SANTA CLARA UNIVERSITY

Entity-aware Multi-task Learning for Query Understanding at Walmart

Zhiyuan Peng@SCU, Vachik Dave@Walmart, Nicole McNabb@Walmart,

Rahul Sharnagat@Walmart, Alessandro Magnani@Walmart,

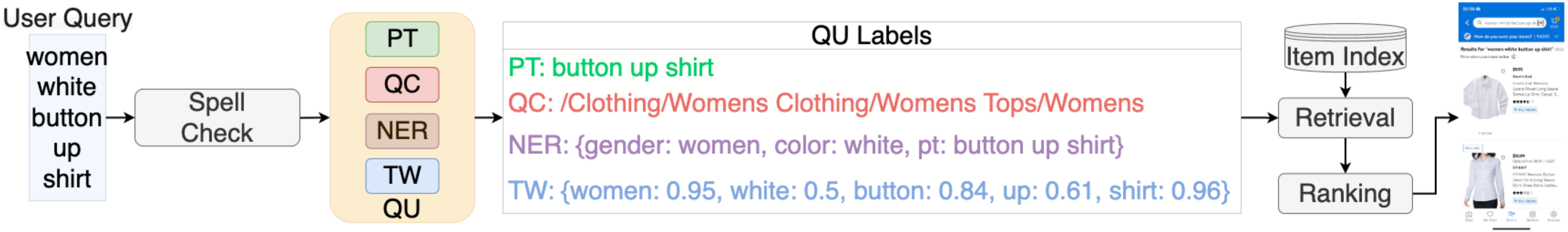
Ciya Liao@Walmart, Yi Fang@SCU, Sravanthi Rajanala@Walmart

Walmart Global Tech

Outlines

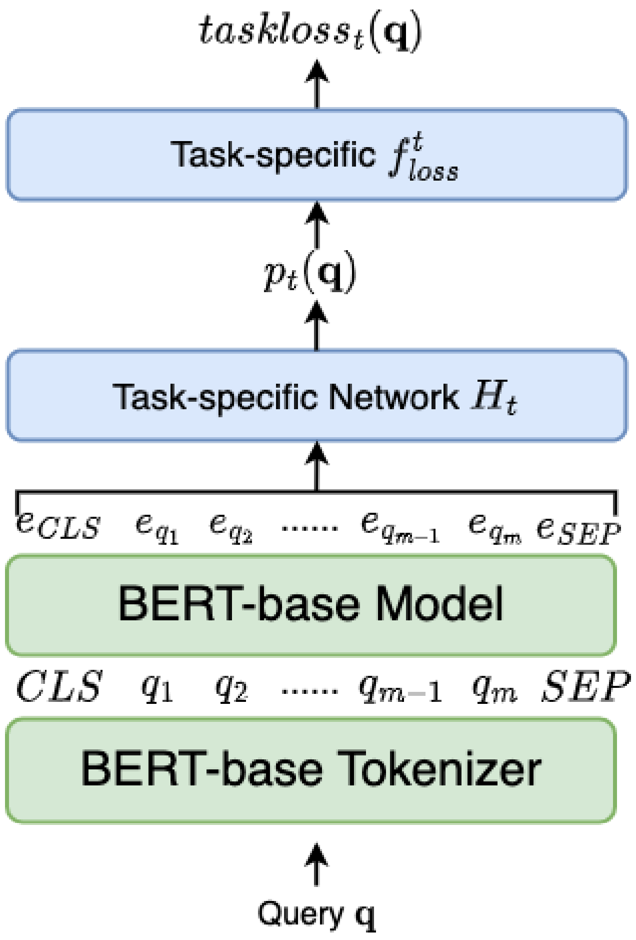
- Background
- Related Work
- Entity-aware Multi-task (EAMT)
- Experiments & Ablation Study
- Conclusion & Future work

Background: Query Understanding (QU) in Search

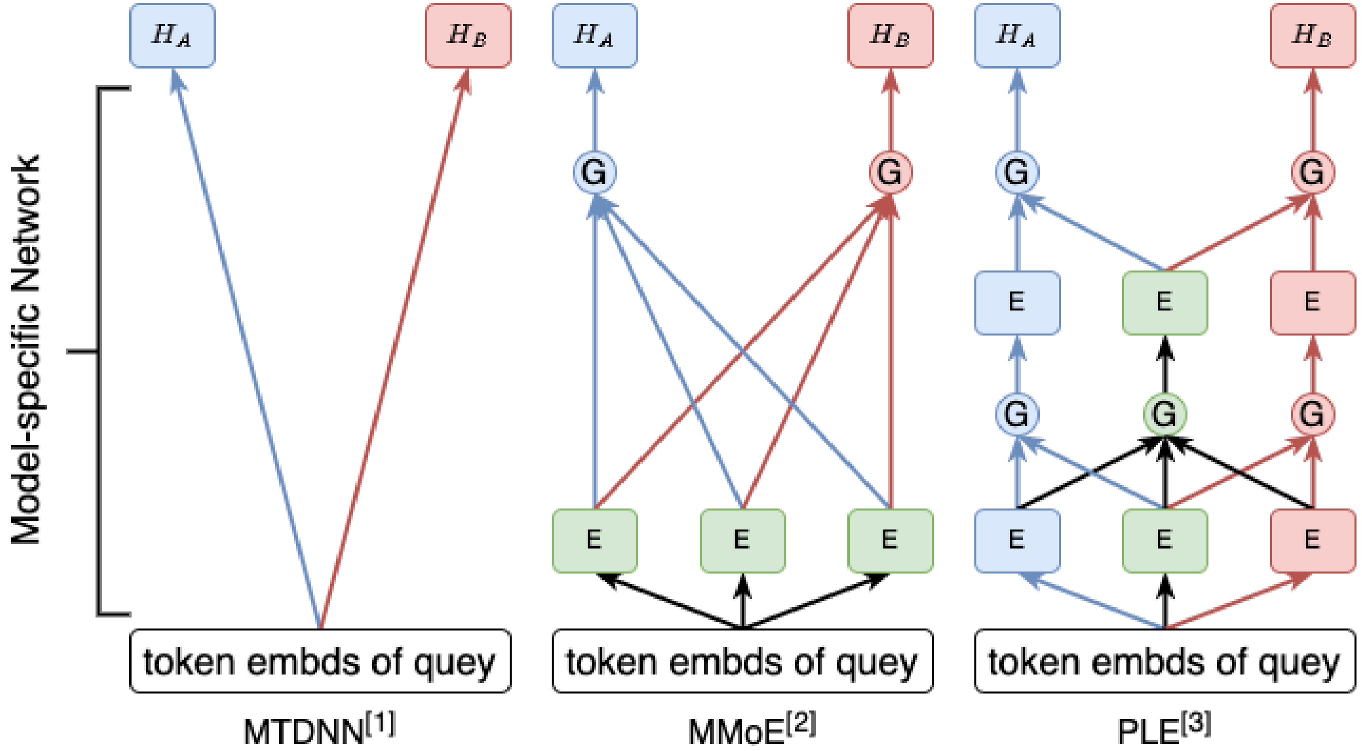


Task	Full Name	Type
PT	Product Type	Sentence-level Multi-label
QC	Query Category	Sentence-level Multi-label
NER	Named Entity Recognition	Token-level Multi-class
TW	Term Weight	Token-level Binary

Background: Single Task Model

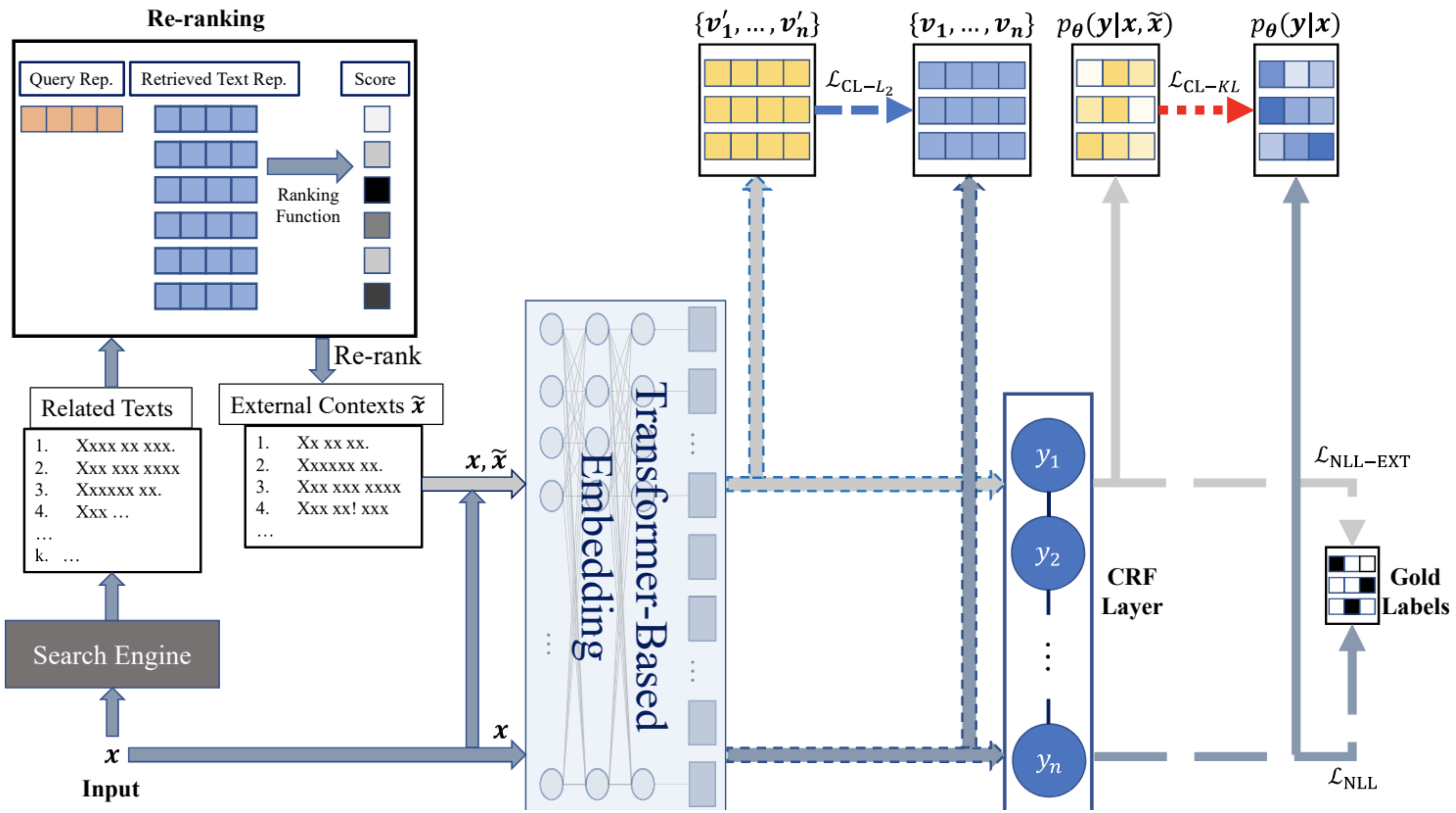


Related Work: Multi-task Learning



[1] Multi-Task Deep Neural Networks for Natural Language UnderstandingMulti-Task Deep Neural Networks for Natural Language Understanding
[2] Sparsely activated mixture-of-experts are robust multi-task learners
[3] Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations

Related Work: CL-KL[1]



[1] Improving Named Entity Recognition by External Context Retrieving and Cooperative Learning

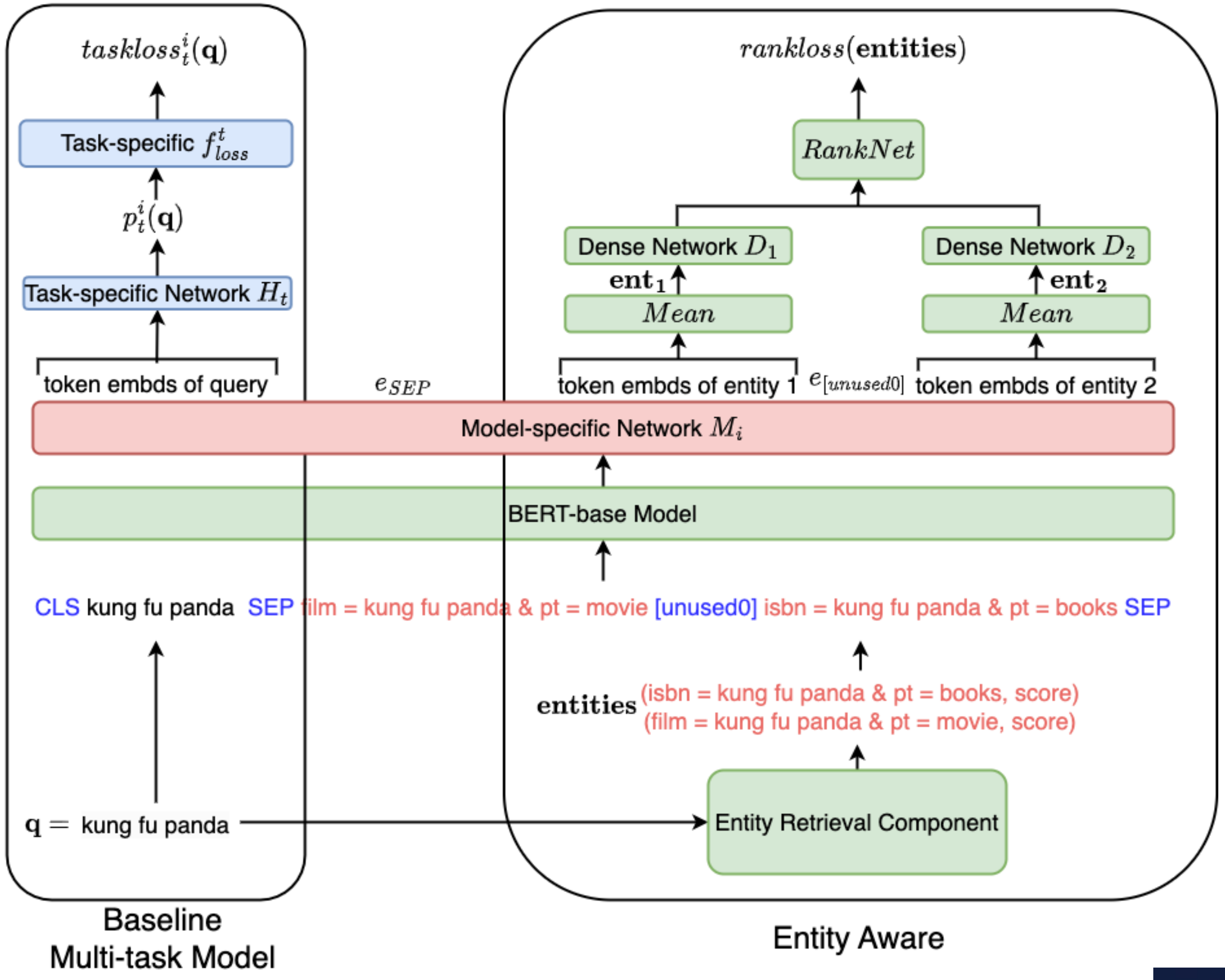
Motivation

- Tackling each task separately:
  - Excessive workload
  - Increase latency and resource usage
- All the tasks share the same input
  - Naturally suited for multi-task learning
- Engagement data
  - Query: {item1: {order: 3, click: 10}, …}

Our Contribution

- First multi-task learning work for e-commerce query understanding.
- Propose a novel large-scale entity-aware multi-task learning model (EAMT) by retrieving entities from engagement data as query context to augment the query representation.
- Comprehensive offline and online experiments to show the effectiveness of EAMT.

Model Structure: Entity-aware Multi-task (EAMT)



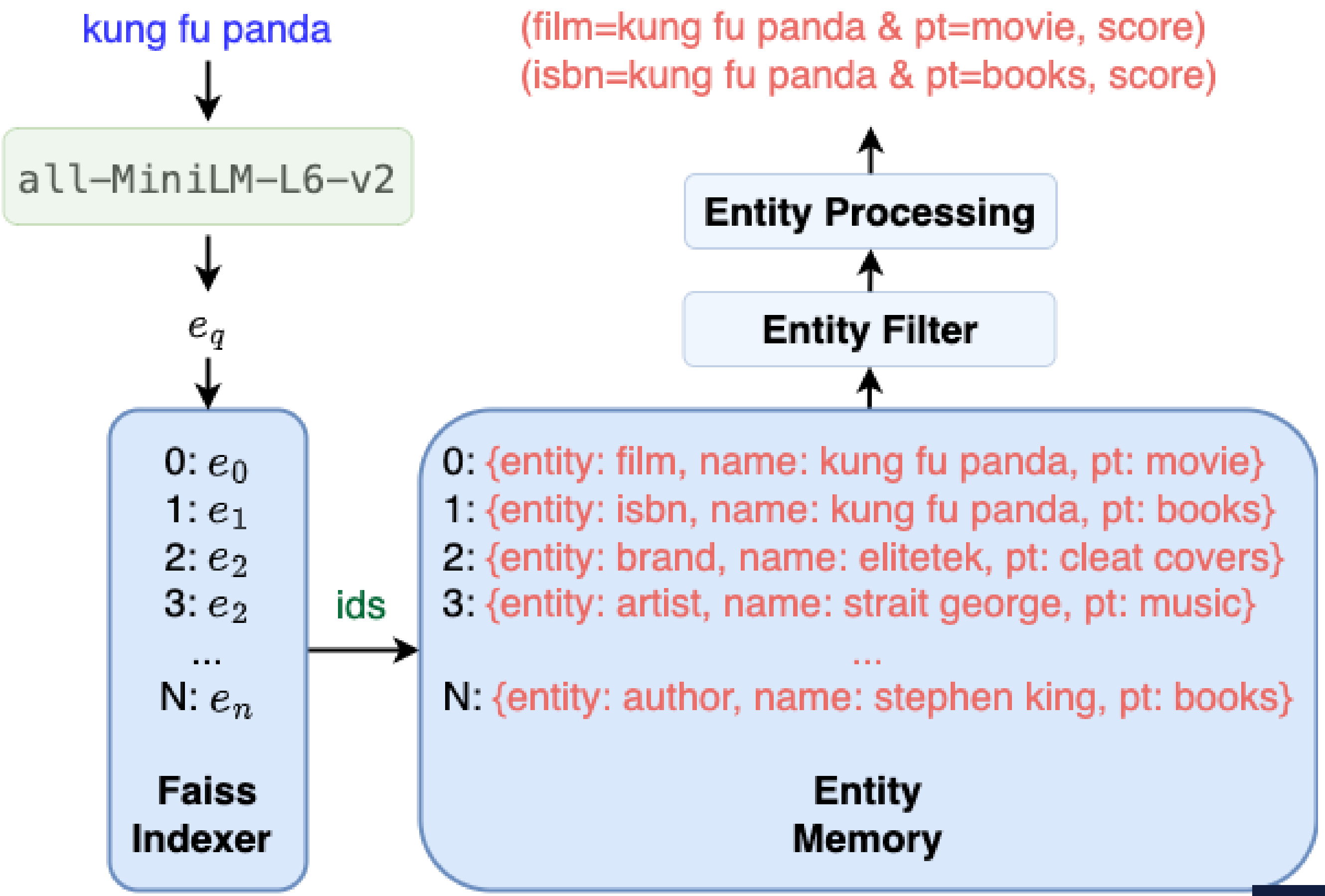
Model Structure: Loss

$$loss_t^i = taskloss_t^i(\mathbf{q}) + rankloss(\mathbf{entities})$$

$$rankloss(\mathbf{entities}) = \sum_{k=1}^{K} -softmax(y_k) \log softmax(s_k)$$

Task	Loss
PT	Binary Cross-entropy
QC	Binary Cross-entropy
NER	Cross-entropy
ST	Binary Cross-entropy

Model Structure: Entity Retrieval Component



Experiments: Datasets

Task	QU-3.75M			QU-965M		
	Train	Dev	Test	Train	Dev	Test
PT	1M	0.1M	0.1M	45.8M	0.46M	0.2M
QC	1M	0.1M	0.1M	44.7M	0.45M	0.2M
NER	0.14M	3.5K	10.5K	0.14M	3.5K	10.5K
TW	1M	0.1M	0.1M	3.2M	0.69M	0.69M

Experiments: Offline Test Results

Table 3: The relative F1 improvements of MTDNN and $EAMT_{mtdnn}$ **over single task models on QU-3.75M and QU-965M, respectively.** $entity\_num = 3$ **for** $EAMT_{mtdnn}$.

Task	QU-3.75M		QU-965M	
	MTDNN	$EAMT_{mtdnn}$	MTDNN	$EAMT_{mtdnn}$
PT	0.81	1.7	1.41	1.94
QC	1.04	2.27	1.85	2.6
NER	-0.27	-0.18	-0.05	0.89
TW	-0.33	0.05	-0.13	-0.27
Micro	0.48	1.29	0.51	0.66
Macro	0.32	0.96	0.77	1.29

Experiments: Online Test Results (MTDNN)

- Gross Merchandise Value (GMV): 0.51%
- Order: 0.65%
- Units: 1.08%
- Add To Cart (ATC): 0.65%
- Resource Usage:
  - 4 GPUs to 1 GPU

- Latency

Avg.	p90	p95	p99
2.10%	26.22%	24.96%	18.66%

Ablation Study: Entity-aware Component

Table 4: Relative F1 improvements of EAMT models over their corresponding baseline multi-task models on QU-3.75M. $entity\_num = 3$ **for all EAMT models.**

Model	PT	QC	NER	TW	ALL	
					Micro	Macro
$EAMT_{mtdnn}$	0.89	1.23	0.09	0.38	0.81	0.65
$EAMT_{mmoe}$	1.14	1.19	0.19	-0.14	0.71	0.59
$EAMT_{ple}$	0.59	1.15	0.16	-0.04	0.55	0.47

Ablation Study: # of Entities (entity_num)

Table 6: Relative F1 improvements of $EAMT_{mtdnn}$ with different entity_num over MTDNN on QU-3.75M. ListNet loss is utilized.

entity_num	PT	QC	NER	TW	ALL Micro	ALL Macro
2	0.76	1.08	0.00	0.19	0.65	0.51
3	0.89	1.23	0.09	0.38	0.81	0.65
5	0.74	1.32	0.29	0.18	0.73	0.63

Ablation Study: Ranking Loss

Table 5: Relative F1 improvements of $EAMT_{mtdnn}$ with ListNet loss over $EAMT_{mtdnn}$ without ListNet loss on different entity_num. Results are reported on QU-3.75M.

entity_num	ALL	
	Micro	Macro
2	-0.01	-0.03
3	0.18	0.09
5	0.10	0.11

Ablation Study: Oversampling

Table 7: Ablation study of oversampling NER dataset on QU-965M. The results are the relative F1 improvements of $EAMT_{mtdnn}$ model over single task models.

Oversample	PT	QC	NER	TW	ALL	
					Micro	Macro
No	1.59	2.22	-4.02	-0.92	0.07	-0.28
Yes	1.94	2.60	0.89	-0.27	0.66	1.29

Ablation Study: Baseline Multi-task Models

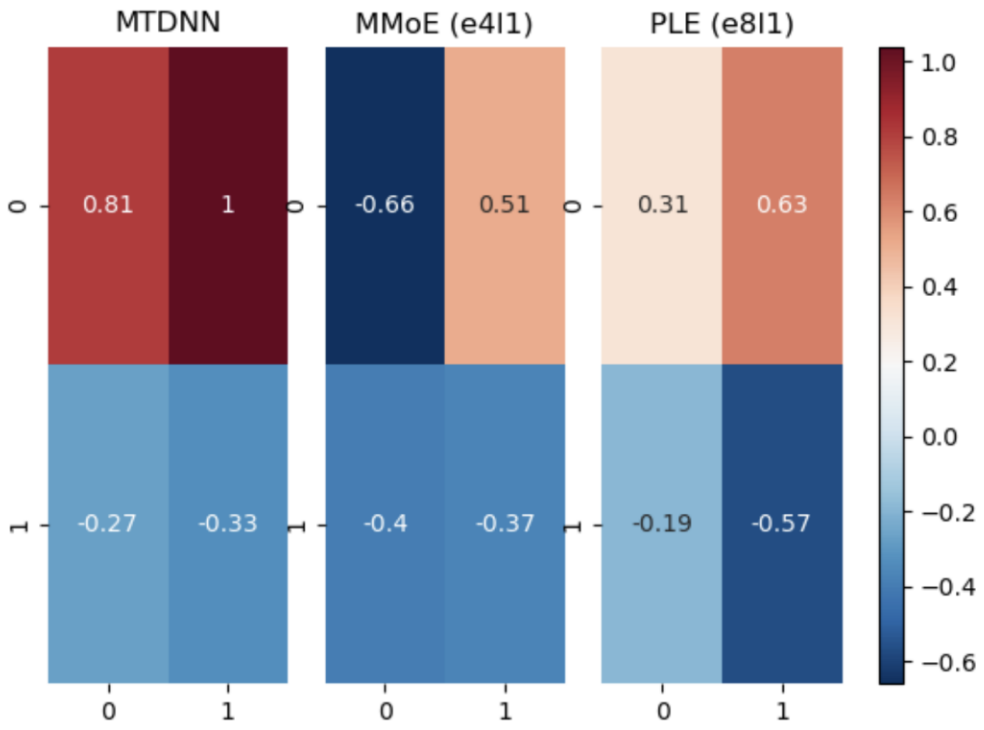


Figure 8: The relative F1 improvements of MTDNN, MMoE, and PLE over a single task. The coordinates of PT, QC, NER, and TW are (0,0), (0,1), (1,0), and (1,1), respectively.

Conclusion & Future work

- Conclusion
  - Multi-task learning improves the performance of QU
  - Entity-aware model does help learn a better representation of query representation
- Future Work
  - Online test for EAMT
  - Mitigate the imbalanced learning issue
  - Reduce the negative transfer learning

Acknowledgments

- Walmart Query Understanding Team
- KDD Student Travel Award

Questions