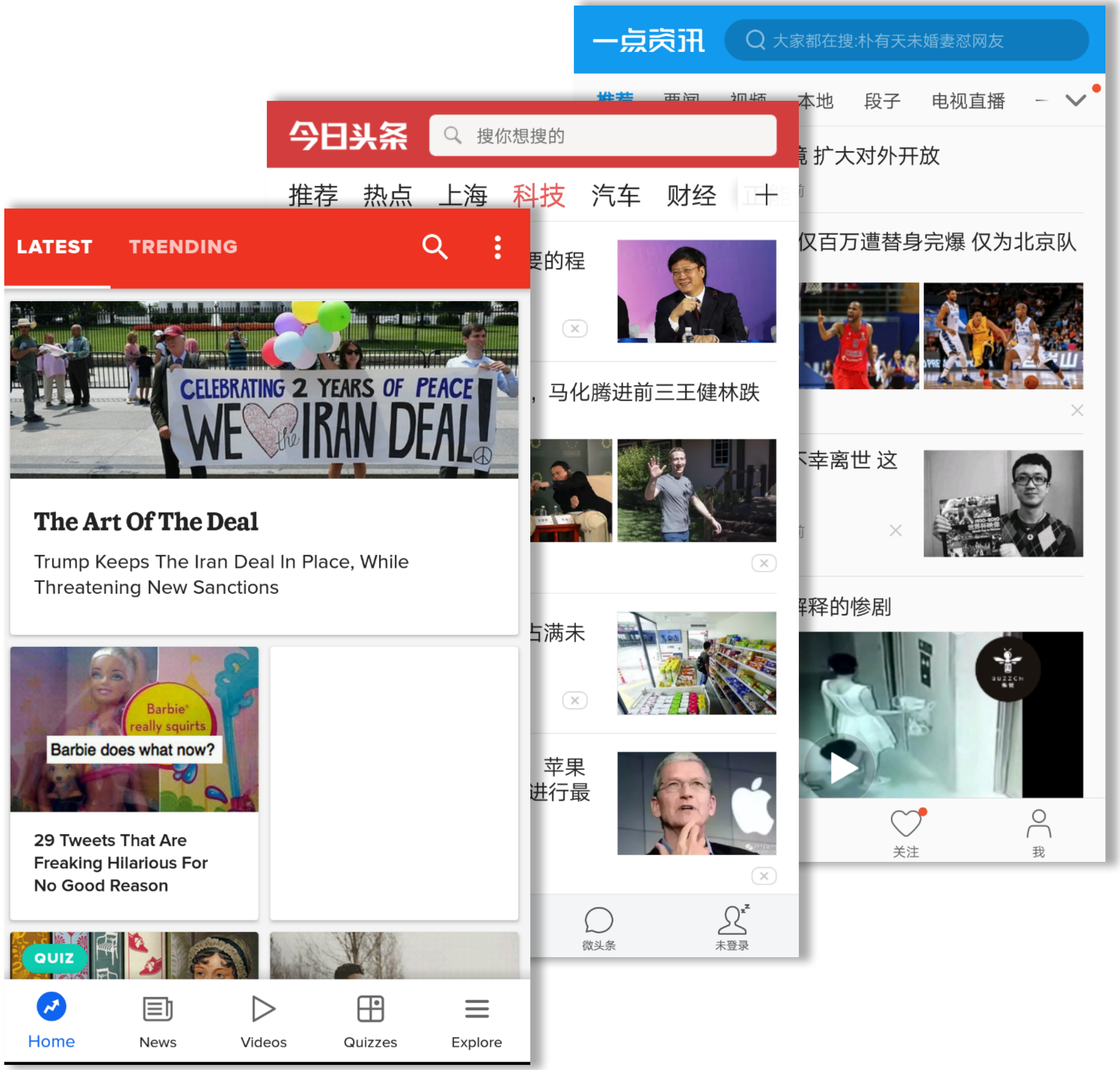
Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration

Xuejian Wang, Lantao Yu, Kan Ren, Guanyu Tao, Weinan Zhang, Jun Wang, Yong Yu

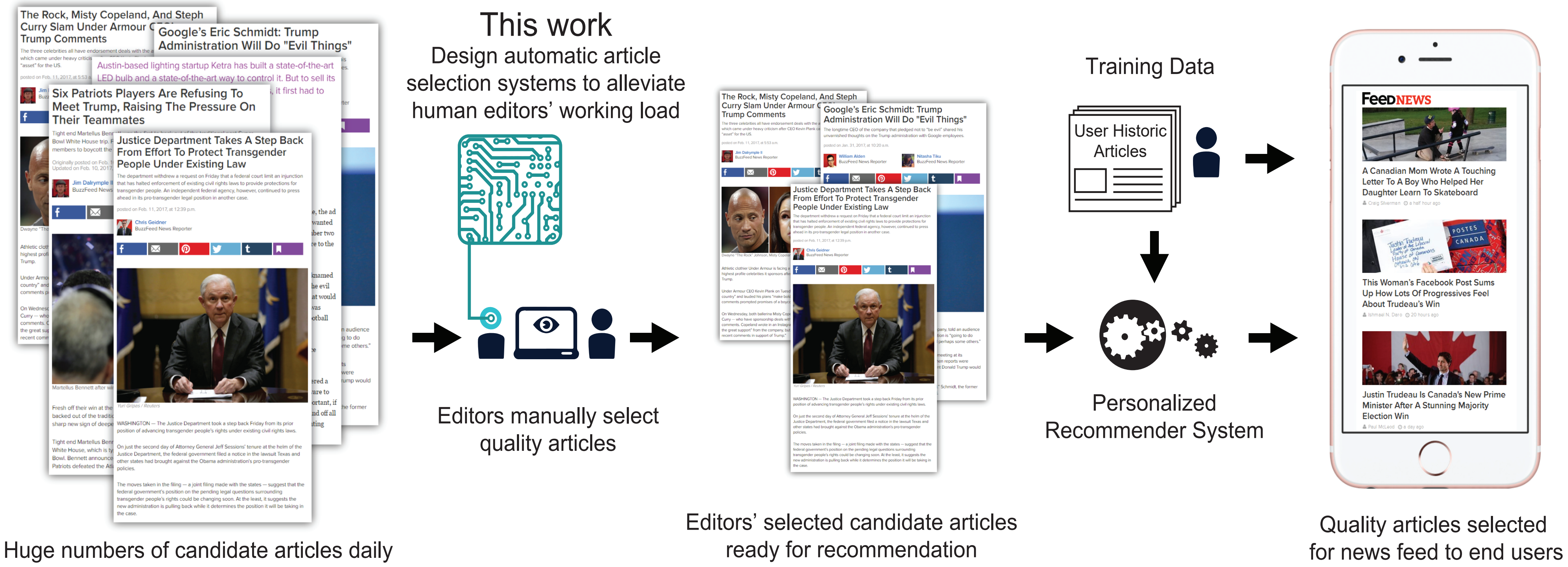
xjwang@apex.sjtu.edu.cn

Streaming Media

- Toutiao
- Buzzfeed
- Yidian
- 36kr
- …



Underlying System of Streaming Media



- How they choose and recommend?

An Example

场景数据

sina 新浪科技

韩国初创公司神奇智能表带 拿手指当电话打

2016年09月19日 15:22　　新浪科技

新浪美股北京时间19日讯　一家从三星分离出来的初创公司开发出了一款全新的智能手表表带，据称能够让人们用自己的手指来接听电话。

自从本月登上众筹网站Kickstarter以来，他们的表带产品Sgnl就风靡了各个科技博客，被认为将开创一种全新的通信方式。这种技术很容易让人联想起间谍影片和科幻小说，甚至被一位博客写手誉为“你梦想中的科幻极件”。业界的兴奋与激动帮助智能表带的开发者，首尔的Innomdle Lab迅速筹集到了大量资金，Quartz报道称，至报道刊发的9月17日，已经超过了86万美元，虽然筹资周期还剩21天，但是5万美元的目标早已轻松超越。该公司预计2017年2月就能交付第一批Sgnl。

→

标准化输入数据

域	类别或序列取值
来源城市	北京
星期	星期一
小时	15点
媒体来源	新浪科技
文章类别	科技
文章标题	韩国初创公司神奇智能表带拿手指当电话打
文章内容	一家从三星分离出来的初创公司开发出了一款全新的智能手表表带，据称能够让人们用自己的手指来接听电话。自从本月登上众筹网站Kickstarter以来...

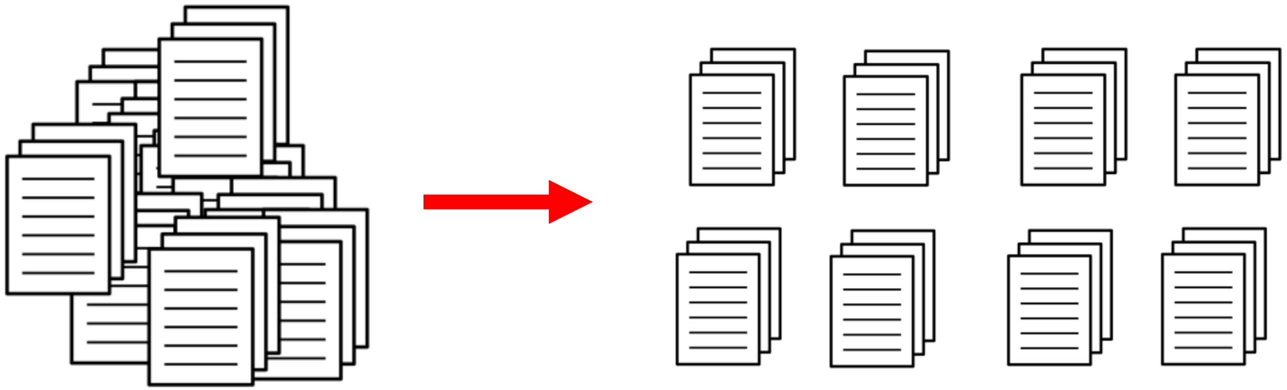
→

预测

文章是否应该入选当日推送

Particularity in This Problem

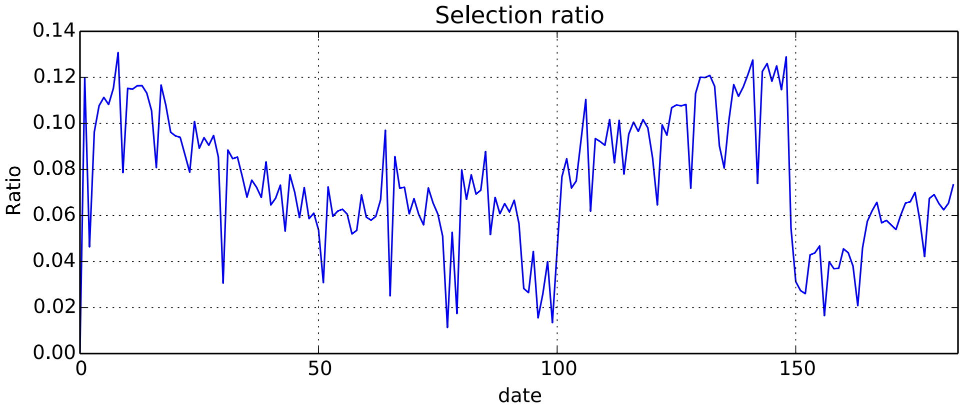
- The underlying criteria for the editors’ selection is non-explicit
  - Attractiveness
  - Humor
  - Stringency



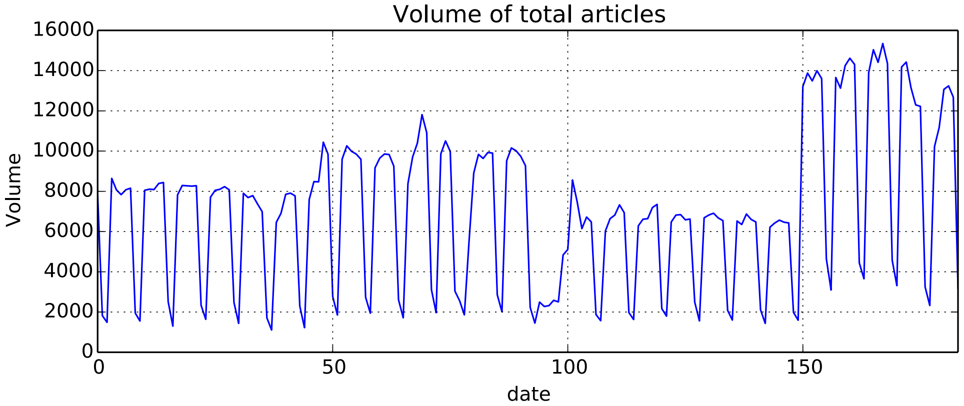
- The crawled or submitted article data distribution and the editors’ article selection behavior on the data are non-stationary
  - Drift of data distribution
  - The editors’ preference varies

Dynamic Characteristics: Some Examples

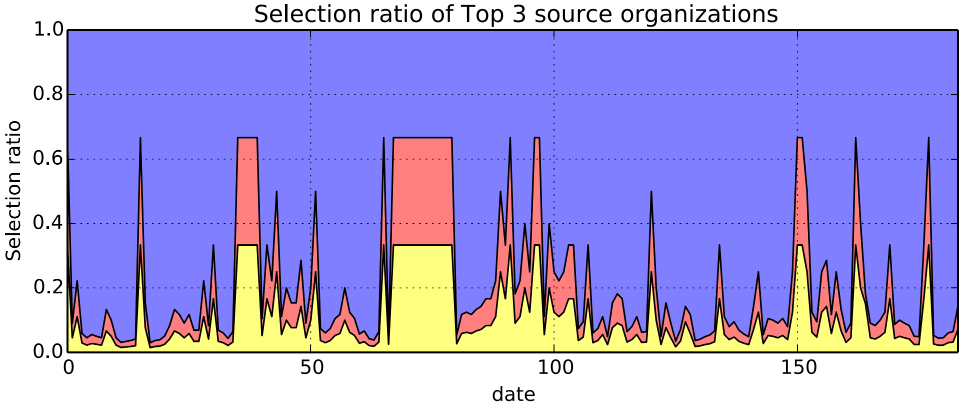
The change of the number of total submitted articles over time:



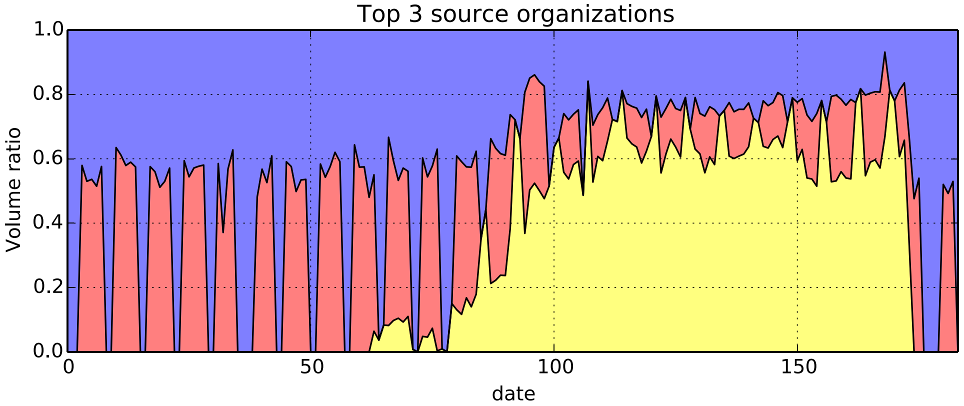
The change of the editors' selection ratio over time:

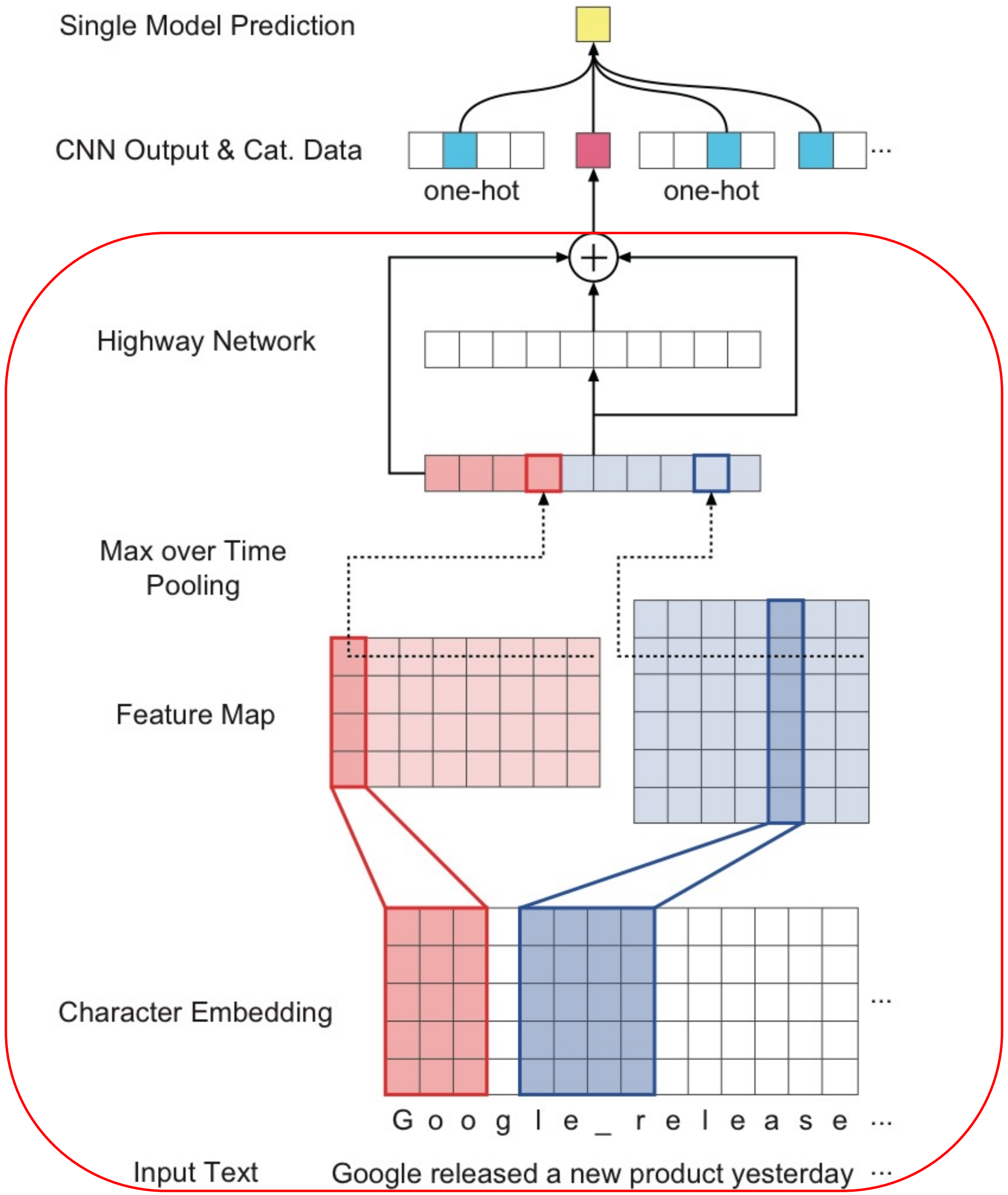


The change of the total volume of 3 main source organizations:



The change of the selection ratio of 3 main source organizations:





Architecture for Single Model Prediction

- The raw textual input of the document
  $\boldsymbol{x} = \{c_1, c_2, \ldots, c_l\}$
- Character embedding set $\mathcal{E} \in \mathbb{R}^{d \times |C|}$
- Embedding function $\Pi : c \longrightarrow \mathbf{e} \in \mathcal{E}$
- Feature map matrix
  $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_k] \in \mathbb{R}^{(l-w+1) \times k}$
- Final part: Highway Network

$$\eta = \sigma(\mathbf{W}_q^H \cdot \mathbf{x}_q + \mathbf{b}_q^H)\,, q \in [1, n]\,,$$
$$\mathbf{x}_{q+1} = \eta \cdot g(\mathbf{W}_q \cdot \mathbf{x}_q + \mathbf{b}_q) + (1 - \eta)\mathbf{x}_q\,,$$
$$o_{\mathrm{D}} = \sigma(\mathbf{w}_{\mathrm{D}} \cdot \mathbf{x}_n + \mathbf{b}_{\mathrm{D}})\,,$$

Attention Mechanism

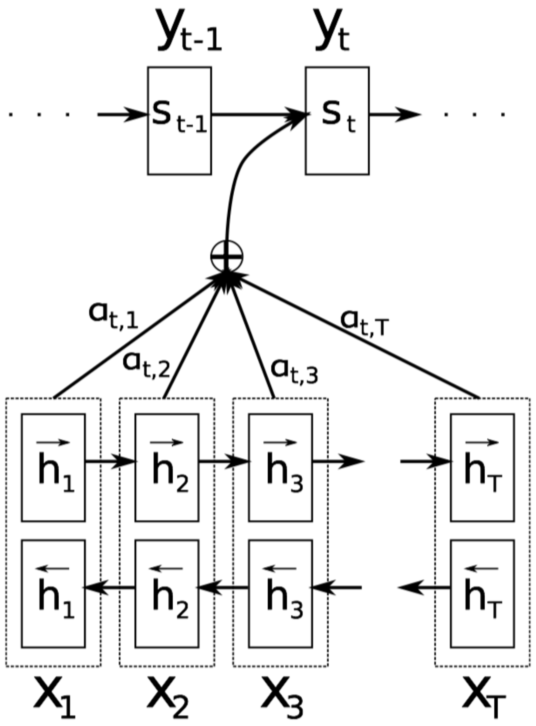


Figure 1: The graphical illustration of the proposed model trying to generate the t-th target word y_t given a source sentence $(x_1, x_2, \ldots, x_T)$.

- Auto-regressive RNN

$$p(y_i | y_1, \ldots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i)$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

- Attention Mechanism

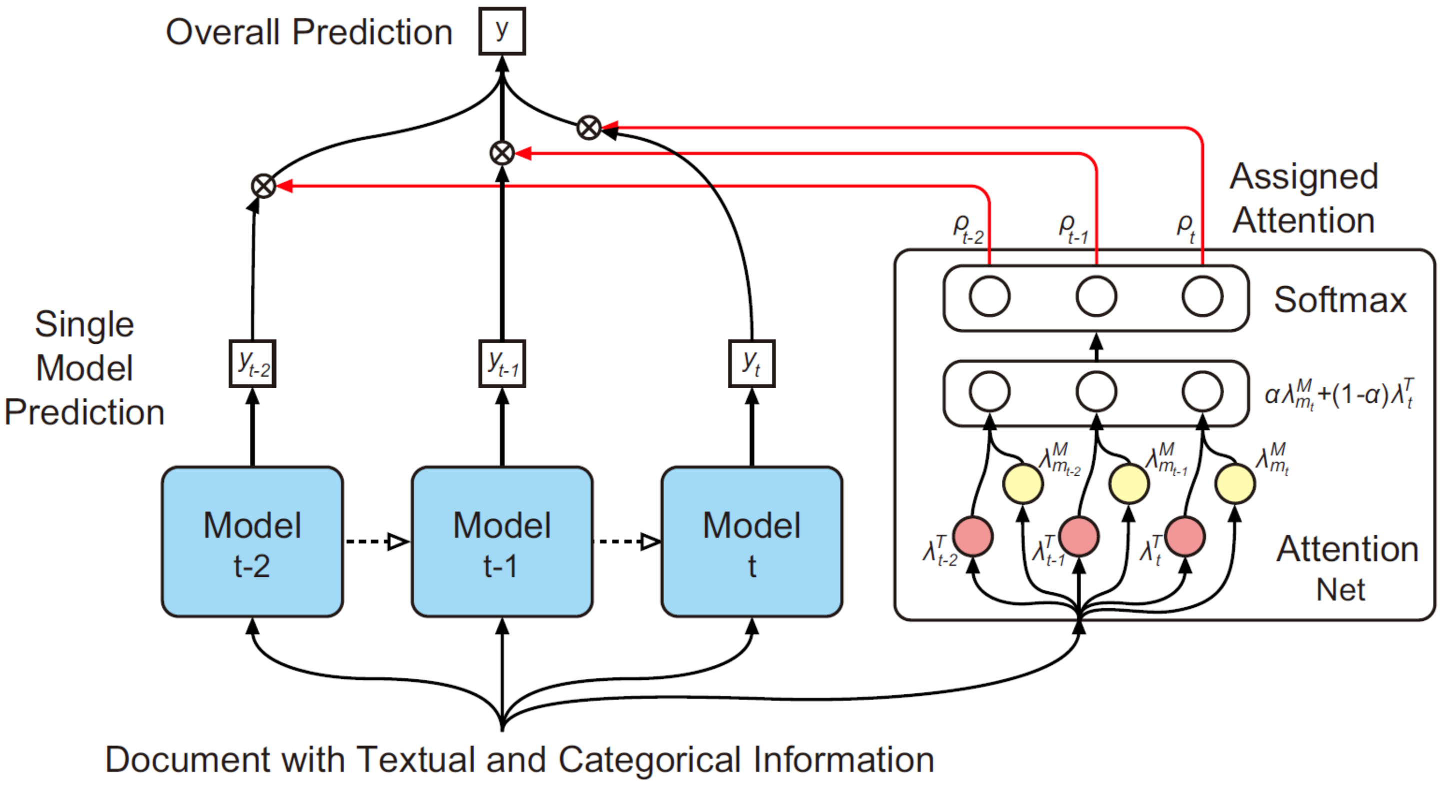
$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$e_{ij} = a(s_{i-1}, h_j)$$

- Neural Machine Translation By Jointly Learning To Align and Translate, Bahdanau, Dzmitry et al., ICLR 2015

Dynamic Attention Deep Model (DADM)



$$\mathbf{o} = [\mathbf{o}_1^{\top} \oplus \mathbf{o}_2^{\top} \oplus, \ldots \oplus \mathbf{o}_{|\mathcal{S}|}^{\top} \oplus \mathbf{o}_{\mathbf{D}}^{\top}]^{\top} \in \mathbb{R}^{l_1+l_2+\ldots+l_{|\mathcal{S}|}+1}$$

$$\hat{y} = \Pr(\mathrm{y}_i | d_i^t; \mathbb{D}_t) = \sigma(\mathbf{w} \cdot \mathbf{o} + b)$$

$$\lambda^{\mathcal{M}}_{m_t} = \mathbf{w}^{\mathcal{M}}_{m_t} \cdot \mathbf{o} + b^{\mathcal{M}}_{m_t} \,,$$

$$\lambda^{\mathcal{T}}_t = \mathbf{w}^{\mathcal{T}}_t \cdot \mathbf{o} + b^{\mathcal{T}}_t \,,$$

$$\rho_t = \mathrm{softmax}(\alpha\lambda^{\mathcal{M}}_{m_t} + (1-\alpha)\lambda^{\mathcal{T}}_t) = \frac{\exp(\alpha \cdot \lambda^{\mathcal{M}}_{m_t} + (1-\alpha) \cdot \lambda^{\mathcal{T}}_t)}{\sum_{\tau \in [0,K]} \exp(\alpha \cdot \lambda^{\mathcal{M}}_{m_\tau} + (1-\alpha) \cdot \lambda^{\mathcal{T}}_\tau)}$$

$$\hat{y} = \Pr(y_i | d_i^{t_0}; \mathbb{D}_{t_0}) = \sum_{\tau=t_0-K+1}^{t_0} \rho_\tau \cdot \hat{y}_\tau$$

- Model specialty
- Timeliness

Experiments

The data statistics over 9 tested days: *

Date	Articles	Selected	Authors	Websites	Orgs.
10-01	2,233	126	499	84	211
10-02	1,449	41	299	62	178
10-03	2,494	66	200	65	200
10-04	2,275	101	365	65	190
10-05	2,319	36	407	68	194
10-06	2,582	67	412	75	186
10-07	2,504	100	488	69	193
10-08	4,837	65	974	122	560
10-09	5,109	228	1,088	137	594
Overall	25,802	830	4,732	747	2,506

* The experiments are based on ULU Technologies article filtering API platform.

Compared Settings

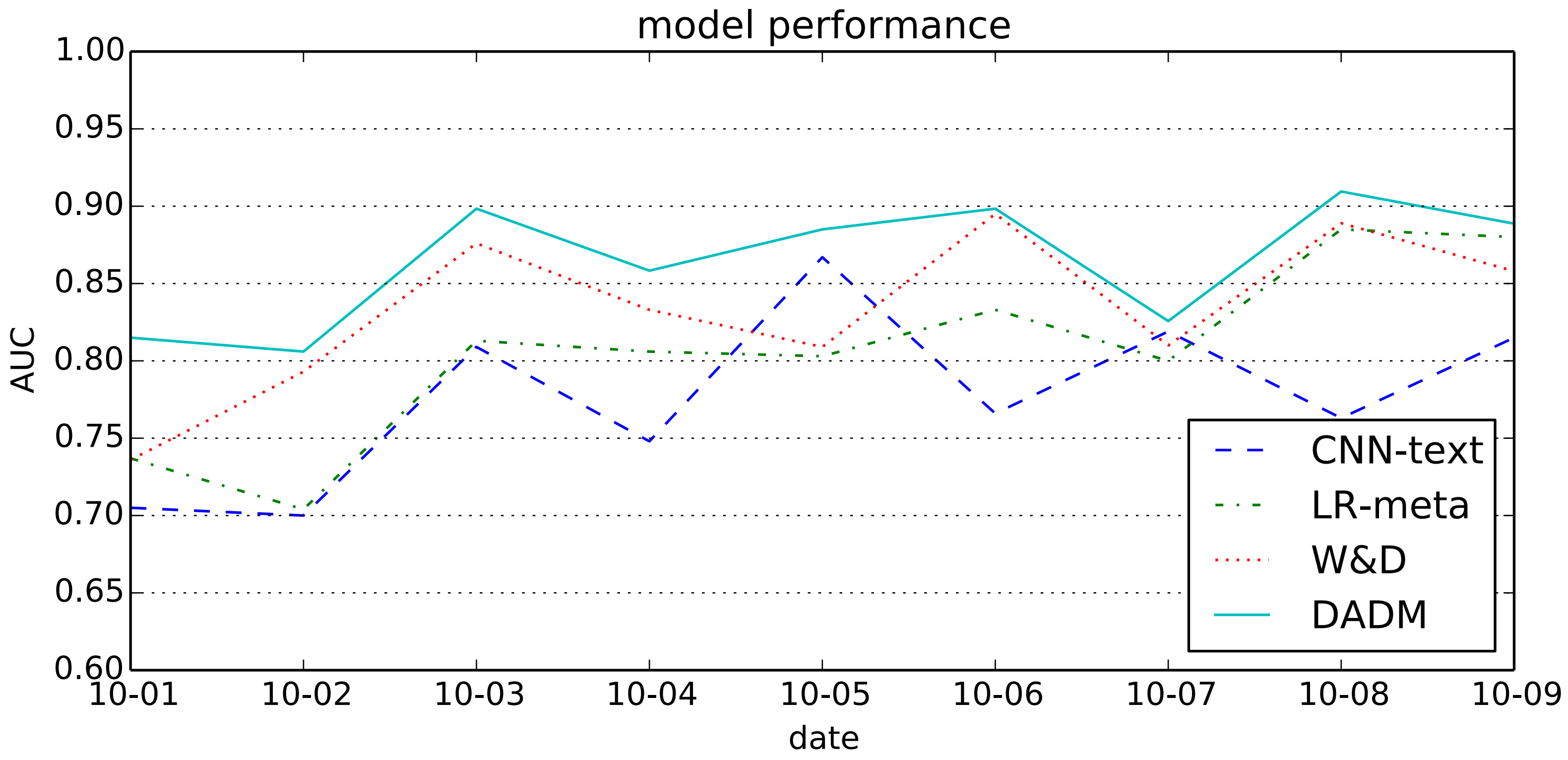
Model	Compared Settings
CNN-text [Yoon Kim, 2014]	convolutional neural network for text representation and classification
LR-meta	logistic regression model
W&D [Heng-Tze Cheng, 2016]	the widely-used wide & deep model which leverages both of the two aspects of information
DADM	our proposed dynamic attention deep model as discussed before

Experiments

Model	AUC	F1	Precision	Recall
CNN-text	0.777±0.052	0.186±0.079	0.170±0.077	0.253±0.143
LR-meta	0.807±0.055	0.255±0.107	0.221±0.118	0.376±0.148
W&D	0.833±0.049	0.284±0.091	0.220±0.094	0.484±0.187
DADM	**0.853±0.036**	**0.317±0.079**	0.258±0.059	0.451±0.202

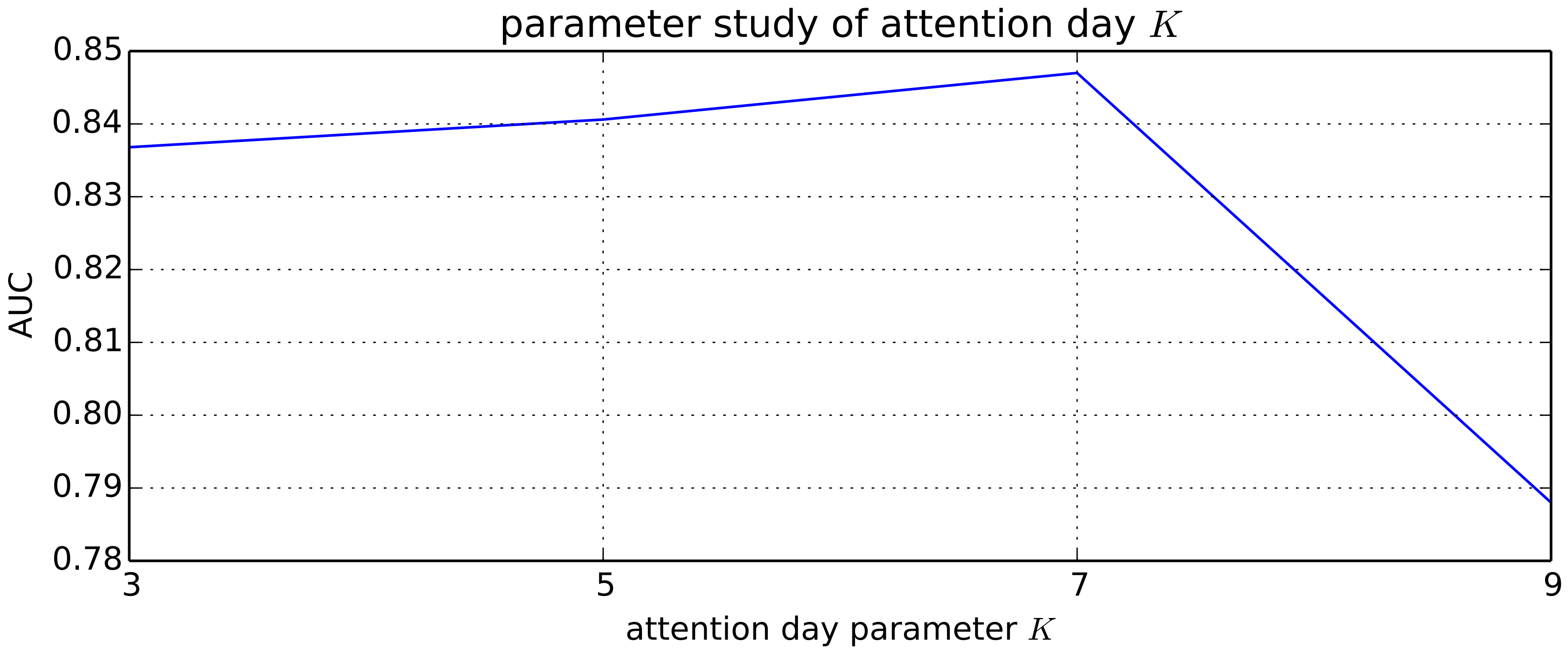
- the overall performance of recommending quality articles over a time period of 9 days.

Experiments



- making more accurate decisions with lower variance.
- robustness and stability

Ablation Study



- the empirical optimal attention day number is 7
- obvious week patterns

Summary

- we have proposed a dynamic attention deep model to deal with the problems of non-explicit selection criteria and non-stationary data in the editors’ article selection stage of content recommendation pipeline.
- A 9-day online A/B testing has shown that our proposed dynamic attention deep model performs the best in terms of both prediction AUC and F1 score as well as the low variance in handling the dynamic data and editors’ behavior.

Further Thinking

- How our recommendations influence the editors’ further actions since their observed data is ‘biased’ due to our provided article ranking?
- Exploitation and exploration problem

- Other application such as sentiment analysis etc.

Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration

Q&A

xjwang@apex.sjtu.edu.cn