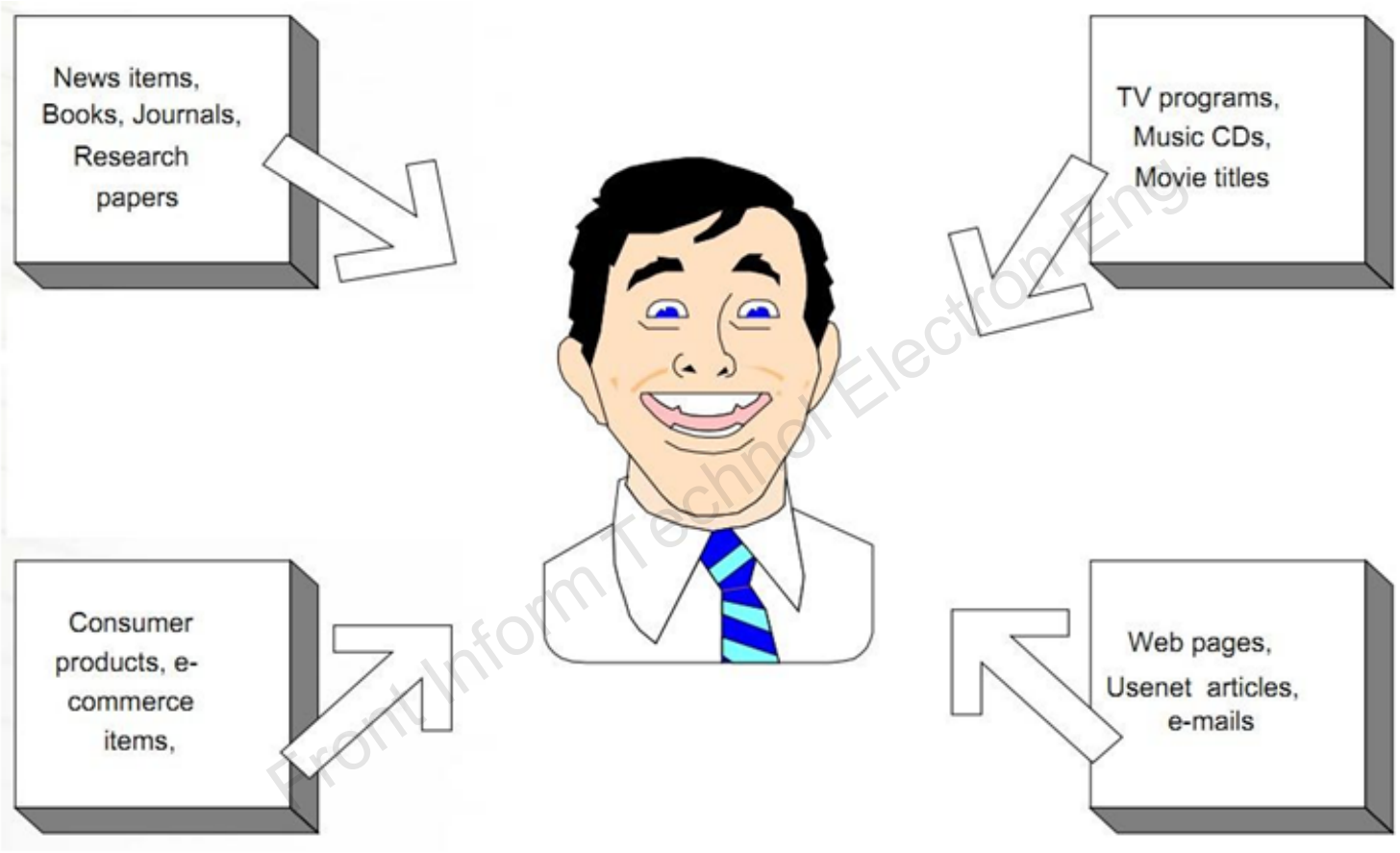
Cheng-wei WANG, Teng-fei ZHOU, Chen CHEN, Tian-lei HU, Gang CHEN, 2020. HAM: a deep collaborative ranking method incorporating textual information. *Frontiers of Information Technology & Electronic Engineering*, 21(8):1206-1216. https://doi.org/10.1631/FITEE.1900382

HAM: a deep collaborative ranking method incorporating textual information

Key words: Deep learning; Recommendation system; Highway network; Block coordinate descent

Corresponding author: Gang CHEN
E-mail: cg@zju.edu.cn
ORCID: https://orcid.org/0000-0002-7483-0045

Background for a recommendation system



A recommendation system, or a recommender system (sometimes replacing “system” with a synonym such as platform or engine), is a subclass of information filtering system that seeks to predict the “rating” or “preference” a user would give to an item.

Background for a recommendation system (Cont'd)

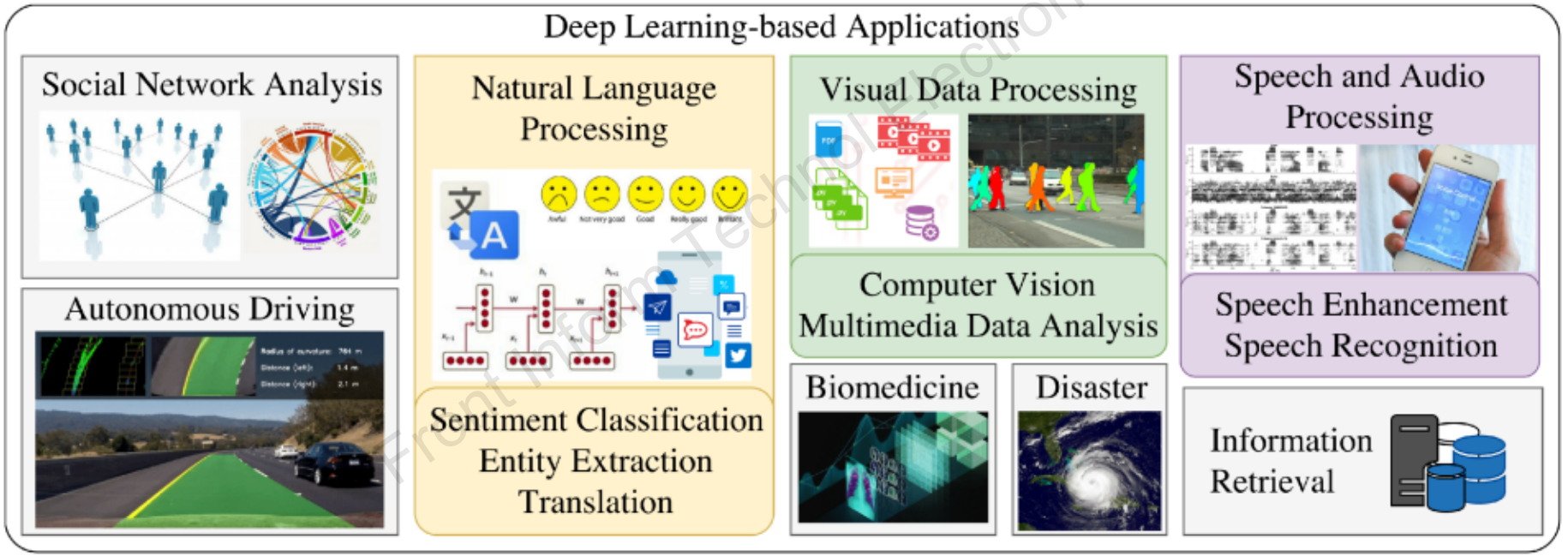
- Amazon.com personalized book recommendation

Background for deep learning

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised, or unsupervised.

- In 1980, Fukushima Kunihiko proposed a new perceptron.
- In 1989, LeCun and other researchers applied the standard. backpropagation algorithm proposed in 1974 to deep neural networks for handwritten postal code recognition.
- In 2012, Hinton significantly reduced the error rate of ImageNet image classification problems.
- In 2016, DeepMind's AlphaGo used deep reinforcement learning to train the Go AI to defeat the human world champion.

Background for deep learning based applications



Contend based recommendation

This method further fully mines the content information of users and items in addition to using feedback data

– content information: **text content,** picture content, audio content, video content

Motivations and challenges

- Noise in text content

- ☐ Topic 1: Gangster
- ☐ Topic 2: love

- Multiple topics

Motivations and challenges (Cont'd)

Motivation I: mining textual content more effectively

Motivation II: designing a very deep neural structure

- The number of neural layers usually determines the neural model's performance. A recommendation system based on a deeper neural network is likely to outperform the shallow ones.
- Due to overparameterization, a deeper neural network generalizes better than shallow ones.

Challenge I

- Textual noise makes a neural network be prone to overfitting.
- It is difficult to identify topics in textual information by neural networks.

Challenge II

- Gradient vanishing/exploding issue.
- Data sparsity of the recommendation problem further exacerbates the instability issue of very deep networks.
- Most state-of-the-art recommenders are based on shallow neural networks.

HighwAy recoMmender (HAM)

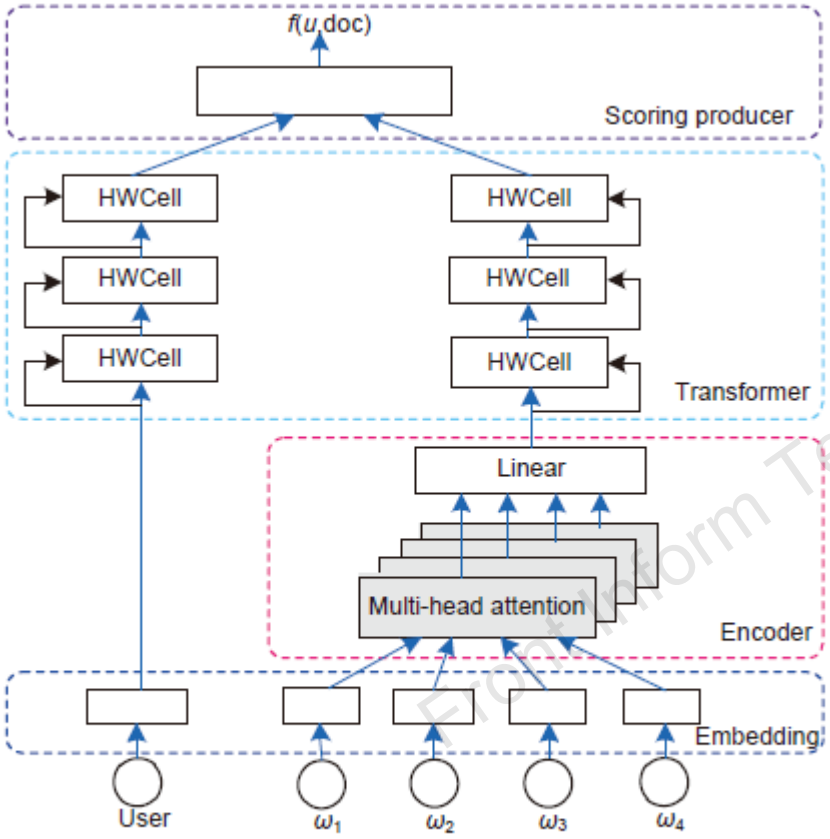


Fig. 1 Architecture of the HighwAy recoMmender (HAM)

Deep transformer

- Deep transformer enhances expressive power of the encoder
- The transformer is constituted of a series of HWCell
- The HWCell $H(x|G,K)$ can be computed by

$$\begin{cases} g = \sigma(Gx), \\ H(x|G,K) = g \odot x + (1-g) \odot \tanh(Kx), \end{cases}$$

  where *g* acts as an information gate.
- The HWCell uses gated skip-connection to stabilize the gradient flow of the neural network
- This kind of neural structure can alleviate gradient vanishing/exploding issue of deep neural networks

Scoring module

- Suppose the latent vectors of users and items are mapped into $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ respectively by the deep transformer
- The scoring module describes the match of users and items via:

$$f(i, \mathrm{doc}_j) = u_i^{\mathrm{T}} v_j.$$

Training the model

Train HAM via multitask learning

Ranking task + Word embedding task

$$\min_{\theta_f, C} \mathcal{L}_{\text{rank}}(\theta_f) + \lambda \mathcal{L}_{\text{embd}}(C, E).$$

Coordinate descent method

The optimization problem can be transformed into the following problem by penalizing constrains

$$\min_{\theta_{\text{opt}}} -\frac{1}{N} \sum_{(i,j,j')} \ln \sigma \left((P_L[i])^{\mathrm{T}} (Q_L[j] - Q_L[j']) \right) + \lambda L_{\text{embd}}(C, E)$$

s.t.

$$\begin{cases} P_l[i] = H(P_{l-1}[i] | \boldsymbol{G}_l, \boldsymbol{K}_l), & i \le m, 1 \le l \le L, \\ Q_l[j] = H(Q_{l-1}[j] | \tilde{\boldsymbol{G}}_l, \tilde{\boldsymbol{K}}_l), & j \le n, 1 \le l \le L, \\ P_0 = P, \\ Q_0[j] = \text{Enc}(\boldsymbol{E}[\text{doc}] | \theta_e), & j \le n \end{cases}$$

$$\min_{\theta_{\text{opt}}} \underbrace{-\frac{1}{N} \sum_{(i,j,j')} \ln \sigma \left((P_L[i])^{\mathrm{T}} (Q_L[j] - Q_L[j']) \right)}_{\mathcal{L}_O(\boldsymbol{P}_L, \boldsymbol{Q}_L)} + \lambda \mathcal{L}_{\text{embd}}(C, E) + \sum_{l=1}^{L} \Bigg\{ \underbrace{\frac{\gamma}{m} \| P_l - H_l(P_{l-1}) \|^2}_{\mathcal{L}_{H(l)}(\boldsymbol{P}_l, \boldsymbol{P}_{l-1}, \boldsymbol{G}_l, \boldsymbol{K}_l)} + \underbrace{\frac{\gamma}{n} \| Q_l - \tilde{H}_l(Q_{l-1}) \|^2}_{\mathcal{L}_{H(l)}(\boldsymbol{Q}_l, \boldsymbol{Q}_{l-1}, \tilde{\boldsymbol{G}}_l, \tilde{\boldsymbol{K}}_l)} \Bigg\} + \underbrace{\frac{\gamma}{m} \| P_0 - P \|^2}_{\mathcal{L}_S(\boldsymbol{P}_0, \boldsymbol{P})} + \underbrace{\frac{\gamma}{n} \| Q_0 - Q - E_{\theta_e}(\{\text{doc}_j\}) \|^2}_{\mathcal{L}_E(\boldsymbol{Q}_0, \boldsymbol{Q}, \boldsymbol{E}_{\text{doc}_j}, \theta_e)}$$

Experimental results

Table 3 Performance of the different methods on the collaborative ranking task

Method	AUC			MAP		
	CiteULike	M1M	M10M	CiteULike	M1M	M10M
CMLST	0.9313±0.003	0.9104±0.013	0.9370±0.013	0.1217±0.002	0.1881±0.004	0.1753±0.006
CRAE	0.9206±0.005	0.9157±0.023	**0.9682±0.024**	0.0580±0.007	0.1856±0.007	0.1585±0.005
CDL	0.9120±0.004	0.9188±0.023	0.9365±0.036	0.1124±0.003	0.1432±0.005	0.1547±0.004
CDAE	0.9348±0.003	0.9107±0.014	0.9366±0.017	0.1211±0.006	0.1630±0.002	0.1712±0.004
CTR	0.9043±0.007	0.9109±0.012	0.9211±0.021	0.0603±0.002	0.1008±0.009	0.1148±0.003
HAM	**0.9406±0.002**	**0.9239±0.009**	0.9671±0.010	**0.1481±0.008**	**0.2332±0.004**	**0.2249±0.007**

The best results are in bold

Table 4 Performance of the different methods on the top-k ranking task

Method	Recall@5			Precision@5		
	CiteULike	M1M	M10M	CiteULike	M1M	M10M
CMLST	0.0937±0.002	0.0722±0.003	0.0667±0.013	0.1100±0.006	0.2291±0.004	0.1822±0.006
CRAE	0.0825±0.003	0.0731±0.003	0.0729±0.024	0.1308±0.004	0.3132±0.002	0.2260±0.007
CDL	0.0895±0.007	0.0758±0.006	0.0840±0.036	0.1404±0.006	0.3782±0.010	0.2773±0.008
CDAE	0.0925±0.004	0.0858±0.002	0.0767±0.017	0.1511±0.002	0.3680±0.009	0.2445±0.003
CTR	0.0363±0.001	0.0434±0.002	0.0520±0.013	0.0430±0.001	0.2059±0.003	0.1356±0.005
HAM	**0.1151±0.002**	**0.0958±0.004**	**0.1038±0.010**	**0.1647±0.006**	**0.4142±0.006**	**0.3532±0.009**

The best results are in bold

Experimental results (Cont'd)

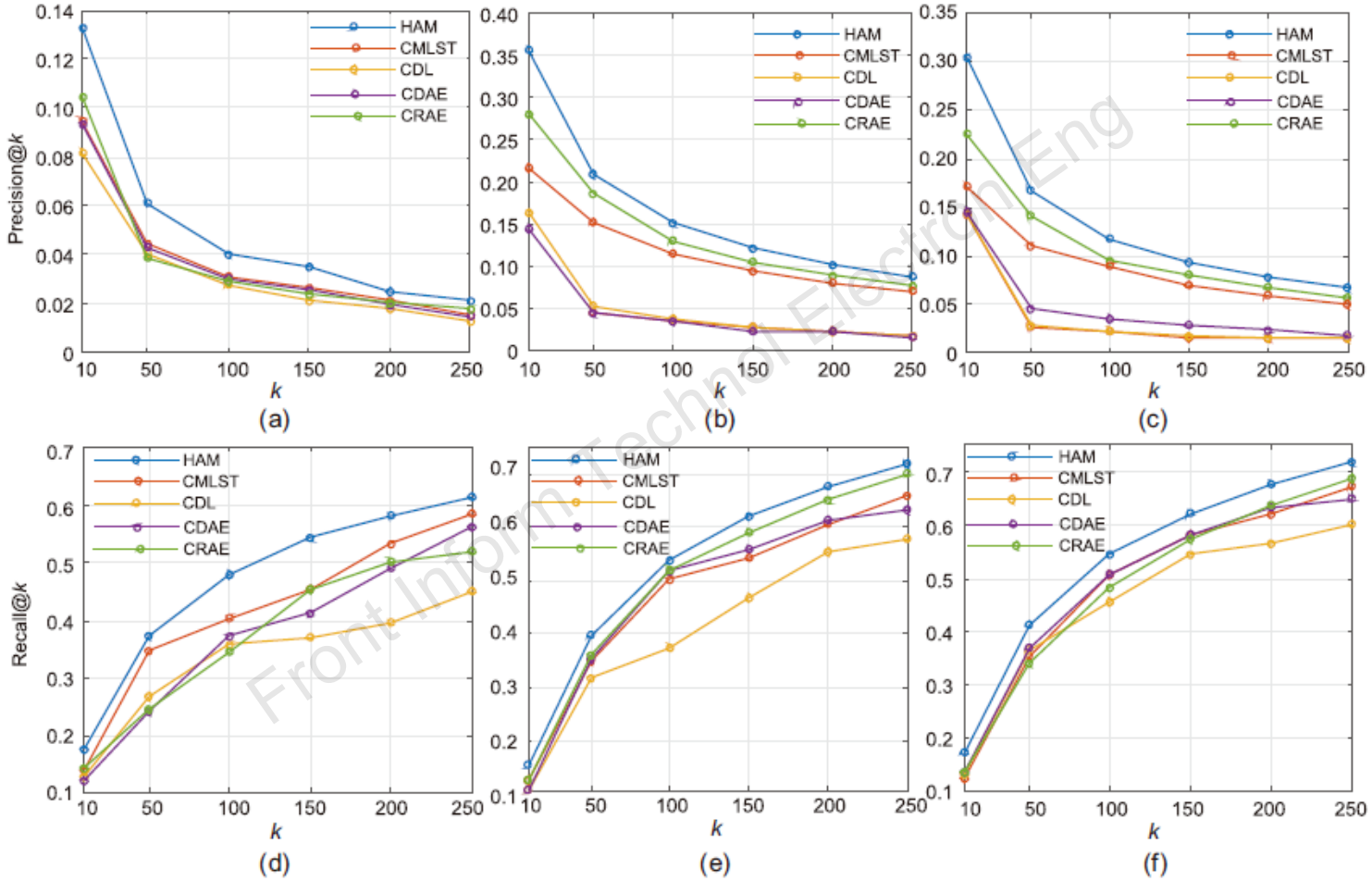


Fig. 2 Precision@k of the different methods under different choices of k with respect to benchmark datasets CiteULike (a), M1M (b), and M10M (c), and Recall@k of the different methods under different choices of k with respect to benchmark datasets CiteULike (d), M1M (e), and M10M (f)

Experimental results (Cont'd)

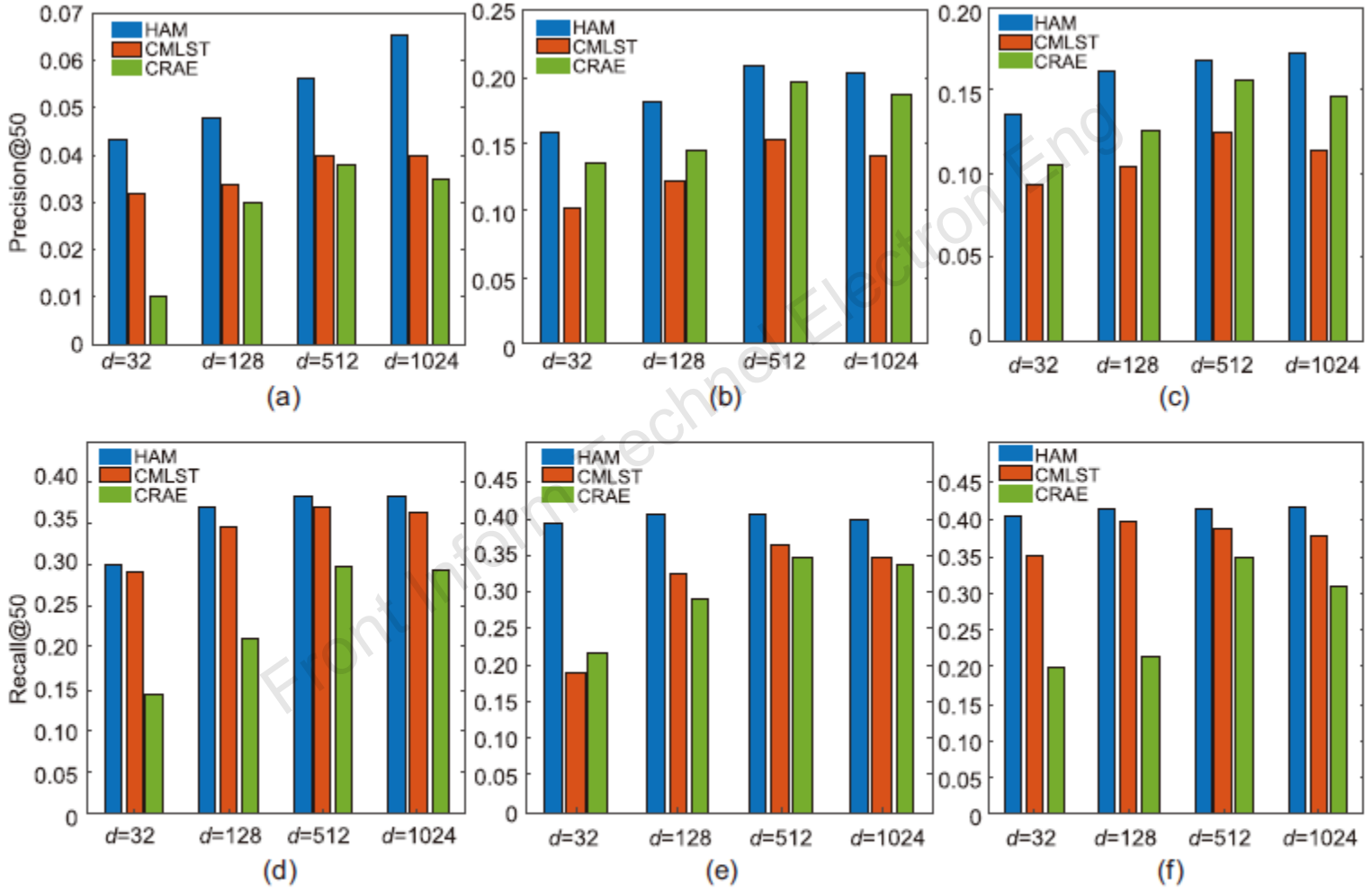


Fig. 3 Precision@50 of the different methods under different choices of dimension with respect to benchmark datasets CiteULike (a), M1M (b), and M10M (c), and Recall@50 of the different methods under different choices of dimension with respect to benchmark datasets CiteULike (d), M1M (e), and M10M (f)

Summary

- We designed a new neural recommendation framework based on the highway network to stabilize the gradient flow of the deep recommender.
- A multi-head attention encoder has been designed to exploit textual information for enhancing the recommendation performance.
- We proposed a novel BCD method to train the deeper recommender effectively.
- Empirical studies showed that the proposed method outperforms state-of-the-art methods significantly in terms of accuracy.

Cheng-wei WANG received the Ph.D. degree in Computer Science from Zhejiang University, Hangzhou, China. He is a data scientist at the Hangzhou Laohe Mountain Technology Corp. His research interests include machine learning, data mining, and recommendation system.

Gang CHEN received the Ph.D. degree in Computer Science from Zhejiang University, Hangzhou, China. He is a professor at the College of Computer Science and Technology, Zhejiang University. He is also Director of the Key Laboratory of Big Data Intelligent Computing of Zhejiang Province. His research interests include database management technology, intelligent computing based big data, and massive Internet system. He is a member of the ACM and a standing member of the China Computer Federation Database Professional Committee.