

Hyperparameter	Value
word embedding dimension	50
KG embedding dimension	50
position embedding dimension	5
CNN window size	3
CNN filter number	230
dropout rate	0.5
learning rate	0.01
batch size	160

Table 1: Hyperparameters used in PCNN+ATT on Medline21 dataset. The experiments are conducted on Nvidia Titan X(Pascal) GPU.

Hyperparameter	Value
word embedding dimension	50
KG embedding dimension	50
position embedding dimension	5
CNN window size	3
CNN filter number	100
dropout rate	0.5
learning rate (for sentences)	0.02
learning rate (for KG)	0.001
batch size	100

Table 2: Hyperparameters used in JointE on Medline21 dataset. The experiments are conducted on Nvidia Titan X(Pascal) GPU.

Hyperparameter	Value
word embedding dimension	50
KG embedding dimension	50
position embedding dimension	5
CNN window size	3
CNN filter number	100
dropout rate	0.5
learning rate (for sentences)	0.1
learning rate (for KG)	0.001
batch size	100

Table 3: Hyperparameters used in RELE on Medline21 dataset. The experiments are conducted on Nvidia Titan X(Pascal) GPU.

Hyperparameter	Value
learning rate	3e-5
hidden size	768
weight decay rate	1e-5
Adam epsilon	1-e8
warmup steps	500
batch size	100
maximum epochs	15

Table 4: Hyperparameters used in BRE+CE on Medline21 dataset. The experiments are conducted on a NVIDIA GeForce GTX 1080 TI GPU.

Hyperparameter	Value (Medline21)	Value (NYT10)
learning rate	3e-5	3e-5
hidden size	768	768
weight decay rate	1e-5	1e-5
Adam epsilon	1-e8	1-e8
warmup steps	500	500
batch size	100	80
maximum epochs	15	10

Table 5: Hyperparameters used in CRE+KA model. The experiments are conducted on a NVIDIA GeForce GTX 1080 TI GPU.

Hyperparameter	Value (Medline21)	Value (NYT10)
learning rate	3e-5	3e-5
hidden size	768	768
weight decay rate	1e-5	1e-5
Adam epsilon	1-e8	1-e8
warmup steps	500	500
batch size	100	80
w	1.0	0.6
λ_t	1.0	1.0
λ_s	1-e4	1-e4
maximum epochs	15	10

Table 6: Hyperparameters used in our proposed XBE model. The experiments (Medline21) are conducted on Nvidia Titan X(Pascal) GPU, and the experiments (NYT10) are conducted on a NVIDIA GeForce GTX 1080 TI GPU.

Bag	Sentence	Target Relation	XBE	XBE - X-stitch	BRE+CE
B1	... ' idol ' finalist arrested jessica sierra , one of the top 10 finalists on " american idol " in 2005 , was arrested early yesterday in tampa , ...	/people/person/ place_lived	✓	✗	✗
B2	... accepted a job , a friend suggested he check out a san francisco start-up , powerset , which, ...	/business/company/ place_founded	✓	✗	✗
B3	..., the coordinator of aurore , a renewable energy service company in auroville , india .	/location/location/ contains	✓	✗	✗
B4	president gloria macapagal-arroyo in a statement released after she voted in her province of pampanga , north of manila , said the country ...	/people/person/ place_lived	✓	✗	✗
B5	..., a track architect known for his work on tracks in bahrain , shanghai and sebang , malaysia .	/location/location/ contains	✓	✗	✗
B6	In the clinical context hair analysis be advantageously used to monitor the abuse of analgesic combinations with C0054234#ent , common among C0018681#ent patients .	may_treat	✓	✗	✗
B7	The solute carrier family 45 a3 member (C1822765#ent) , known also as C0965053#ent , has been implicated with ...	gene_encodes_ gene_product	✓	✗	✗
B8	... and investigated the effects of C1098320#ent on ID remodelling during development of C0018801#ent	may_treat	✓	✗	✗
B9	C1832024#ent interferes with anaerobic glycolysis pentose cycle C0035547#ent .	chemical_or_drug_ affects_gene_product	✓	✗	✗
B10	... to minimize acute and C0030201#ent has been provided by microspheres that slowly release C0006400#ent (MS-Bup) ...	may_be_ treated_by	✓	✗	✗

Table 7: Some practical results, where each bag contains one sentence, ✓(or ✗) represents the correct (or incorrect) prediction of the target relation. We can observe that the proposed X-stitch can facilitate XBE to correctly identify the relation especially when a sentence bag is noisy or implicitly represents the target relation.