# A Deep Relevance Model for Zero-Shot Document Filtering

Chenliang Li[1], Wei Zhou[2], Feng Ji[2], Yu Duan[1] and Haiqing Chen[2]
[1]School of Cyber Science and Engineering, WHU [2]Alibaba Group, China

## Overview:

➢**Document filtering** is the task to separate relevant documents from the irrelevant ones for a specific topic

➢**Zero-shot** means the instances of targeted categories are unseen during training phase

## Motivation and Objective:

➢In the era of big data, the potentially possible categories covered by documents would be limitless, we have to do filtering for **unseen** categories in many situations

➢We propose a novel **d**eep relev**a**nce model for **ze**ro-shot document filt**er**ing (named **DAZER**) by modeling the hidden feature interactions between the documents and category in the word embedding space

## Approach:

➢Interaction-based representation

- A category is represented by a set of seed words $|S_c|$, the embedding of category can be obtained as follow:

$$e_c = \frac{1}{|S_c|}\sum_{s \in S_c} e_s$$

- Interaction-based representation between word and category can be got as follow:

$$e_{c,w}^{diff} = e_c - e_w \ , \ e_{c,w}^{prod} = e_c \odot e_w$$
$$e_w^c = [e_w \oplus e_{c,w}^{diff} \oplus e_{c,w}^{prod}]$$

➢Relevance signal extraction

- Convolution with k-max pooling:

$$r_i = W_1 e_{i-l:i+l}^c + b_1 \ , \ r_d = [r_{k-max}^1 \oplus r_{k-max}^2 \oplus \cdots \oplus r_{k-max}^m]$$

- Category-specific gate mechanism is applied after convolution operation to control the information flow:

$$a_c = \sigma(W_2 e_c + b_2) \ , \ r_i = a_c \odot (W_1 e_{i-l:i+l}^c + b_1)$$

➢Relevance aggregation

- Two fully-connected layers are applied to get relevance score:

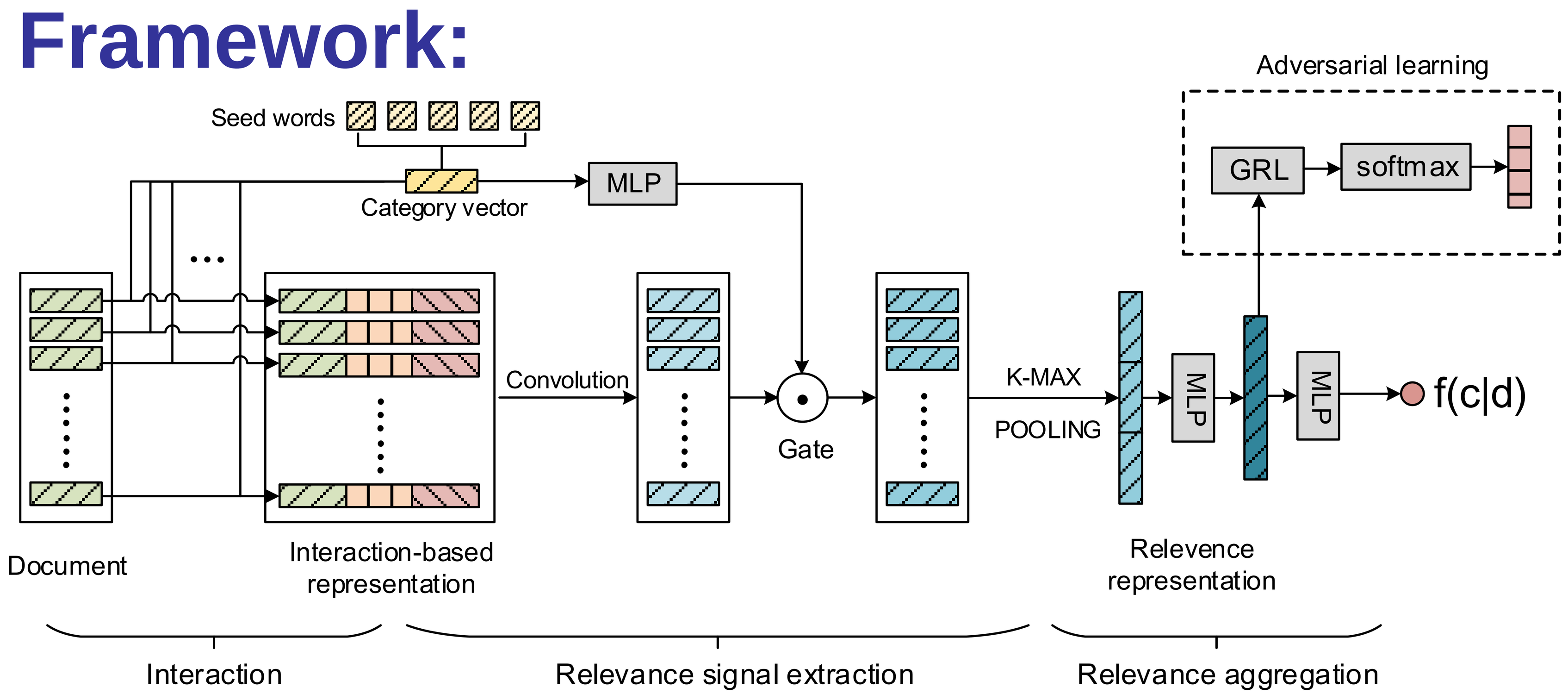
$$h_{c,d} = g_a(W_3 r_{c,d} + b_3) \ , \ f(c|d) = \tanh(W^T h_{c,d} + b)$$

➢Adversarial learning

- To ensure the extracted hidden feature $h_{c,d}$ contain no category-specific information, we introduce an adversarial classifier over hidden feature $h_{c,d}$:

$$p_{cat}(\cdot|h_{c,d}) = softmax(W_4 h_{c,d} + b_4)$$

## Framework:




## Experimental Results

➢Dataset: 20-Newsgroup (18,846 documents) and Movie Review (5,006 movie reviews)

➢Mean average precision

| Dataset | Category | DAZER | DRMM | K-NRM | DeepRank | DSSM | SSVM | BM25 |
|---|---|---|---|---|---|---|---|---|
| 20NG | pc | 0.535 | 0.382† | 0.369† | 0.144† | 0.222† | 0.117 | 0.313 |
| | med | 0.826 | 0.662† | 0.645† | 0.033† | 0.192† | 0.104 | 0.403 |
| | baseball | 0.764 | 0.731† | 0.735† | 0.294† | 0.373† | 0.291 | 0.414 |
| | space | 0.780 | 0.593† | 0.671† | 0.285† | 0.142† | 0.140 | 0.641 |
| | med-space | 0.805 | 0.640† | 0.666† | 0.101† | 0.174† | 0.122 | 0.522 |
| | atheism-electronics | 0.464 | 0.242† | 0.346† | 0.418† | 0.219† | 0.132 | 0.263 |
| | christian-mideast | 0.712 | 0.662† | 0.657† | 0.298† | 0.327† | 0.161 | 0.579 |
| | baseball-hockey | 0.782 | 0.642† | 0.736† | 0.332† | 0.135† | 0.438 | 0.444 |
| | pc-windowsx-electronics | 0.489 | 0.274† | 0.379† | 0.183† | 0.278† | 0.120 | 0.314 |
| Movie Review | very negative | 0.290 | 0.119† | 0.114† | 0.097† | 0.216† | 0.080 | 0.134 |
| | negative | 0.807 | 0.528† | 0.557† | 0.423† | 0.478† | 0.236 | 0.090 |
| | neural | 0.798 | 0.764† | 0.749† | 0.686† | 0.678† | 0.365 | 0.007 |
| | Positive | 0.862 | 0.696 | 0.706† | 0.655† | 0.753† | 0.300 | 0.090 |
| | very positive | 0.479 | 0.250† | 0.339† | 0.217† | 0.271† | 0.063 | 0.066 |
| Avg | | 0.671 | 0.513 | 0.548 | 0.298 | 0.318 | 0.191 | 0.306 |

➢Ablation test

| Setting | very negative | negative | neural | positive | very positive |
|---|---|---|---|---|---|
| **DAZER** | 0.29 | 0.807 | 0.798 | 0.862 | 0.479 |
| **-$e_{c,w}^{diff}$** | 0.246 | 0.773 | 0.776 | 0.847 | 0.453 |
| **-$e_{c,w}^{prod}$** | 0.258 | 0.779 | 0.785 | 0.847 | 0.43 |
| **-Gate** | 0.278 | 0.755 | 0.785 | 0.848 | 0.429 |
| **-Adv** | 0.261 | 0.779 | 0.776 | 0.827 | 0.444 |

➢Zero-shot document filtering is more like a semantic matching task

➢Proposed DAZER significantly achieves much better filtering performance on all tasks

➢Each component contributes significantly positive for the filtering performance