MGIMN: Multi-Grained Interactive Matching Network for Few-shot Text Classification

Alibaba Group DAMO ALIBABA DAMO ACADEMY

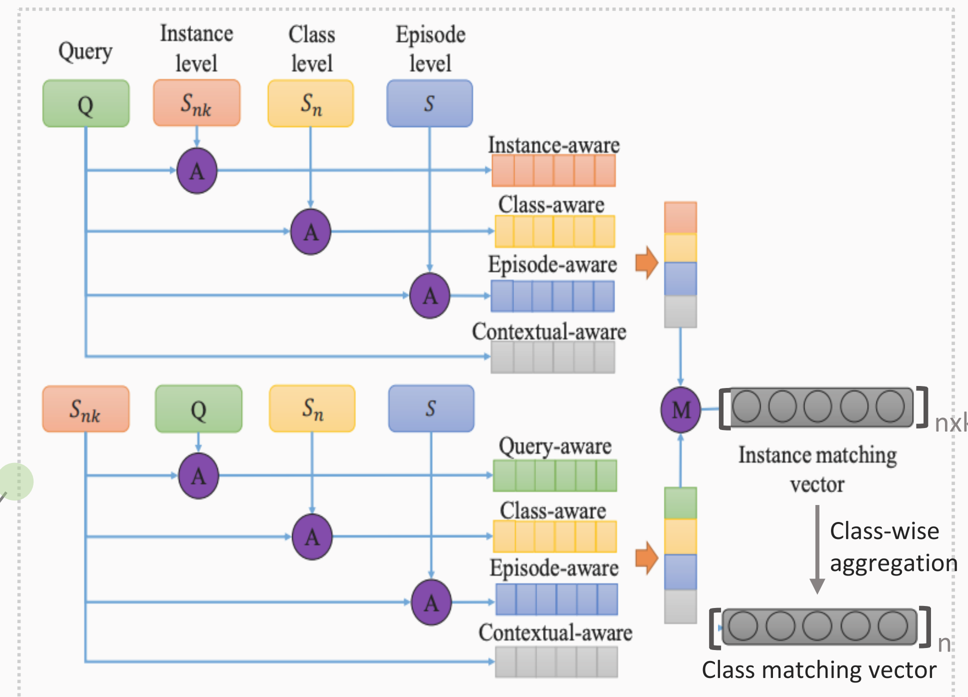
Jianhai Zhang, Mieradilijiang Maimaiti, Xing Gao, Yuanhang Zheng and Ji Zhang*

Introduction

WHAT Few-shot classification aims to predict the label of query given on query set and support set with few samples.

WHY Constructing a compact prototype of class is challenging-yet-unnecessary. And Inter-dependency between query and support set is much important.

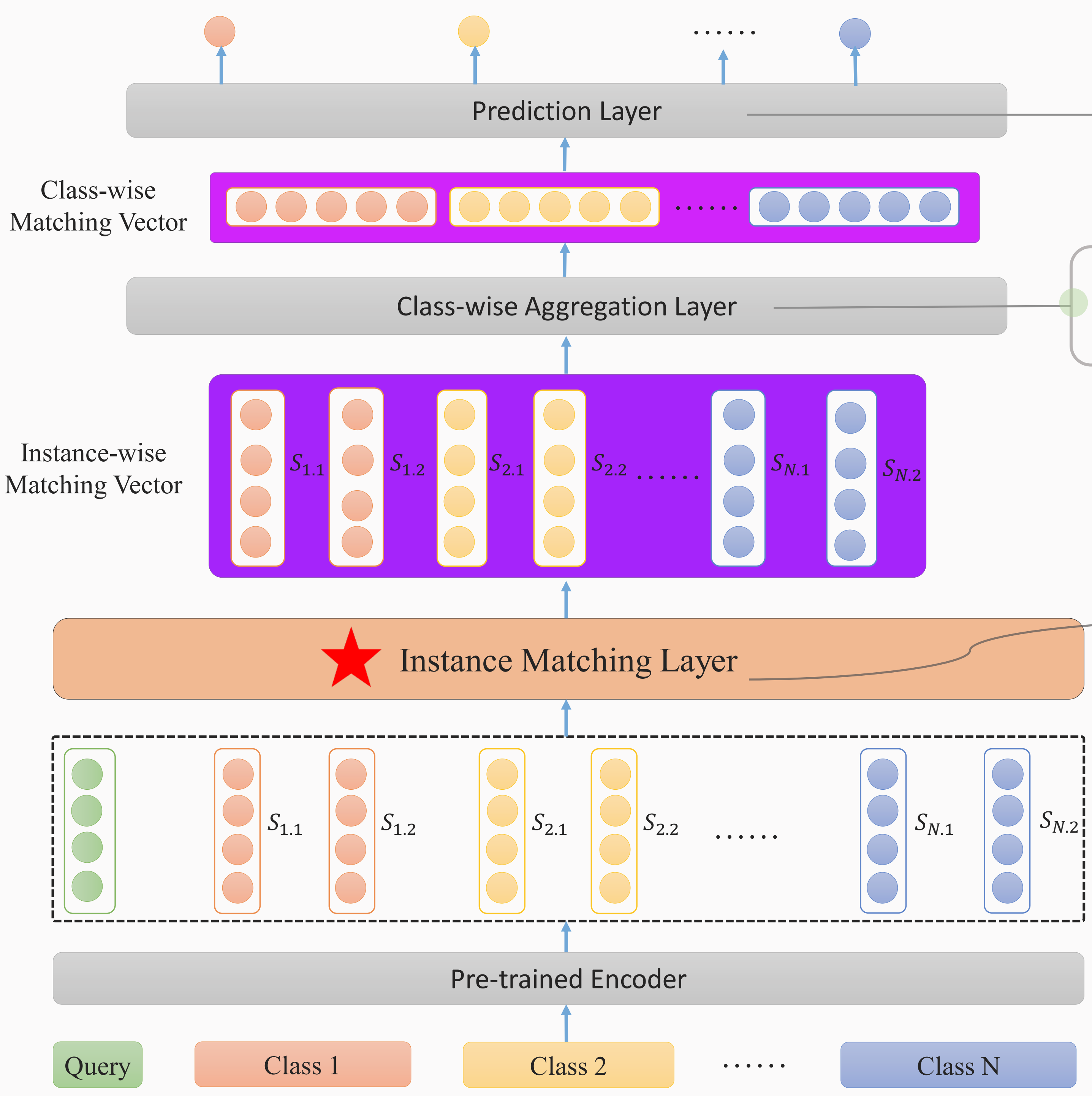
HOW Perform instance-level comparisons (multi-grained interactive matching) followed by class-wise aggregation.



Highlights

- Propose a new few-shot text classification framework(MGIMN).
- State-of-the-art on five different datasets with FSL and GFSL metrics.
- Introduce Retrieval-then-classify method to improve the inference performance in realistic scenery.

Approach



$$logit_n = MLP(\vec{c_n}), n = 1, .., N$$

$$\vec{c_n} = [max(\{\vec{m_{nk}}\}_{k=1}^{K}); avg(\{\vec{m_{nk}}\}_{k=1}^{K})]$$

$$\widehat{a}, \widehat{b} = BiAlign(a, b)$$
$$e_{ij} = \mathbf{F}(a_i)^{\mathrm{T}}\mathbf{F}(b_j)$$
$$\widehat{a}_i = \sum_{j=1}^{l_b} \frac{exp(e_{ij})}{\sum_{k=1}^{l_b} exp(e_{ik})} b_j$$
$$\widehat{b}_j = \sum_{i=1}^{l_a} \frac{exp(e_{ij})}{\sum_{k=1}^{l_a} exp(e_{kj})} a_i$$

Align with multi-grained context

$$q_{nk}^{'}, s_{nk}^{'} = BiAlign(q, s_{nk})$$
$$q_n^{''}, \_ = BiAlign(q, S_n)$$
$$s_{nk}^{''}, \_ = BiAlign(s_{nk}, S_n)$$
$$q^{'''}, \_ = BiAlign(q, S)$$
$$s_{nk}^{'''}, \_ = BiAlign(s_{nk}, S)$$

Fusion

$$\boldsymbol{q_{nk}^{'}} = H_1(q; q_{nk}^{'}; \left|q - q_{nk}^{'}\right|; q \odot q_{nk}^{'})$$
$$\boldsymbol{q_{nk}^{''}} = H_2(q; q_n^{''}; \left|q - q_n^{''}\right|; q \odot q_n^{''})$$
$$\boldsymbol{q_{nk}^{'''}} = H_3(q; q^{'''}; \left|q - q^{'''}\right|; q \odot q^{'''})$$
$$\boldsymbol{s_{nk}^{'}} = H_1(s_{nk}; s_{nk}^{'}; \left|s_{nk} - s_{nk}^{'}\right|; s_{nk} \odot s_{nk}^{'})$$
$$\boldsymbol{s_{nk}^{''}} = H_2(s_{nk}; s_{nk}^{''}; \left|s_{nk} - s_{nk}^{''}\right|; s_{nk} \odot s_{nk}^{''})$$
$$\boldsymbol{s_{nk}^{'''}} = H_3(s_{nk}; s_{nk}^{'''}; \left|s_{nk} - s_{nk}^{'''}\right|; s_{nk} \odot s_{nk}^{'''})$$
$$\boldsymbol{q_{nk}} = H(\boldsymbol{q_{nk}^{'}}; \boldsymbol{q_{nk}^{''}}; \boldsymbol{q_{nk}^{'''}})$$
$$\boldsymbol{s_{nk}} = H(\boldsymbol{s_{nk}^{'}}; \boldsymbol{s_{nk}^{''}}; \boldsymbol{s_{nk}^{'''}})$$

Instance-wise comparison

$$\vec{q_{nk}} = [max(\{\boldsymbol{q_{nk}}\}); avg(\{\boldsymbol{q_{nk}}\})]$$
$$\vec{s_{nk}} = [max(\{\boldsymbol{s_{nk}}\}); avg(\{\boldsymbol{s_{nk}}\})]$$
$$\vec{m_{nk}} = G(\vec{q_{nk}}; \vec{s_{nk}}; |\vec{q_{nk}} - \vec{s_{nk}}|; \vec{q_{nk}} \odot \vec{s_{nk}})$$

Dataset Statics

Datasets	Standard FSL Setting		Generalized FSL Setting						
	#sentences	$C_{tr}/C_{val}/C_{test}$	C	K	#sc	#uc	$\#D_{tr}$	$\#D_{val}$	$\#D_{test}$
OOS	22500	50/50/50	150	5	50	100	7000	1250	1250
Liu	25478	18/18/18	54	5	18	36	8312	450	450
Amzn	3057	106/106/106	318	5	106	212	1043	530	530
Huffpost	41000	14/13/14	41	5	14	27	13860	340	340
FaqIr	1233	17/16/17	50	5	17	33	309	381	381

- Standard FSL Setting: Widely used in most studies(Snell et al., 2017 etc.)
- Generalized FSL Setting(GFSL, Nguyen et al., 2020 etc.) : A more challenging-yet-realistic evaluation method. In this setting, we reform task a C-way K-shot classification in which only subset of classes are seen in training phase.

Experiments

Main Experiments

Methods	OOS			Liu			FaqIr		
	5-way	10-way	GFSL	5-way	10-way	GFSL	5-way	10-way	GFSL
Proto	92.20	87.91	61.94	82.46	73.23	47.66	89.83	81.56	60.78
Matching	89.78	84.41	58.34	78.25	67.45	41.95	86.74	78.77	53.85
Induction	80.44	70.92	34.00	65.58	51.56	24.73	71.62	56.99	20.10
Proto-HATT	92.84	89.11	65.52	82.38	75.29	51.27	85.01	76.17	62.62
MLMAN	95.99	93.41	74.39	87.39	79.82	57.24	94.77	89.49	74.42
MGIMN(ours)	**96.36**	**94.00**	**76.23**	**87.84**	**80.60**	**57.66**	**95.14**	**90.69**	**75.81**

Methods	Amzn		
	5-way	10-way	GFSL
Proto	78.40	69.02	41.03
Matching	75.73	64.17	38.34
Induction	64.02	50.12	20.09
Proto-HATT	78.05	69.00	41.81
MLMAN	85.64	79.39	46.71
MGIMN(ours)	**85.96**	**80.07**	**49.46**

Methods	Huffpost		
	5-way	10-way	GFSL
Proto	51.57	36.74	16.47
Matching	49.77	34.28	14.18
Induction	44.69	29.35	10.40
Proto-HATT	51.23	36.65	16.06
MLMAN	52.76	38.22	16.78
MGIMN(ours)	**54.98**	**40.12**	**19.61**

* All the results are the averaged over 15 runs(different seen-unseen class 365 splits and random seeds)

Ablation Study

Methods	Liu			Huffpost		
	5-way	10-way	GFSL	5-way	10-way	GFSL
MGIMN(ours)	**87.84**	**80.60**	**57.66**	**54.98**	**40.12**	**19.61**
w/o episode	86.22	78.99	56.67	54.14	39.53	18.69
w/o class	84.56	76.89	54.62	54.09	39.10	17.53
w/o instance	87.74	79.93	57.39	53.65	38.86	18.67
w/o instance&class&episode	80.53	70.94	42.54	51.81	37.10	16.48

Retrieval-then-classify

Methods	Liu(c=50)		OOS(c=150)		Amzn(c=318)	
	score	speed	score	speed	score	speed
MGIMN-overall	57.66	315	76.23	757	49.46	1630
RTC-BM25	54.97	55	74.80	56	44.76	58
RTC-oribert	52.93	60	70.55	65	31.09	70
RTC-mgimnbert	56.21	60	75.58	65	46.80	70

* Processing 100 queries on Intel Core i7 CPUs(ms/query)
* Sequence length=20

Email: tanfan.zjh@alibaba-inc.com; mieradilijiang.mea@alibaba-inc.com;