenario. The 2-class GRU+Entity works best with 10 tokens around the character mentions.