Title: Early-Exit offloading for Embedded QA Applications

Author: Yihua Liu, Shuocheng Chen, Yiming Ju

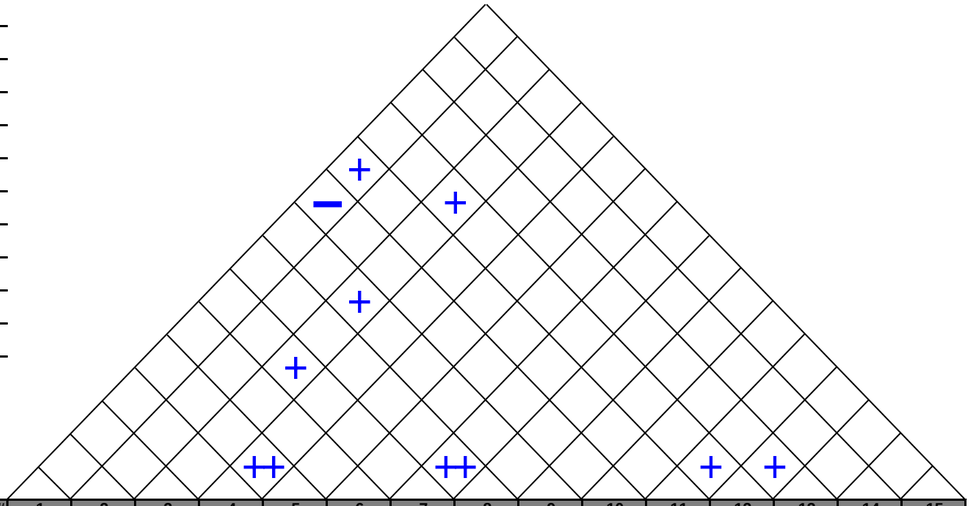
Date: 2021/12/14

Notes: Shanghai Jiao Tong University

UM-JI

ECE4730

Final Project



Legend		
Θ	Strong Relationship	9
O	Moderate Relationship	3
▲	Weak Relationship	1
++	Strong Positive Correlation	
+	Positive Correlation	
−	Negative Correlation	
▼	Strong Negative Correlation	
▼	Objective Is To Minimize	
▲	Objective Is To Maximize	
X	Objective Is To Hit Target	

Row #	Max Relationship Value in Row	Relative Weight	Weight / Importance	Demanded Quality (a.k.a. "Customer Requirements" or "Whats")	Column #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Competitive Analysis (0=Worst, 5=Best)					
				Direction of Improvement: Minimize (▼), Maximize (▲), or Target (x)		X	X	▼	▼	▼	X	▼	▼	▼	X	▼	▲	X	X	X							
				Quality Characteristics (a.k.a. "Functional Requirements" or "Hows")		Train set on the cloud	To predict on embedded system	Prediction time	Dataset Size on the cloud	Training time	Support of look-up tables (LUTs)	Number of model parameters	Size of model	Power consumption	Protection of user information	Network latency	Frequency to update local model	Support of subsequent model adjustments	Support of predicting offline	User guide and progammers' manua	Our project	BERT	ALBERT	EdgeBERT	DistilBERT	TinyBERT
1	9	12.3	10.0	Run question answering task		Θ	O														5	5	5	5	5	5
2	9	12.3	10.0	Can apply to embedded system			Θ														4	1	3	5	3	4
3	9	7.4	6.0	Have a good documentation for reference															Θ		4	5	5	3	4	3
4	9	11.1	9.0	Run fast for prediction			O	Θ	O							O	▲				4	5	5	4	4	5
5	9	11.1	9.0	Run precisely for prediction					Θ	O	Θ	Θ	▲					O			4	5	4	4	4	3
6	9	9.9	8.0	Small power										Θ						Θ						
7	9	9.9	8.0	Protect personal sensitive data											Θ											
8	3	8.6	7.0	Have a good transportability													▲	O			3	2	5	5	4	4
9	9	9.9	8.0	Occupy small storage space								Θ	Θ					Θ			4	2	4	4	4	5
10	9	7.4	6.0	Support update in terms of user need							▲						Θ									
				Target or Limit Value				1.0 s	100 MiB	70 min/epoch		15 MiB	100 MiB	10 W		0.1s	24 hours										
				Difficulty (0=Easy to Accomplish, 10=Extremely Difficult)		3	9	10	9	6	6	3	7	8	9	5	10	9	6	6							
				Max Relationship Value in Column		9	9	9	9	3	9	9	9	9	9	3	9	9	9	9							
				Weight / Importance		111.1	181.5	100.0	133.3	33.3	107.4	188.9	100.0	88.9	88.9	33.3	86.4	148.1	66.7	88.9							
				Relative Weight		7.1	11.7	6.4	8.6	2.1	6.9	12.1	6.4	5.7	5.7	2.1	5.6	9.5	4.3	5.7