

A Supplemental Material

A.1 Data Statistics

The statistics of the data set is shown in Table 5. This shows that the data is highly unbalanced and includes more negative pairs than positive pairs.

	DDI type	Train	Test
Positive	<i>Mechanism</i>	1,319	302
	<i>Effect</i>	1,687	360
	<i>Advice</i>	826	221
	<i>Int</i>	189	96
	Total	4,021	979
Negative		23,771	4,737
Total		27,792	5,716

Table 5: Statistics of the DDIExtraction 2013 shared task data set

A.2 Hyper-parameter Settings

We show hyper-parameters for training text-based model and molecular-based model in Tables 6 and 7, respectively.

Parameter	Value
Word embedding size	200
Word position embedding size	20
Convolution window size	[3, 5, 7]
Convolution filter size	100
Hidden layer size	500
Initial learning rate	0.001
Mini-batch size	50
L2 regularization parameter	0.0001

Table 6: Hyperparameters for text-based model

Parameter	Value
Molecular vector size	50
Number of steps	4
Hidden layer size	1,000
Initial learning rate	0.001
Mini-batch size	100
Hidden layer size of NFP	50
GRU unit size of GGNN	50

Table 7: Hyperparameters for molecular-based model