A FramNet Frames Used for Relation Extraction on TACRED Dataset

Relation Types	FrameNet Frames
org:alternate_names, per:alternate_names	Being_named, Name_conferral, Namesake, Referring_by_name, Simple_naming
org:city_of_headquarters, org:country_of_headquarters, org:stateorprovince_of_headquarters	Being_located, Locale, Locale_by_characteristic_entity, Locale_by_collocation, Locale_by_event, Locale_by_ownership, Locale_by_use, Locale_closure, Locating, Locative_relation, Spatial_co-location
org:founded, org:founded_by	Intentionally_create
org:dissolved	Location_in_time, Relative_time, Time_vector, Timespan, Temporal_collocation, Temporal_subregion
org:member_of, org:members	Becoming_a_member, Membership
org:political/religious_affiliation, per:religion	People_by_religion, Religious_belief, Political_locales
org:subsidiaries, org:parents	Part_whole, Partitive, Inclusion
org:shareholders	Capital_stock
org:top_members/employees, org:number_of_employees/members, per:employee_of	Leadership, Working_a_post, Employing, People_by_vocation, Cardinal_numbers
per:age	Age
per:cause_of_death, per:date_of_death, per:city_of_death, per:country_of_death, per:stateorprovince_of_death	Death, Cause_harm
per:charges	Notification_of_charges, Committing_crime, Criminal_investigation
per:children, per:parents, per:other_family, per:siblings	Kinship
per:cities_of_residence, per:countries_of_residence, per:stateorprovinces_of_residence	Expected_location_of_person, Residence
per:date_of_birth, per:city_of_birth, per:country_of_birth, per:stateorprovince_of_birth	Being_born, Giving_birth
per:origin	Origin, People_by_origin
per:schools_attended	Education_teaching
per:spouse	Personal_relationship, Forming_relationships
per:title	Performers_and_roles

B Hyper-parameter Settings and Training Details

The dimension of POS and position embeddings are both 30. The inner layer dimension in position-wise feed-forward network is 130. The dimension of the relative positional encoding within knowledge-attention and self-attention is 50. The attention dimension is 200 in position-aware attention and 100 in multi-channel attention. The fully connected network before softmax has a dimensionality of 100. We use ReLU for all the nonlinear activation functions. Dropout rate is 0.4 for input embeddings, attention outputs and position-wise feed-forward outputs, and 0.1 for attention weights dropout.

The models are trained using Stochastic Gradient Descent with learning rate of 0.1 and momentum of 0.9. The learning rate is decayed with a rate of 0.9 after 15 epochs if F_1 score on dev set does not improve. The batch size is set to 100 and we train the model for 70 epochs.