Task-Aware Representation of Sentences for Generic Text Classification

This is a supplementary material for our submission with the above mentioned title at COLING 2020.

Appendix

The class label names used to train TARS for each dataset is provided in Tables 1, 2, 3, 4, 5, and 6.

Dataset	Surface Level Class Name	Curated Class Name
AGNEWS	World	World
	Sports	Sports
	Business	Business
	Sci/Tech	Science Technology

Table 1: Description of class label names used by TARS for AGNEWS corpus.

Dataset	Surface Level Class Name	Curated Class Name
DBPEDIA	Company	Company
	EducationalInstitution	Educational Institution
	Artist	Artist
	Athlete	Athlete
	OfficeHolder	Office Holder
	MeanOfTransportation	Mean Of Transportation
	Building	Building
	NaturalPlace	Natural Place
	Village	Village
	Animal	Animal
	Plant	Plant
	Album	Album
	Film	Film
	WrittenWork	Written Work

Table 2: Description of class label names used by TARS for DBPEDIA corpus.

Dataset	Surface Level Class Name	Curated Class Name
TREC-6	ABBR	question about abbreviation
	DESC	question about description
	ENTY	question about entity
	HUM	question about person
	LOC	question about location
	NUM	question about number

Table 3: Description of class label names used by TARS for TREC-6 corpus. It follows the taxonomy found at `https://cogcomp.seas.upenn.edu/Data/QA/QC/definition.html`

Dataset	Surface Level Class Name	Curated Class Name
TREC-50	ENTY:sport	question about entity sport
	ENTY:dismed	question about entity diseases medicine
	LOC:city	question about location city
	DESC:reason	question about description reasons
	NUM:other	question about number other
	LOC:state	question about location state
	NUM:speed	question about number speed
	NUM:ord	question about number order ranks
	ENTY:event	question about entity event
	ENTY:substance	question about entity elemetnt substance
	NUM:perc	question about number percentage fractions
	ENTY:product	question about entity product
	ENTY:animal	question about entity animal
	DESC:manner	question about description manner of action
	ENTY:cremat	question about entity creative pieces inventions books
	ENTY:color	question about entity color
	ENTY:techmeth	question about entity technique method
	NUM:dist	question about number distance measure
	NUM:weight	question about number weight
	LOC:mount	question about location mountains
	HUM:title	question about person title
	HUM:gr	question about person group organization of persons
	HUM:desc	question about person description
	ABBR:abb	question about abbreviation abbreviation
	ENTY:currency	question about entity currency
	DESC:def	question about description definition
	NUM:code	question about number code
	LOC:other	question about location other
	ENTY:other	question about entity other
	ENTY:body	question about entity body organ
	ENTY:instru	question about entity musical instrument
	ENTY:termeq	question about entity term equivalent
	NUM:money	question about number money prices
	NUM:temp	question about number temperature
	LOC:country	question about location country
	ABBR:exp	question about abbreviation expression
	ENTY:symbol	question about entity symbol signs
	ENTY:religion	question about entity religion
	HUM:ind	question about person individual
	ENTY:letter	question about entity letters characters
	NUM:date	question about number date
	ENTY:lang	question about entity language
	ENTY:veh	question about entity vehicle
	NUM:count	question about number count
	ENTY:word	question about entity word special property
	NUM:period	question about number period lasting time
	ENTY:plant	question about entity plant
	ENTY:food	question about entity food
	NUM:volsize	question about number volume size
	DESC:desc	question about description description

Table 4: Description of class label names used by TARS for TREC-50 corpus. It follows the taxonomy found at `https://cogcomp.seas.upenn.edu/Data/QA/QC/definition.html`

Dataset	Surface Level Class Name	Curated Class Name
Yelp-full	1	very negative restaurant sentiment
	2	negative restaurant sentiment
	3	neutral restaurant sentiment
	4	positive restaurant sentiment
	5	very positive restaurant sentiment

Table 5: Description of class label names used by TARS for YELP-FULL corpus.

Dataset	Surface Level Class Name	Curated Class Name
AMAZON-FULL	1	very negative product sentiment
	2	negative product sentiment
	3	neutral product sentiment
	4	positive product sentiment
	5	very positive product sentiment

Table 6: Description of class label names used by TARS for AMAZON-FULL corpus.