

## A Hyper-parameter Setting

For our framework, word embeddings and other trainable parameters are randomly initialized with Xavier (Glorot and Bengio, 2010). The dimension of word embedding, the dimension of hidden layers in both CNN and TCVAE and the number of convolutional filters in both CNN and TCVAE are 100. The non-linear mapping function is tanh everywhere. In CNN, the width of convolutional filters is 3 and the number of memory is 128. Moreover, the dimension of pooling stride is 2 for WDtext and 4 for DBPtext. In TCVAE, the dimension of latent layer is 50. Other hyper-parameters can be viewed in Table 1. In order to make a fair comparison, for DKRL and ConMask, their dimensions of word embedding and hidden layer, parameters of convolutional filters and strides of pooling layer are same as what we used in our framework, and other hyper-parameters are set to ones mentioned in their papers. For GMatching, we randomly initialize its embeddings like other models, and other hyper-parameters are their default values.

Name of hyper-parameter	Symbol in paper	Value in WDtext	Value in DBPtext
number of generated triplet	K	8	128
batch size	B		8
number of layers in CNN	N		3
number of steps in inner iteration	S		5
weight of $L_{kld}$	$\lambda_1$		1.0
weight of $L_{reg}$	$\lambda_2$		1.0
learning rate of inner iteration	$\alpha_1$		0.001
learning rate of Reptile	$\alpha_2$		0.001

Table 1: Hyper-parameters of our framework.