Dynamic Memory Induction Networks for Few-Shot Text Classification

Ruiying Geng, Binhua Li, Yongbin Li, Jian Sun, Xiaodan Zhu

Motivation

Observation: good prototypes yield better performance

Literature:

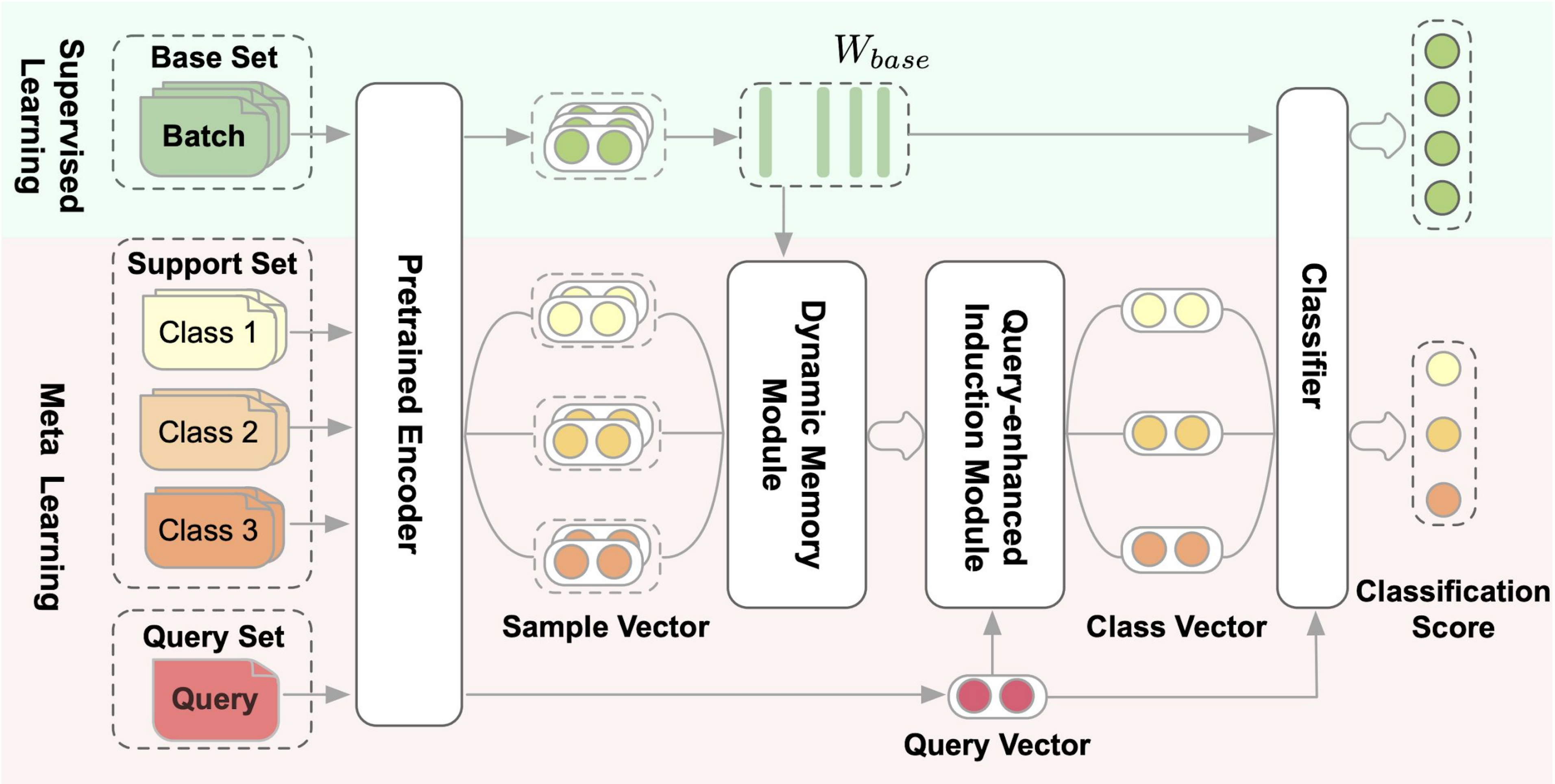
Average prototype (Prototypical Network)

Query-based weighted sum (Hybrid Attention Prototypical network)

Induction network (Dynamic Routing)

This paper try to advance the prototypes using dynamic memory with knowledge from supervised task

Overview



Supervised Learning
Base Set
Batch
W_{base}
Meta Learning
Support Set
Class 1
Class 2
Class 3
Query Set
Query
Pretrained Encoder
Sample Vector
Dynamic Memory Module
Query-enhanced Induction Module
Class Vector
Classifier
Classification Score
Query Vector

Three-stage matching network

1. Dynamic Memory Instance Encoder

$q' = DMR(M, q)$

a memory matrix M (here W_{base})

sample vector $q \in R^d$,

2. Dynamic Memory Query-based Prototype

$$e_c = DMR(\{e'_{c,s}\}_{s=1,\dots,K}, e_q).$$

3. Classifier

$$s_{q,c} = \tau \cdot cos(e_q, e_c) = \tau \cdot \overline{e}_q^T \overline{e}_c.$$

Dynamic Memory Routing

$q' = DMR(M, q)$

a memory matrix M (here W_{base})

sample vector $q \in R^d$,

1. Linear transformation and squash

$$
\begin{aligned}
\hat{m}_{ij} &= squash(W_j m_i + b_j), \\
\hat{q}_j &= squash(W_j q + b_j),
\end{aligned}
$$

2. Pearson Correlation Coefficient

$$
\begin{aligned}
p_{ij} &= tanh(PCCs(\hat{m}_{ij}, \hat{q}_j)), \\
PCCs &= \frac{Cov(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}}.
\end{aligned}
$$

3. Routing iteration

$$
\begin{aligned}
d_i &= softmax\left(\alpha_i\right), \\
\alpha_{ij} &= \alpha_{ij} + p_{ij}\hat{m}_i v_j.
\end{aligned}
$$

4. Vector representation

$$
\begin{aligned}
\hat{v}_j &= \sum_{i=1}^{n}(d_{ij} + p_{ij})m_{ij}, \\
v_j &= squash(\hat{v}_j).
\end{aligned}
$$

Algorithm

Algorithm 1 Dynamic Memory Routing Process

Require: r, q and memory $M = \{m_1, m_2, ..., m_n\}$

Ensure: $\boldsymbol{v} = v_1, v_2, ..., v_l, q'$

for all m_i, v_j **do**

  $\hat{m}_{ij} = squash(W_j m_i + b_j)$

  $\hat{q}_j = sqush(W_j q + b_j)$

  $\alpha_{ij} = 0$

  $p_{ij} = tanh(PCCs(\hat{m}_{ij}, \hat{q}_j))$

end for

for r iterations **do**

  $d_i = softmax\,(\alpha_i)$

  $\hat{v}_j = \sum_{i=1}^{n}(d_{ij} + p_{ij})\hat{m}_{ij}$

  $v_j = squash(\hat{v}_j)$

  for all i, j: $\alpha_{ij} = \alpha_{i,j} + p_{ij}\hat{m}_{ij}v_j$

  for all j: $\hat{q}_j = \frac{\hat{q}_j + v_j}{2}$

  for all i, j: $p_{ij} = tanh(PCCs(\hat{m}_{ij}, \hat{q}_j))$

end for

$q' = concat[\boldsymbol{v}]$

Return q'

Result

Model	5-way Acc.		10-way Acc.	
	1-shot	5-shot	1-shot	5-shot
BERT	30.79±0.68	63.31±0.73	23.48±0.53	61.18±0.82
ATAML	54.05±0.14	72.79±0.27	39.48±0.23	61.74±0.36
Rel. Net	59.19±0.12	78.35±0.27	44.69±0.19	67.49±0.23
Ind. Net	60.97±0.16	80.91±0.19	46.15±0.26	69.42±0.34
HATT	60.40±0.17	79.46±0.32	47.09±0.28	68.58±0.37
LwoF	63.35±0.26	78.83±0.38	48.61±0.21	69.57±0.35
DMIN	**65.72**±0.28	**82.39**±0.24	**49.54**±0.31	**72.52**±0.25

Ablation study

Model	**Iteration**	**1 Shot**	**5 Shot**
w/o DMM	3	81.79	90.19
w/o QIM	3	82.37	90.57
DMIN	1	82.70	90.92
DMIN	2	82.95	91.18
DMIN	3	**83.46**	**91.75**