Explicit Interaction Mechanism towards Text Classification

Cunxiao Du[1], Zhaozheng Chen[1], Fuli Feng[2], Lei Zhu[3], Tian Gan[1], Liqiang Nie[1]

[1]Shandong University
[2]National University of Singapore
[3]Shandong Normal University

AAAI 2019
Honolulu, Hawaii, USA

Overview

1. Background
   - Text Classification
   - Related Works
   - Essence of Baselines
2. Model
   - Interaction Mechanism
   - Our Model
3. Experiment
   - Multi-Class Datasets
   - Multi-Label Datasets
   - Convergence
   - Reproduction
4. Conclusion and Future Work

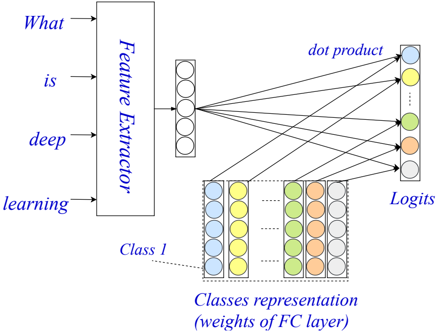
Background: Text Classification

- Text classification is one of the fundamental tasks in natural language processing, targeting at classifying a piece of text content into one or multiple categories.
- According to the number of desired categories, text classification can be divided into two groups, namely, multi-label (multiple categories) and multi-class (unique category).

Background: Related Works

- Supervised Models
  - Feature Engineering
  - Machine Learning Models
  - **Representation Learning Models**
    - Word-Based Models: Fasttext, LSTM, W.C. Region Embedding, etc.
    - Char-Based Models: VD-CNN, Char-CNN, Char-CRNN, etc.
- Unsupervised Models

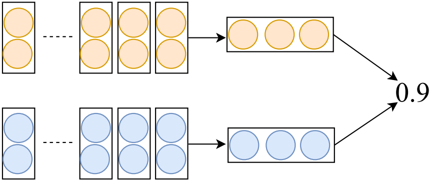
Background: Essence of Baselines

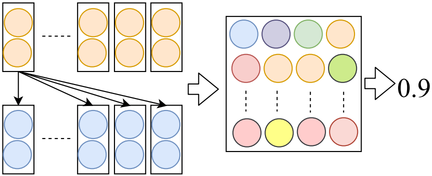


- [Press and Wolf EACL 2017] interprets the parameter matrix of the topmost FC layer as a set of class representations.
- Baseline is the same as the text matching model which performs matching between raw text input and the learned class representations.

Model: Interaction Mechanism of Text Matching

- Encoding-Based Methods
  - find vector representation for each sentence
  - classify the relation by using the concatenation of
    - two vector representation
    - absolute element-wise difference
    - element-wise product
- Interaction-Based Methods
  - use the interaction mechanism (e.g., dot product and element-wise multiplication) to model the similarity between each words
  - aggregate the similarity features into a scalar as the final prediction





Model: Our Model

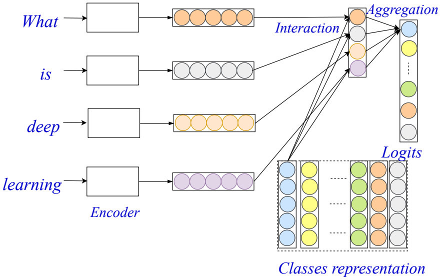


Figure: Illustration of proposed EXAM method with word-level matching.

- Encode Layer
  project the input text into a word-level representation
- Interaction Layer
  compute the interaction signals between the words and classes
- Aggregation Layer
  aggregate the interaction signals for each class and make the final predictions

Experiment: Multi-Class Datasets

Table: Test Set Accuracy [%] on multi-class document classification tasks.

Model	Amz. P.	Amz. F.	AG	Yah. A.	DBP
BoW	90.4	<u>54.6</u>	88.8	<u>68.9</u>	96.6
N-grams	<u>92.0</u>	54.3	92.0	68.5	98.6
N-grams TFIDF	91.5	52.4	<u>92.4</u>	68.5	<u>98.7</u>
Char-CNN	94.5	59.6	87.2	71.2	98.3
Char-CRNN	94.1	59.2	<u>91.4</u>	71.7	98.6
VDCNN	<u>**95.7**</u>	<u>**63.0**</u>	91.3	<u>73.4</u>	<u>98.7</u>
Small word CNN	94.2	56.3	89.1	70.0	98.2
Large word CNN	94.2	54.1	91.5	71.0	98.3
LSTM	93.9	59.4	86.1	70.8	98.6
Bigram-FastText	94.6	60.2	92.5	72.3	98.6
W.C RegionEmb	95.1	60.9	92.8	73.7	98.9
EXAM (Ours)	<u>95.5</u>	<u>61.9</u>	<u>**93.0**</u>	<u>**74.8**</u>	<u>**99.0**</u>

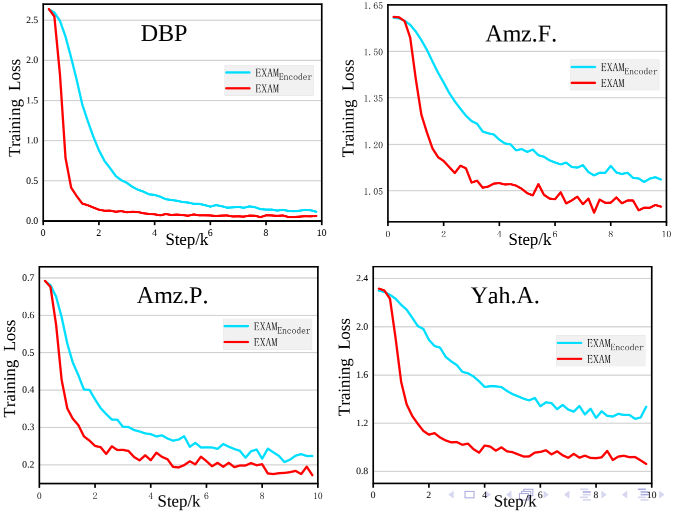
Experiment: Multi-Label Datasets

Table: Performance comparison between EXAM and baselines.

	Kanshan-Cup Dataset			Zhihu Dataset		
Model	Precision	Recall@5	F_1	Precision	Recall@5	F_1
Char-CNN	1.299	0.536	0.379	-	-	-
Char-TextRNN	1.304	0.537	0.380	-	-	-
Fasttext	1.325	0.546	0.387	1.235	0.564	0.387
TextCNN	1.331	0.550	0.389	1.241	0.566	0.389
Word-TextRNN	1.345	0.555	0.393	1.240	0.566	0.389
EXAM (Ours)	**1.360**	**0.561**	**0.397**	**1.267**	**0.578**	**0.397**

Experiment: Convergence

- We observed that EXAM converges faster than the one without EXAM with respect to all the datasets.



Experiment: Visualization

- To illustrate the effectiveness of explicit interaction, we visualized the interaction feature.

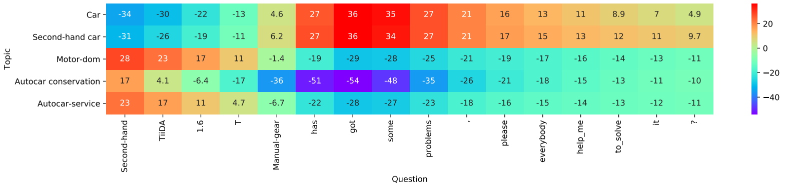


Figure: The visualization of interaction features of EXAM.

Experiment: Reproduction

Impelement Details

- Learning Framework & Environment: MXNet 1.2 & 1 TITAN Xp
- Optimizer & Learning Rate & Batch Size : Adam & 0.0001 & 16

Open-Source Website

`https://github.com/NonvolatileMemory/AAAI_2019_EXAM`

Tricks

- Small Batch Size is **VERY IMPROTANT**!
- More regularize methods (like dropout and L2 Norm) will make the results better!
- Pre-trained embedding like glove can help a lot.

Conclusion and Future Work

- A novel framework enhanced by interaction mechanism.
- Robust and reproducible experiments.
- Pay more attention to interaction feature.
- More complex interaction methods can be used.

Thanks for your attention!

Q & A

Feel free to contact me!

cnsdunm@gmail.com