Architecture		
Model	**Parameter**	**Value**
Chr, Wrd	BiLSTM layers	3
Mt	BiLSTM layers	1
Chr, Wrd, Mt	BiLSTM size	400
Chr, Wrd, Mt	Dropout LSTMs	0.33
Chr, Wrd, Mt	Dropout MLP	0.33
Wrd	Dropout embeddings	0.33
Chr	Dropout embeddings	0.05
Chr, Wrd, Mt	Nonlinear act. (MLP)	ELU
Initialization		
Model	**Parameter**	**Value**
Wrd	embeddings	Zero
Chr	embeddings	Gaussian
Chr, Wrd, Mt	MLP	Gaussian
Training		
Model	**Parameter**	**Value**
Chr, Wrd, Mt	Optimizer	Adam
Chr, Wrd, Mt	Loss	Cross entropy
Chr, Wrd, Mt	Learning rate	0.002
Chr, Wrd, Mt	Decay	0.999994
Chr, Wrd, Mt	Adam epsilon	1e-08
Chr, Wrd, Mt	beta1	0.9
Chr, Wrd, Mt	beta2	0.999