Team Enigma at ArgMining-EMNLP 2021: Leveraging Pre-trained Language Models for Key Point Matching

Manav Nitin Kapadnis*, **Sohan Patnaik***,
Siba Smarak Panigrahi*, **Varun Madhavan***, **Abhilash Nandy**
Indian Institute of Technology Kharagpur
{iammanavk, sohanpatnaik106, sibasmarak.p, varun.m.iitkgp, nandyabhilash}@gmail.com

Appendix : Supplementary Material

9.1 Appendix A

The Tables 1 and 2 show the list of top ten frequently occurring dependency feature tags and POS (Part Of Speech) tags along with their brief explanations.

Dependency Tag	Short Explanation
aux	Auxiliary
nsubj	Nominal subject
amod	Adjectival modifier
dobj	Direct object
prep	Prepositional modifier
pobj	Object of preposition
ROOT	Root
compound	Compound
conj	Conjunct
ccomp	Clausal complement

Table 1: Top ten frequent dependency features

POS Tag	Short Explanation
VERB	Verb
NOUN	Noun
AUX	Auxiliary
ADP	Adposition
ADJ	Adjective
PROPN	Proper noun
SCONJ	Subordinating conjunction
ADV	Adverb
PRON	Pronoun
DET	Determiner

Table 2: Top ten frequent POS features

*Equal contribution.

9.2 Appendix B

Tables 3, 4 and 5 report complete results[1] of the transformer models with additional feature vectors, whose best results were reported in the Results and Discussion section of the paper.

Model	Feature Type	mAP Strict	mAP Relaxed
BERT-large	Dep[2]	0.764 ± 0.035	0.901 ± 0.017
RoBERTa-large	Dep	0.802 ± 0.028	0.952 ± 0.021
BART-large	Dep	**0.868 ± 0.023**	**0.977 ± 0.015**
DeBERTa-large	Dep	0.851 ± 0.029	0.957 ± 0.018

Table 3: Results with dependency features

Model	Feature Type	mAP Strict	mAP Relaxed
BERT-large	POS[3]	0.808 ± 0.012	0.942 ±0.012
RoBERTa-large	POS	0.859 ± 0.047	0.964 ±0.002
BART-large	POS	**0.906 ± 0.011**	**0.987 ± 0.005**
DeBERTa-large	POS	0.880 ± 0.040	0.968 ± 0.012

Table 4: Results with parts of speech features

Model	Feature Type	mAP Strict	mAP Relaxed
BERT-large	Tf-idf	0.830 ± 0.023	0.937 ± 0.018
RoBERTa-large	Tf-idf	0.870 ± 0.041	0.967 ± 0.013
BART-large	Tf-idf	0.880 ± 0.029	0.969 ± 0.027
DeBERTa-large	Tf-idf	**0.911 ± 0.005**	**0.987 ± 0.008**

Table 5: Results with tf-idf features

[1] All the reported results are according to the default evaluation method for *mAP* strict and *mAP* relaxed scores (explained in section 5 of main paper)

[2] Encoded dependency features

[3] Encoded POS features

9.3 Appendix C

Table 6 reports the complete results[4] of transformer models pre-trained on the additional datasets with each of the additional feature and further fine tuned on the main shared task dataset with the same type of additional feature.

Some of the results which are marked with a dashed line could not be reported due to resource constraints.

Model	Additional Dataset	Feature type	mAP Strict	mAP Relaxed
BART-large	STS	POS	0.904 ± 0.031	0.985 ± 0.015
DeBERTa-large	STS	POS	0.877 ± 0.017	0.978 ± 0.011
BART-large	STS	Dep	0.899 ± 0.029	0.984 ± 0.014
DeBERTa-large	STS	Dep	0.874 ± 0.042	0.936 ± 0.073
BART-large	STS	Tf-idf	0.910 ± 0.008	0.985 ± 0.003
DeBERTa-large	STS	Tf-idf	0.903 ± 0.017	0.980 ± 0.013
BART-large	IBM Args30k	POS	0.912 ± 0.018	0.984 ± 0.002
DeBERTa-large	IBM Args30k	POS	–[5]	–
BART-large	IBM Args30k	Dep	0.890 ± 0.007	0.978 ± 0.006
DeBERTa-large	IBM Args30k	Dep	–	–
BART-large	IBM Args30k	Tf-idf	**0.919 ± 0.007**	**0.986 ± 0.002**
DeBERTa-large	IBM Args30k	Tf-idf	–	–

Table 6: Results with pretraining on additional datasets with additional features

[4] All the reported results are according to the default evaluation method for *mAP* strict and *mAP* relaxed scores (explained in section 5 of main paper)

[5] Due to resource constraints, we could not report values of these experimental settings