# MDFEND: Multi-domain Fake News Detection

Qiong Nan[1,2,3], Juan Cao[1,2,*], Yongchun Zhu[1,2], Yanyan Wang[1,2], Jintao Li[1]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China [2]University of Chinese Academy of Sciences, Beijing, China
[3]State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, China
{nanqiong19z,caojuan,zhuyongchun18s,wangyanyan18s,jtli}@ict.ac.cn

CIKM'21

220217 Chia-Chun Ho

# Outline

Introduction

Related Works

Methodology

Experiments

Conclusions

Comments

# Introduction

## Single domain fake news detection

- Most of approaches focus on single domain fake news detection (SFND).
  - e.g., politics (PolitiFact), health (CoAID)
- However, for a certain domain, the amount of fake news can be extremely limited.
- Therefore, based on such inadequate single-domain data, the performance of these detection models is unsatisfying.

# Introduction

## Multi domain fake news detection

- In practical scenarios, the real-world news platforms release various news in different domains everyday.
- Therefore, it's promising to solve the data sparsity problem and improve the performance of all domains.
  - By exploiting data from multiple domains, called multi-domain fake news detection (MFND).

# Introduction

## Multi domain fake news detection

- Data distributions vary from domain to domain, called domain shift.
  - Different domain have different word usage, the most commonly used words in
    - Military news: “navy”, “army”
    - Educational news: “students”, “university”, “teacher”
  - Propagation patterns vary a lot in different domains.
- Facing the problem of serious domain shift, MFND can therefore be quite challenging.

# Introduction

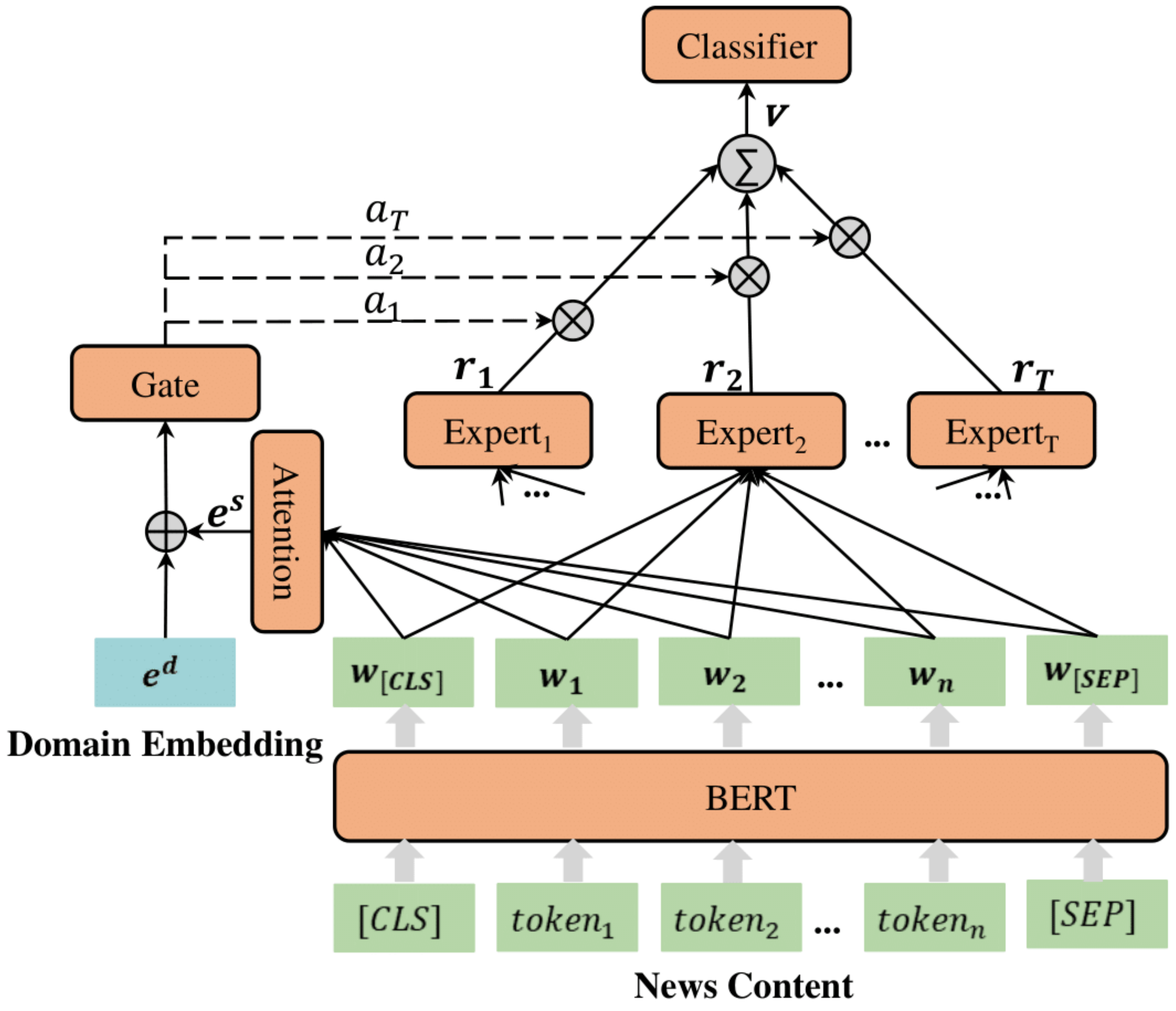
## Weibo21 MFND dataset

- Some domains only contain very little labels data, and this phenomenon further increases the difficulty in MFND, which remain unsolved yet with existing methods.
- The authors build a comprehensive dataset Weibo21, which contain news from 9 domains (Science, Military, Education, Disasters…).
- Every domain contains news content, released timestamp, corresponding pictures and comments.
- Weibo21 contains 4488 fake news and 4640 real news from 9 different domains.

# Introduction

## MDFEND

- Proposed a simple but effective Multi-domain Fake News Detection Model, namely MDFEND.
  - Utilizes a domain gate to aggregate multiple representations extracted by mixture of experts.
- Experiments demonstrate the significant effectiveness improvement of the proposed MDFEND compared with the baselines.



# Introduction

## Contributions

- Construct Weibo21, an MFND dataset.
  - This data repository is the first MFND dataset collected from one platform and contains the richest domains.
- Proposed a simple but effective method named MDFEND for MFND.
- Systematically evaluate MFND performance of different methods on proposed dataset.

# Related Works

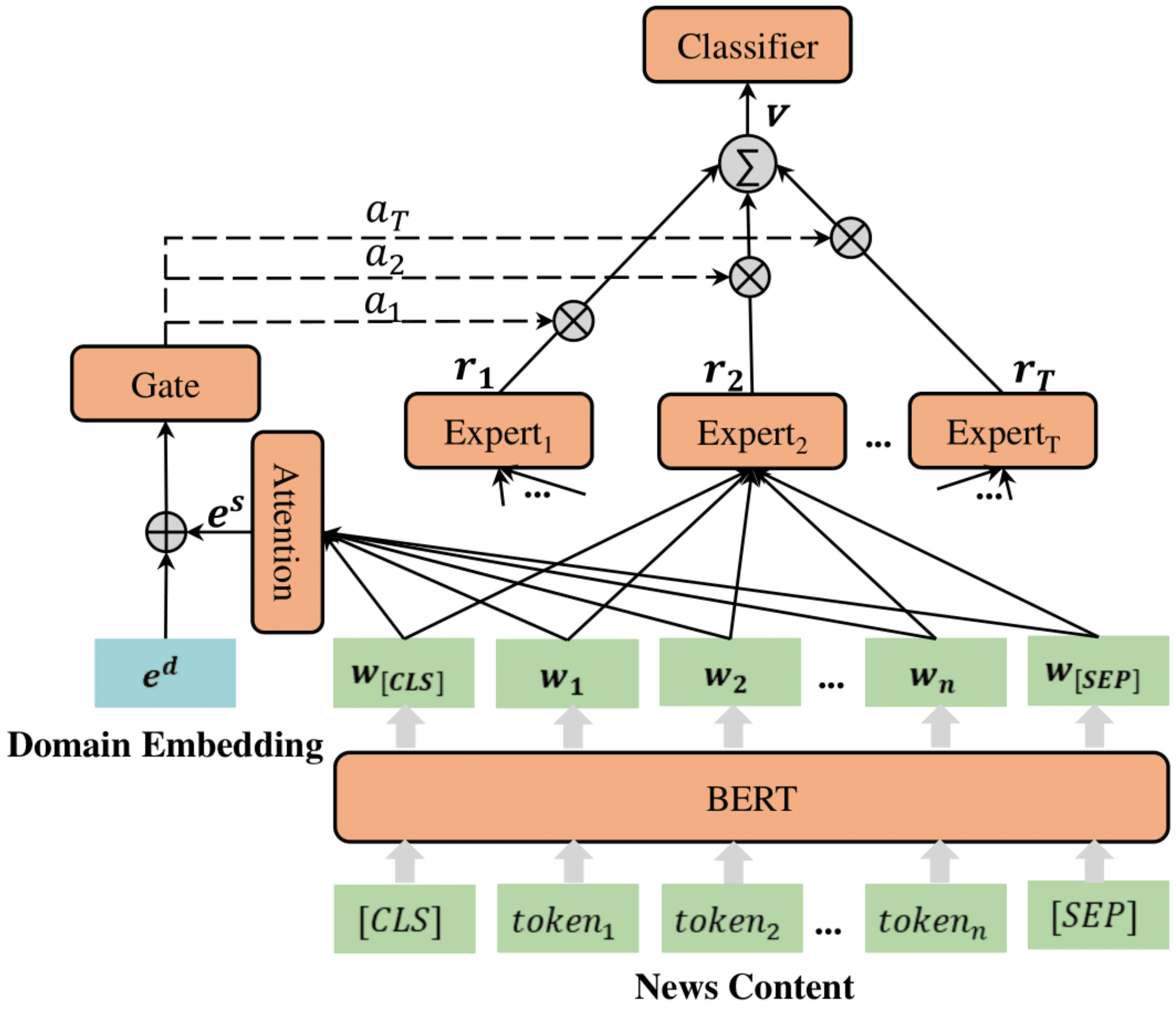
## Fake News Detection

- Earlier studies use hand-craft features.
- Some recent research works use propagation patterns for structural modeling.
- Others jointly used both textual and visual features for multi-modal modeling.
- Some methods incorporated related works to assist fake news detection.
- Silver et al. (AAAI'21) proposed to jointly preserve domain-specific as well as cross-domains, but they didn't make full use of the domain information explicitly.

# Methodology

## MDFEND

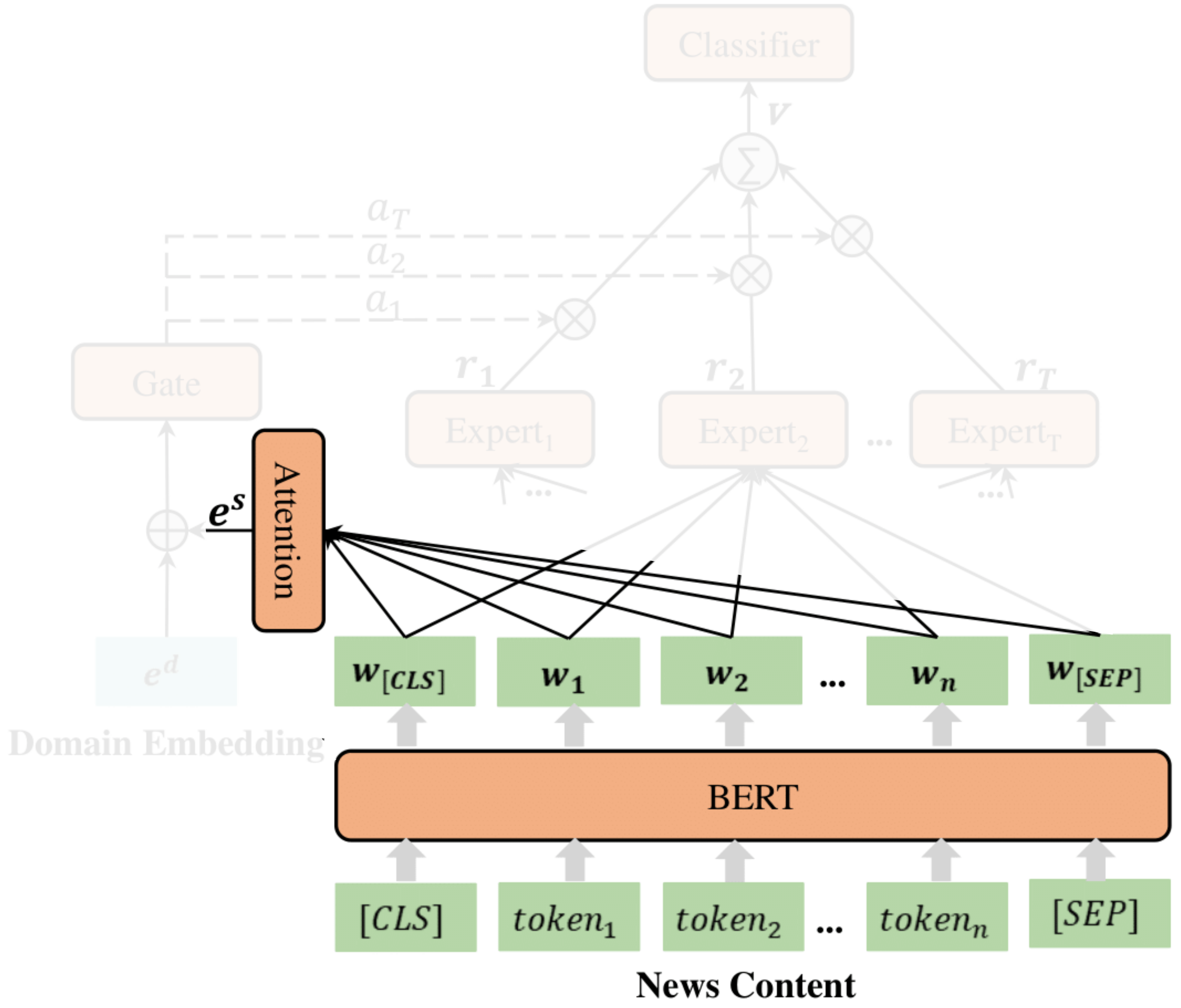
- Propose a novel framework, namely MDFEND, for MFND.
- Same as the single-domain methods, treat MFND as a binary classification problem.



# Methodology

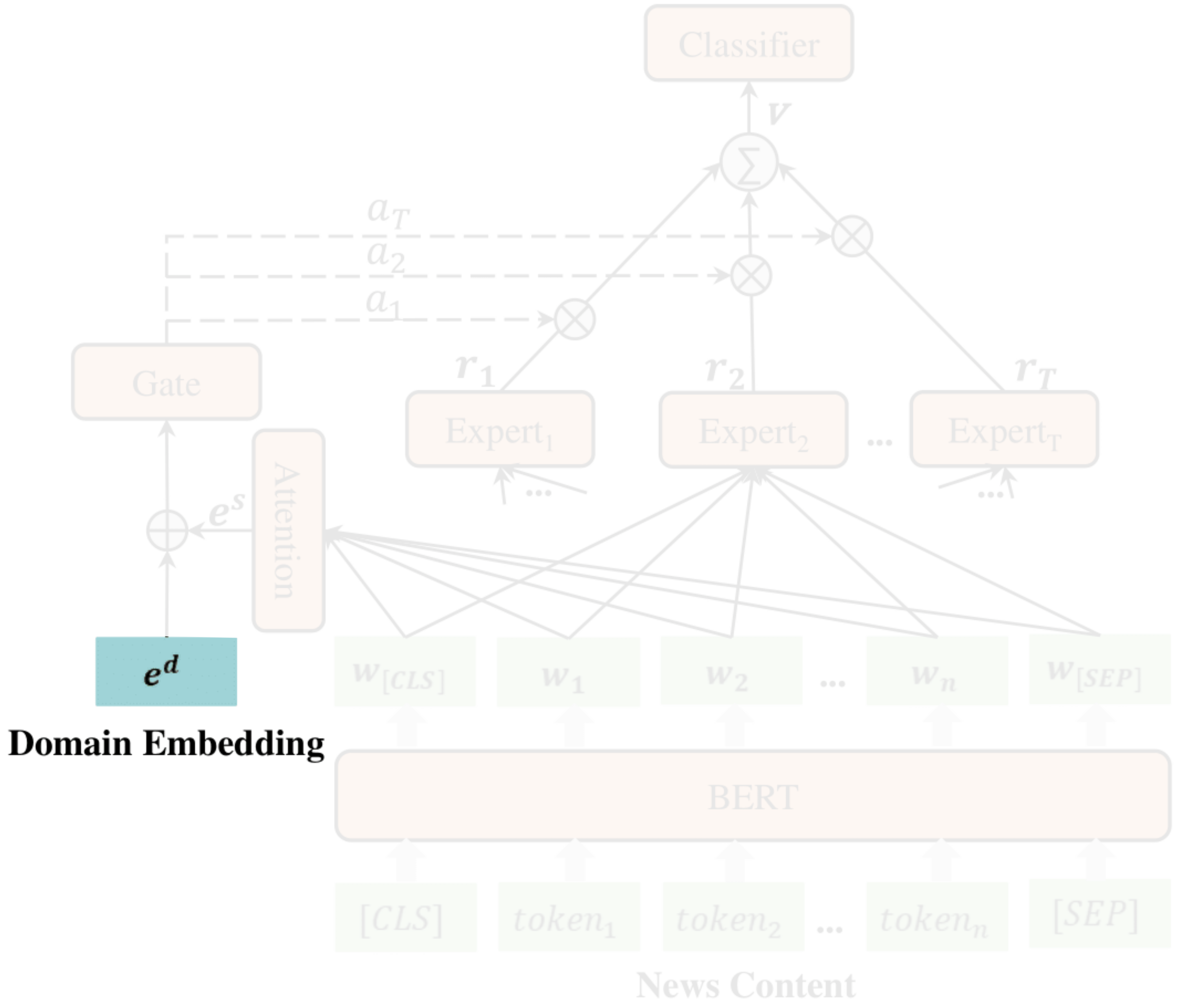
## Representation Extraction

- For a piece of news, tokenize its content with BertTokenizer.
- After adding special tokens for classification [CLS] as well as separation [SEP], obtain a list of token [`CLS`, token_1, ..., token_n, `SEP`].
- These tokens are then fed into BERT to obtain word embeddings $W$ are processing by a Mask-Attention network to get the sentence-level embedding $e^s$.



# Methodology
## Domain Embedding

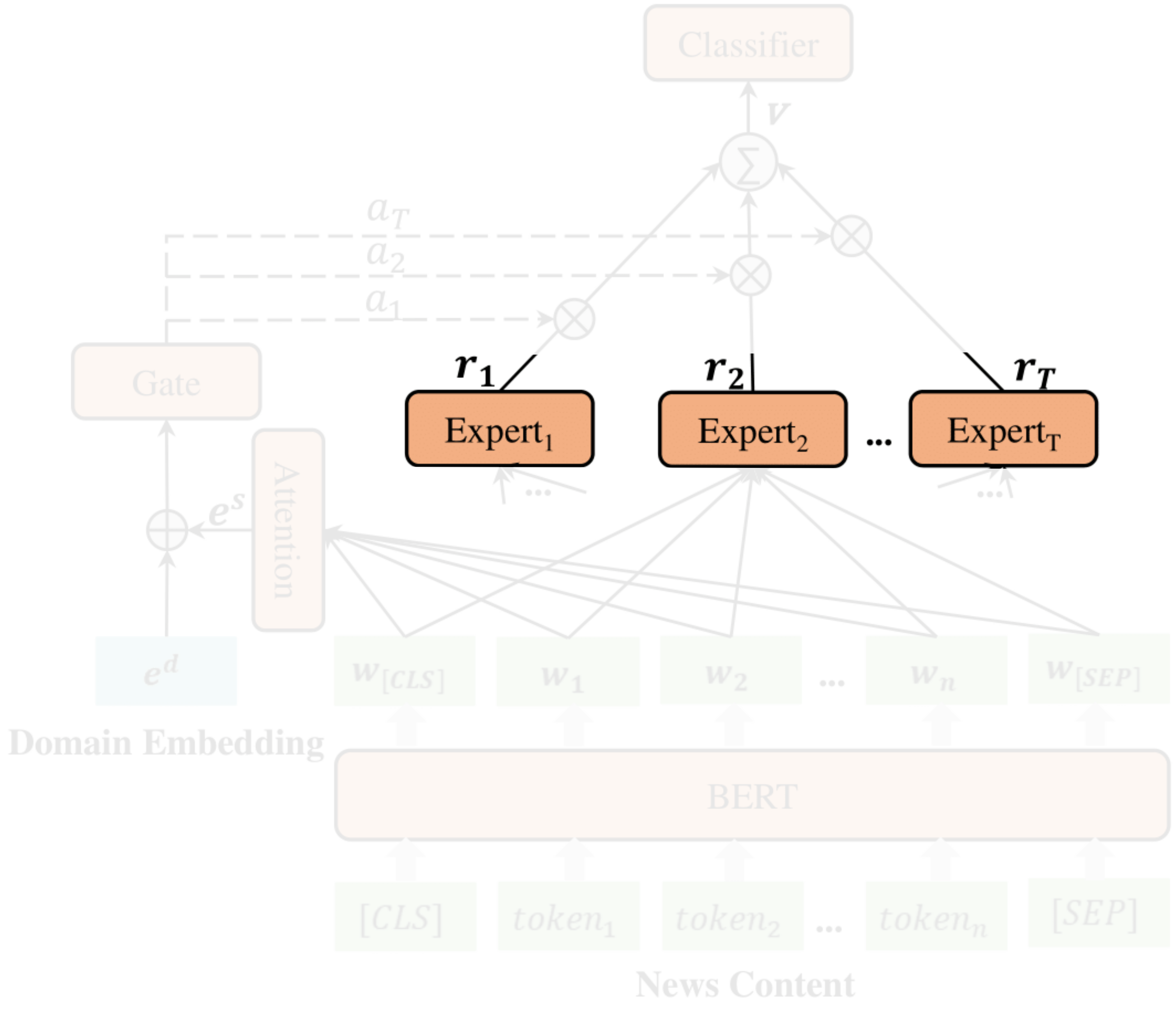
- To handle each domain specially, define a learnable vector $e^d$ → domain embedding.
- To help individualize representation extraction for each domain.
- Thus, a $e^d$ will be learned for each domain.



# Methodology

## Mixture of Expert

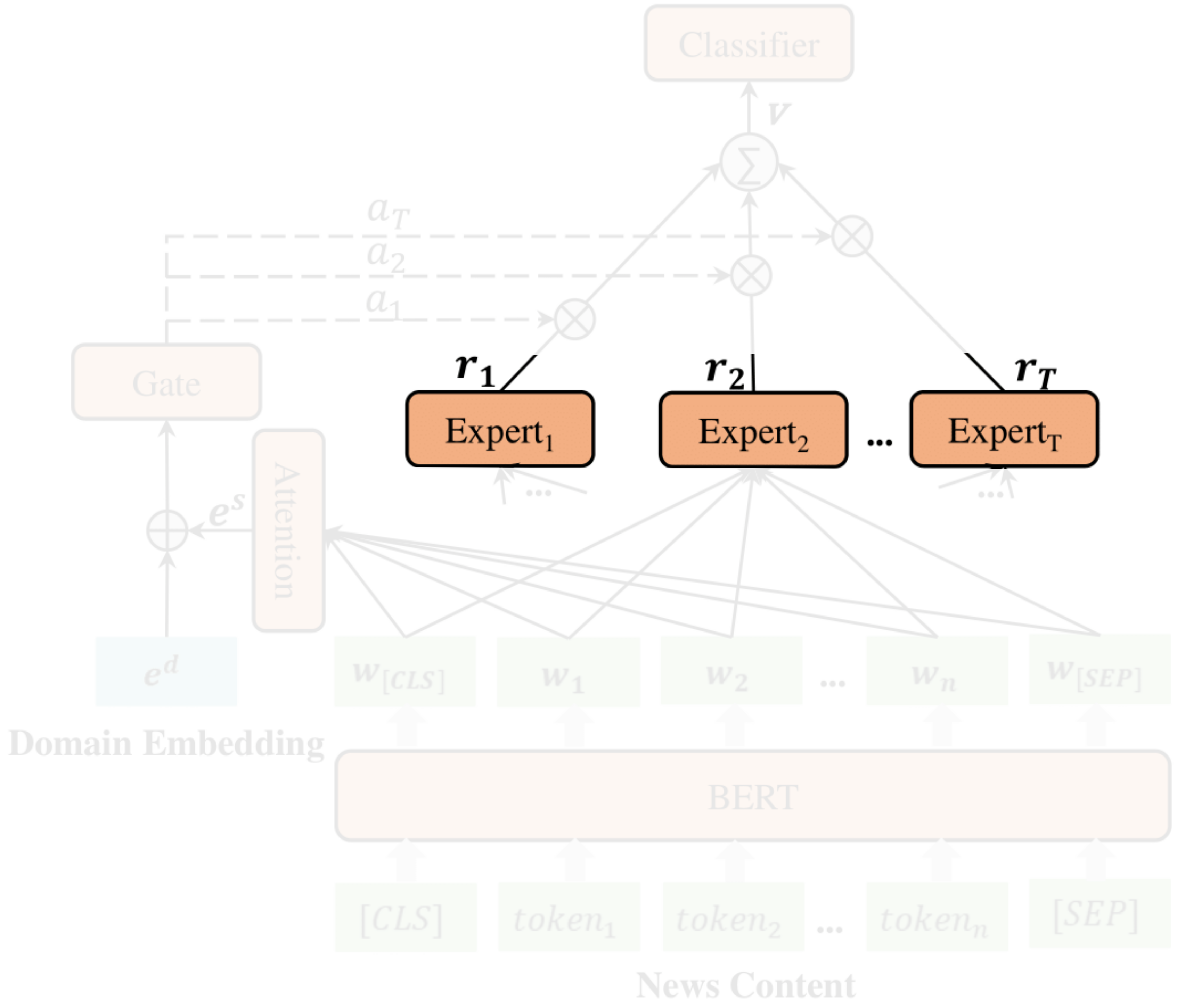
- Employ multiple experts (i.e., networks) to extract various representations of news.
- Intuitively, can employ one expert to extract the news' representations for multiple domains.
- However, one expert only specializes in one area.
  - Could only contain partial information.
  - Cannot completely cover the characteristic of news contents.



# Methodology

## Mixture of Expert

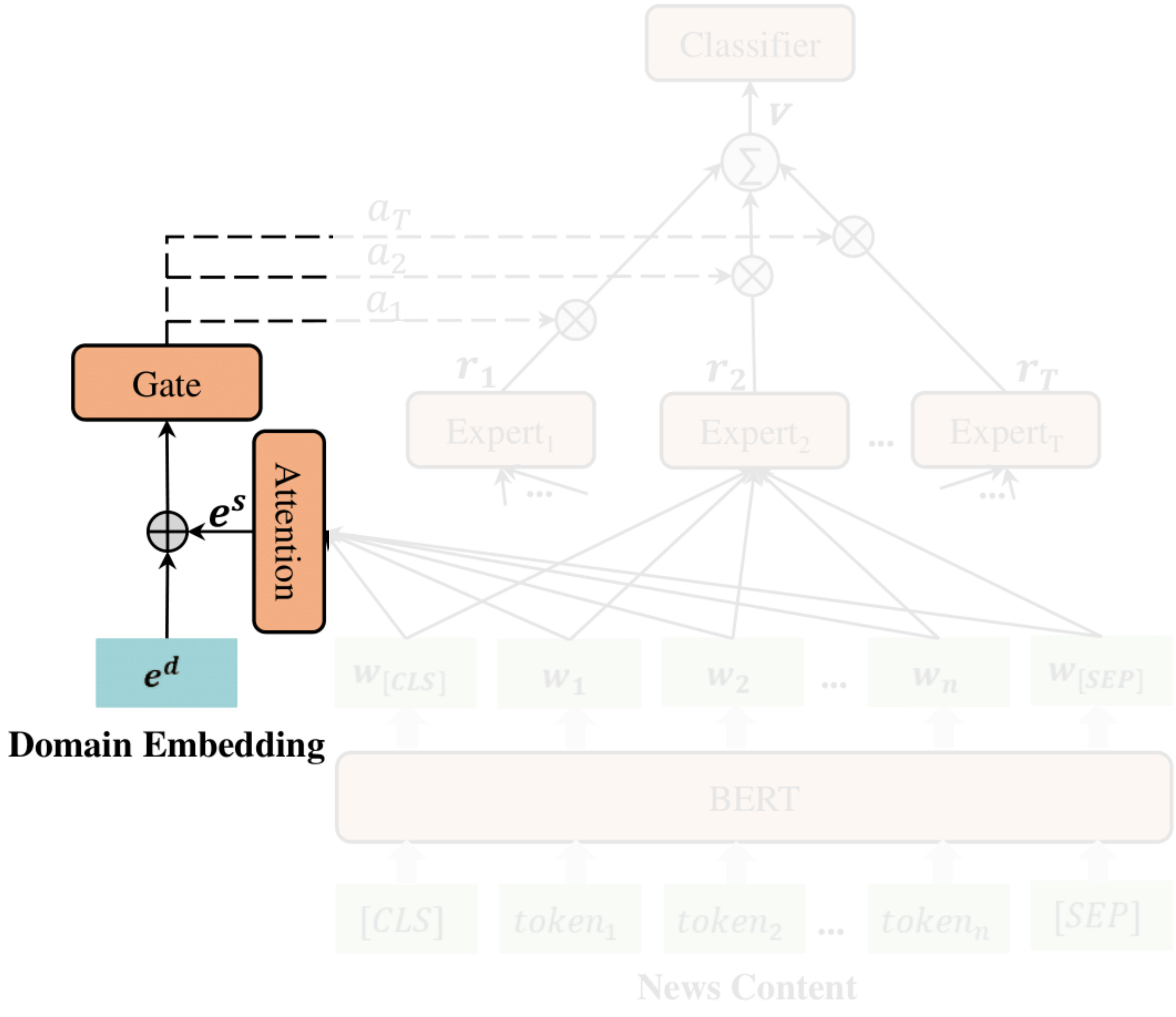
- Employ multiple experts for the sake of comprehensiveness.
- Output of the "expert" network:
  - $r_i = \Psi_i(W; \theta_i) \; (1 \leq i \leq T)$
  - $W$: word embeddings (input)
  - $T$: # of expert networks (hyper-parameter)
- Each "expert" network is a TextCNN in authors' design.



# Methodology

## Domain gate

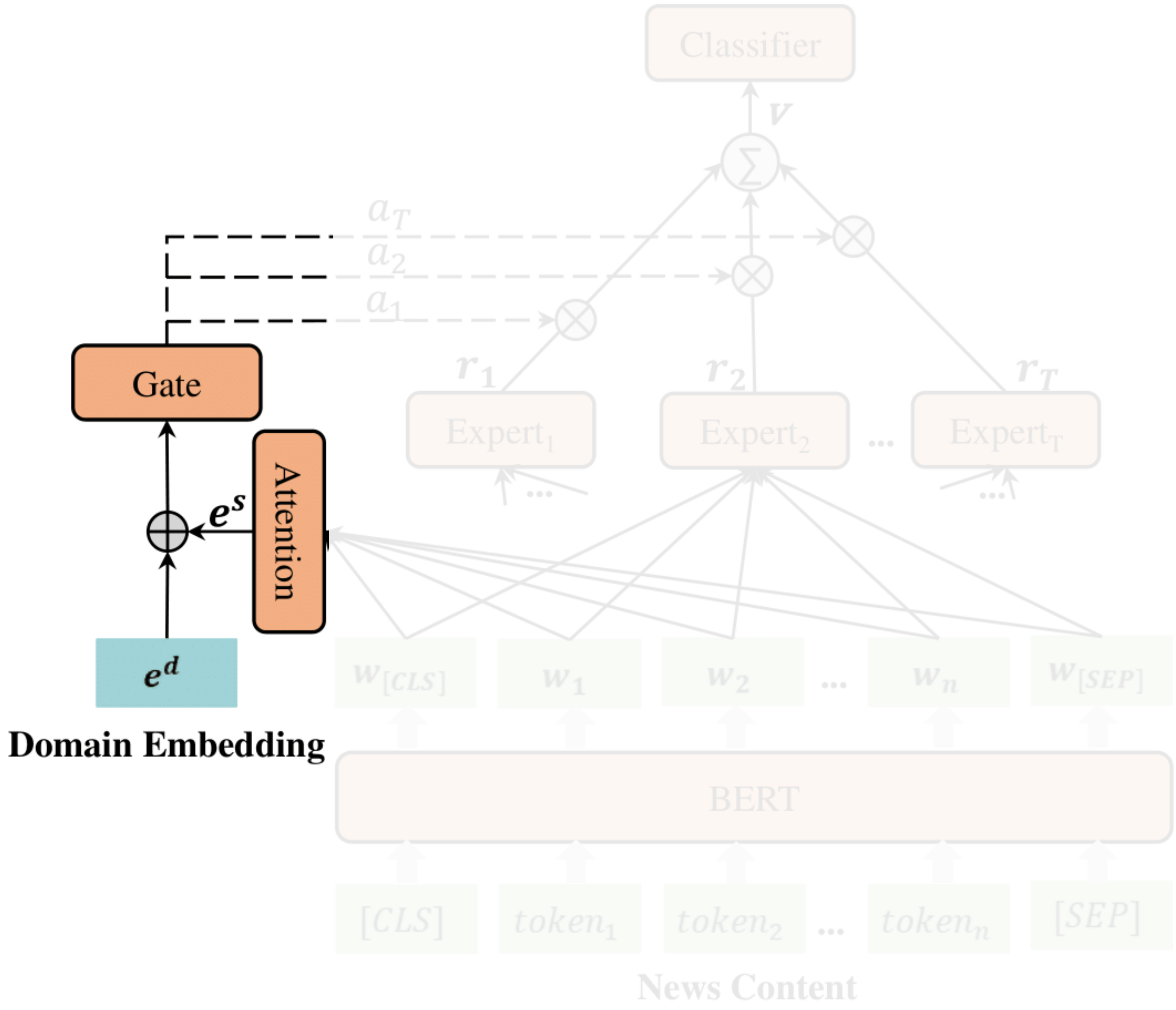
- It's necessary to generate high-quality news representations that can represent news from different domains appropriately.
- Intuitively, can average representations from all experts.
    - Simple average operation will remove the domain-specific information.
    - Synthetic representation may not be good for MFND.



# Methodology

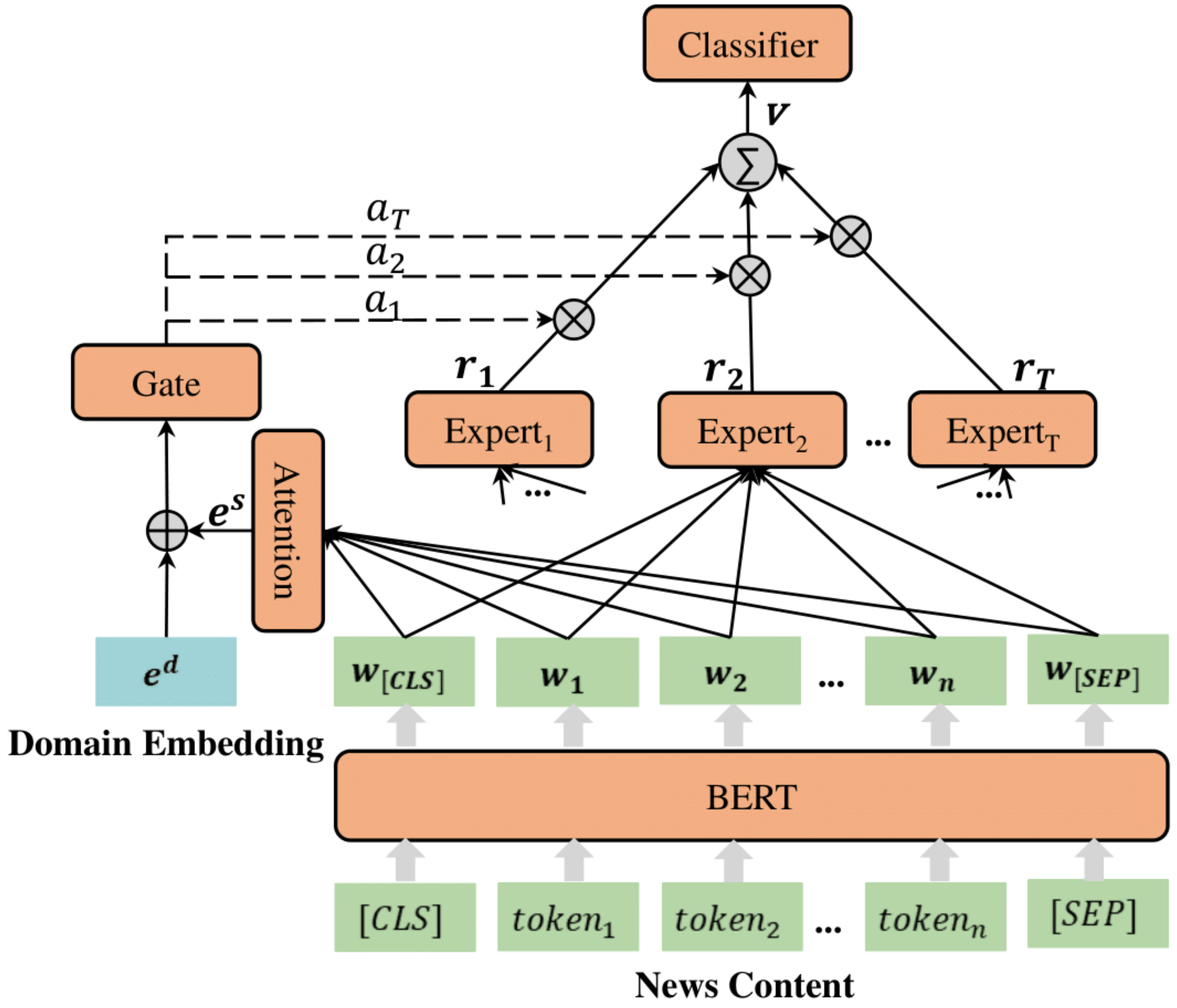
## Domain gate

- Different experts specialize in different areas, and they're good at handling different domains.
- For MFND, would like to select experts adaptively.
- Proposed a domain gate with $e^d$ & $e^s$ as input to guide the selection process.



# Methodology
## Domain gate

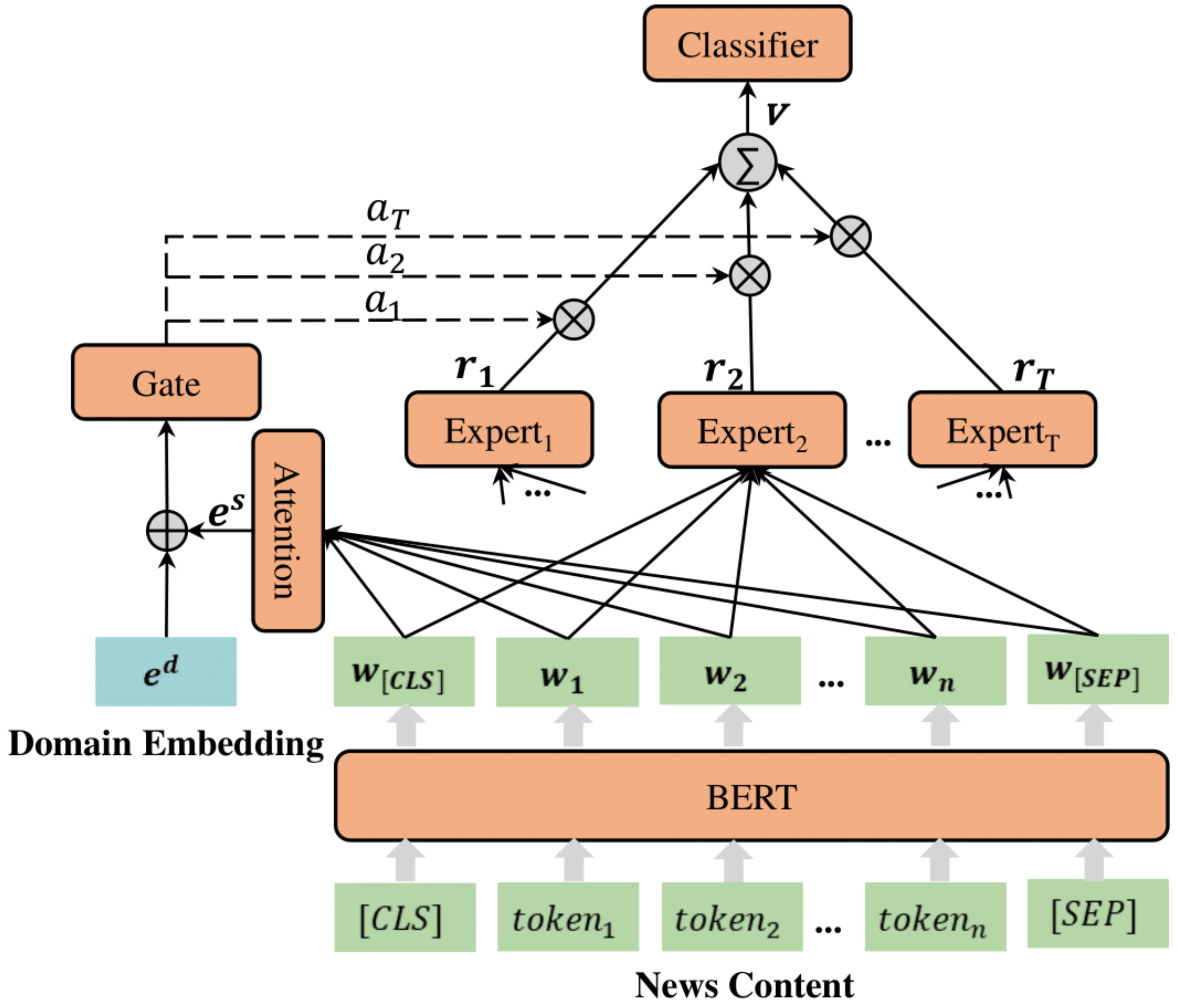
- Vector indicating the weight ratio of each expert.
    - $a = \mathrm{softmax}(G(e^d \oplus e^s; \phi))$
    - Use softmax function to normalize the output of $G(\cdot)$.
    - $a \in \mathbb{R}^n$: weight vector denoting the importance of different experts.
- With the domain gate, the news' final feature vector is obtained: $\boldsymbol{v} = \sum_{i=1}^{T} a_i \boldsymbol{r_i}$



# Methodology

## Prediction

- Final feature vector $\boldsymbol{v}$ is fed into the classifier:
  - MLP network + softmax output layer
  - $\hat{y} = \text{softmax}(MLP(v))$
- Employ binary cross-entropy loss:
  - $L = -\sum_{i=1}^{N} (y^i \log \hat{y}^i + (1 - y^i)\log(1 - \hat{y}^i))$



# Experiments

## Dataset: Weibo21

| domain | Science | Military | Education | Disasters | Politics |
|---|---|---|---|---|---|
| real | 143 | 121 | 243 | 185 | 306 |
| fake | 93 | 222 | 248 | 591 | 546 |
| all | 236 | 343 | 491 | 776 | 852 |

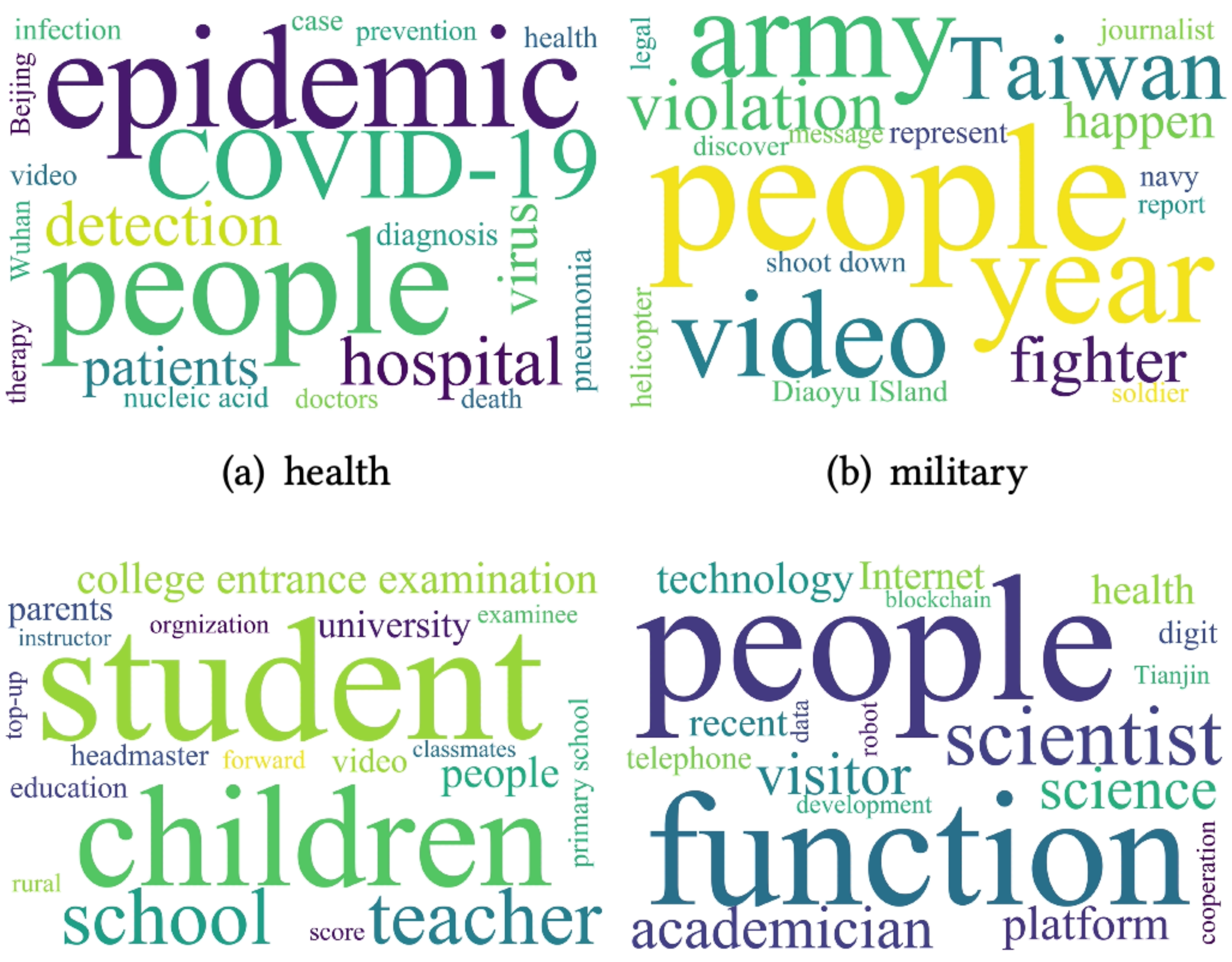
| domain | Health | Finance | Entertainment | Society | All |
|---|---|---|---|---|---|
| real | 485 | 959 | 1000 | 1198 | 4640 |
| fake | 515 | 362 | 440 | 1471 | 4488 |
| all | 1000 | 1321 | 1440 | 2669 | 9128 |

- Multi-domain fake news dataset in Chinese.
- News on Weibo ranging from Dec. 2014 ~ Mar. 2021.
- Collect news content, pictures, timestamp, comments.
  - Further, gather judgement information for fake news, provide evidence to people and increase the credibility of this dataset.
- Perform deduplication in one-pass clustering.
- 4488 : 4640 = fake : true news are obtained.

# Experiments

## Dataset: Weibo21

- Categorization on crowd-sourcing.
- Authors perform the topic distribution of news among different domains.
- Figure shows significant difference in the frequently used words.




(a) health (b) military

(c) education (d) science

# Experiments

## Baselines

- TextCNN, BiGRU, BERT:
  - Single-domain: train and test on specific single domain.
  - Mixed-domain: train model for all domains.
- EANN: to extract domain-independent features. (Use text branch only.)
- MMOE, MOSE: proposed for multi-task learning, assume that MFND are different tasks.
- EDDFN: a model for cross-domain FND, which models different domains implicitly and jointly preserves domain specific and cross-domain knowledge.

# Experiments

## Results (f1-score)

| model | Science | Military | Education | Accidents | Politics | Health | Finance | Entertainment | Society | All |
|---|---|---|---|---|---|---|---|---|---|---|
| TextCNN_single | 0.7470 | 0.778 | 0.8882 | 0.8310 | 0.8694 | 0.9053 | 0.7909 | 0.8591 | 0.8727 | 0.8380 |
| BiGRU_single | 0.4876 | 0.7169 | 0.7067 | 0.7625 | 0.8477 | 0.8378 | 0.8109 | 0.8308 | 0.6067 | 0.7342 |
| BERT_single | 0.8192 | 0.7795 | 0.8136 | 0.7885 | 0.8188 | 0.8909 | 0.8464 | 0.8638 | 0.8242 | 0.8272 |
| TextCNN_all | 0.7254 | 0.8839 | 0.8362 | 0.8222 | 0.8561 | 0.8768 | 0.8638 | 0.8456 | 0.8540 | 0.8686 |
| BiGRU_all | 0.7269 | 0.8724 | 0.8138 | 0.7935 | 0.8356 | 0.8868 | 0.8291 | 0.8629 | 0.8485 | 0.8595 |
| BERT_all | 0.7777 | 0.9072 | 0.8331 | 0.8512 | 0.8366 | 0.9090 | 0.8735 | 0.8769 | 0.8577 | 0.8795 |
| EANN | 0.8225 | 0.9274 | 0.8624 | 0.8666 | 0.8705 | 0.9150 | 0.8710 | 0.8957 | 0.8877 | 0.8975 |
| MMOE | **0.8755** | 0.9112 | 0.8706 | 0.8770 | 0.8620 | 0.9364 | 0.8567 | 0.8886 | 0.8750 | 0.8947 |
| MOSE | 0.8502 | 0.8858 | 0.8815 | 0.8672 | 0.8808 | 0.9179 | 0.8672 | 0.8913 | 0.8729 | 0.8939 |
| EDDFN | 0.8186 | 0.9137 | 0.8676 | 0.8786 | 0.8478 | 0.9379 | 0.8636 | 0.8832 | 0.8689 | 0.8919 |
| **MDFEND** | 0.8301 | **0.9389** | **0.8917** | **0.9003** | **0.8865** | **0.9400** | **0.8951** | **0.9066** | **0.8980** | **0.9137** |

- Mixed & multi-domain > single domain: additional data is great importance.
- Multi > mixed (in general): multi-domain learning is useful and necessary for MFND.

# Experiments

## Results (f1-score)

| model | Science | Military | Education | Accidents | Politics | Health | Finance | Entertainment | Society | All |
|---|---|---|---|---|---|---|---|---|---|---|
| TextCNN_single | 0.7470 | 0.778 | 0.8882 | 0.8310 | 0.8694 | 0.9053 | 0.7909 | 0.8591 | 0.8727 | 0.8380 |
| BiGRU_single | 0.4876 | 0.7169 | 0.7067 | 0.7625 | 0.8477 | 0.8378 | 0.8109 | 0.8308 | 0.6067 | 0.7342 |
| BERT_single | 0.8192 | 0.7795 | 0.8136 | 0.7885 | 0.8188 | 0.8909 | 0.8464 | 0.8638 | 0.8242 | 0.8272 |
| TextCNN_all | 0.7254 | 0.8839 | 0.8362 | 0.8222 | 0.8561 | 0.8768 | 0.8638 | 0.8456 | 0.8540 | 0.8686 |
| BiGRU_all | 0.7269 | 0.8724 | 0.8138 | 0.7935 | 0.8356 | 0.8868 | 0.8291 | 0.8629 | 0.8485 | 0.8595 |
| BERT_all | 0.7777 | 0.9072 | 0.8331 | 0.8512 | 0.8366 | 0.9090 | 0.8735 | 0.8769 | 0.8577 | 0.8795 |
| EANN | 0.8225 | 0.9274 | 0.8624 | 0.8666 | 0.8705 | 0.9150 | 0.8710 | 0.8957 | 0.8877 | 0.8975 |
| MMOE | **0.8755** | 0.9112 | 0.8706 | 0.8770 | 0.8620 | 0.9364 | 0.8567 | 0.8886 | 0.8750 | 0.8947 |
| MOSE | 0.8502 | 0.8858 | 0.8815 | 0.8672 | 0.8808 | 0.9179 | 0.8672 | 0.8913 | 0.8729 | 0.8939 |
| EDDFN | 0.8186 | 0.9137 | 0.8676 | 0.8786 | 0.8478 | 0.9379 | 0.8636 | 0.8832 | 0.8689 | 0.8919 |
| **MDFEND** | 0.8301 | **0.9389** | **0.8917** | **0.9003** | **0.8865** | **0.9400** | **0.8951** | **0.9066** | **0.8980** | **0.9137** |

- Some single-domain > mixed domain model: simply combining data different domains may result in negative effect from additional data.

# Experiments

## Results (f1-score)

| model | Science | Military | Education | Accidents | Politics | Health | Finance | Entertainment | Society | All |
|---|---|---|---|---|---|---|---|---|---|---|
| TextCNN_single | 0.7470 | 0.778 | 0.8882 | 0.8310 | 0.8694 | 0.9053 | 0.7909 | 0.8591 | 0.8727 | 0.8380 |
| BiGRU_single | 0.4876 | 0.7169 | 0.7067 | 0.7625 | 0.8477 | 0.8378 | 0.8109 | 0.8308 | 0.6067 | 0.7342 |
| BERT_single | 0.8192 | 0.7795 | 0.8136 | 0.7885 | 0.8188 | 0.8909 | 0.8464 | 0.8638 | 0.8242 | 0.8272 |
| TextCNN_all | 0.7254 | 0.8839 | 0.8362 | 0.8222 | 0.8561 | 0.8768 | 0.8638 | 0.8456 | 0.8540 | 0.8686 |
| BiGRU_all | 0.7269 | 0.8724 | 0.8138 | 0.7935 | 0.8356 | 0.8868 | 0.8291 | 0.8629 | 0.8485 | 0.8595 |
| BERT_all | 0.7777 | 0.9072 | 0.8331 | 0.8512 | 0.8366 | 0.9090 | 0.8735 | 0.8769 | 0.8577 | 0.8795 |
| EANN | 0.8225 | 0.9274 | 0.8624 | 0.8666 | 0.8705 | 0.9150 | 0.8710 | 0.8957 | 0.8877 | 0.8975 |
| MMOE | **0.8755** | 0.9112 | 0.8706 | 0.8770 | 0.8620 | 0.9364 | 0.8567 | 0.8886 | 0.8750 | 0.8947 |
| MOSE | 0.8502 | 0.8858 | 0.8815 | 0.8672 | 0.8808 | 0.9179 | 0.8672 | 0.8913 | 0.8729 | 0.8939 |
| EDDFN | 0.8186 | 0.9137 | 0.8676 | 0.8786 | 0.8478 | 0.9379 | 0.8636 | 0.8832 | 0.8689 | 0.8919 |
| **MDFEND** | 0.8301 | **0.9389** | **0.8917** | **0.9003** | **0.8865** | **0.9400** | **0.8951** | **0.9066** | **0.8980** | **0.9137** |

- MDFEND model performs better than other multi-domain models.
- By feeding both domain embedding and content to the gate, combining multiple domain softly is better the decouple domain-shared and domain-specific features roughly.

# Conclusions

- Study the problem of multi-domain fake news detection (MFND).
- Construct Weibo21, a MFND dataset. (First dataset collected richest domains.)
- Proposed a simple but effective method - MDFEND for MFND.
  - Utilizes domain gate to aggregate multiple representations extracted by mixture of expert.
- Evaluated MFND performance with different methods on Weibo21, experiments shows the effectiveness of MDFEND model.

# Comments

## of MDFEND

- Focus on multi-domain fake news detection.
- Use concept of mixture of expert.
- Proposed method is text-only.
- No detail information about domain embedding design.