A Label Mapping

ATIS Slot	OntoNotes Label
AIRLINE_NAME	ORG
AIRPORT_NAME	FAC
ARRIVE_DATE, DAY_NAME, DAY_NUMBER, DEPART_DATE, DEPART_TIME, FLIGHT_DAYS,RETURN_DATE, TIME_RELATIVE, TO-DAY_RELATIVE	DATE
ARRIVE_TIME, MONTH_NAME, PE-RIOD_OF_DAY, RE-TURN_TIME, TIME	TIME
CITY_NAME, FROM_LOC, STATE_CODE, STATE_NAME, STOP_LOC, TO_LOC	GPE
COST_RELATIVE, FARE_AMOUNT	MONEY
DAYS_CODE, ECONOMY, FARE_BASIS_CODE, FLIGHT_MOD, MEAL, MEAL_CODE, MEAL_DESCRIPTION, MOD, FLIGHT_STOP, FLIGHT_MOD, OR, RESTRICTION_CODE, ROUNDTRIP, TRANS-PORT_TYPE	O
FLIGHT_NUMBER	CARDINAL

Table 3: Label Mapping from ATIS to OntoNotes.

B Hyperparameters

Hyperparameter	Value
LSTM cell size	100
Dropout	0.5
Word embedding dimension	300
Character embedding dimension	100
Mini-batch size	128
Optimizer	Adam
Learning rate	X
Number of epoch	20
Early stopping	10

Table 4: Hyperparameters for the neural models used in STL and MTL

Parameter	Adopted value
Surrogate model	Gaussian Processes with MCMC sampling
Acquisition function	Expected Logarithmic Improvement
Number of initial evaluation points	3
Search space upper bound	1
Search space lower bound	-1
Number of iterations	50

Table 5: Parameters used by the Bayesian Optimizer.

C Domain Similarity

$\mathcal{D}_{\mathcal{T}}$	$\mathcal{D}_{\mathcal{S}}$						Avg	Δ
	TC	NW	BC	BN	WB	MZ		
TC	-	0.74	0.84	0.80	0.83	0.77	0.80	1.7
NW	0.74	-	0.85	0.91	0.91	0.90	0.86	0.7
BC	0.84	0.85	-	0.90	0.90	0.86	0.87	0.02

Table 6: Domain Similarity (JSD) for each $\mathcal{D}_{\mathcal{T}}$ and $\mathcal{D}_{\mathcal{T}}$, and the Δ gap between the best data selection method and DS$_{all}$.