Table 3: Top six keywords of an example user's dynamic profile with the time being five quarters from April 2014 to May 2015. The keywords from the DGT ground truth, generated by the best baseline DPDR and our DUWE are presented for the user in the rows, respectively.

	Apr. 2014 to Jun. 2014	Jul. 2014 to Sep. 2014	Oct. 2014 to Dec. 2014	Jan. 2015 to Mar. 2015	Apr. 2015 to May 2015
Ground Truth	badminton leaf basketball flower bicycling root	muscle apple heart kiwi lungs pomelo	freezer fly toaster cockroach cabinet ant	injury clothes joint slacks immune slippers	school macbook teacher ipad assignment iphone
DPDR	badminton sky basketball herb coach grass	heart apple ankle pomelo finger peach	freezer water muffin fly toaster cockroach	injury clothes dose slacks food slippers	class ipad garden update teacher system
DUWE	badminton flower basketball leaf bicycling fruit	heart apple muscle kiwi breath pomelo	freezer ant dishwaster fly toaster cockroach	injury clothes ankle trousers doctor slacks	teacher laptop student apple school ipad

where $\mathbf{A} = \left(1/\boldsymbol{\sigma}_{t-1}^2 + 1/\boldsymbol{\alpha}_{t-1}^2\right)^{-1}\left(\boldsymbol{\mu}_{t-1}/\boldsymbol{\sigma}_{t-1}^2 + \mathbf{U}_t/\boldsymbol{\alpha}_{t-1}^2\right)$. The integral in (32) leads to a constant factor (independent of $\mathbf{U}_t$) because it is invariant under a constant shift of the integration variable, which will be cancelled out as well. Thus, (32) becomes:

$$\begin{aligned}
&\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\boldsymbol{\alpha}_{t-1}^2} + \frac{1}{\overline{\boldsymbol{\alpha}}_0^2}\right)\mathbf{U}_t^2\right] \times \exp\left[\frac{1}{2}\left(\frac{1}{\boldsymbol{\sigma}_{t-1}^2} + \frac{1}{\boldsymbol{\alpha}_{t-1}^2}\mathbf{A}^2\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\overline{\boldsymbol{\alpha}}_0^2} + \frac{1}{\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2}\right)\mathbf{U}_t^2 + \frac{\boldsymbol{\mu}_{t-1}}{\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2}\mathbf{U}_t\right] \\
&\propto \mathcal{N}(\mathbf{U}_t; \widetilde{\boldsymbol{\gamma}}_t, \widetilde{\boldsymbol{\psi}}_t^2\mathbf{I}),
\end{aligned} \tag{33}$$

where we let $\widetilde{\boldsymbol{\gamma}}_t$ and $\widetilde{\boldsymbol{\psi}}_t^2$ abbreviate for the means and variances for all users' embeddings $\mathbf{U}_t$, respectively, and let:

$$\frac{1}{\widetilde{\boldsymbol{\psi}}_t^2} = \frac{1}{\overline{\boldsymbol{\alpha}}_0^2} + \frac{1}{\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2}, \quad \text{and} \quad \frac{\widetilde{\boldsymbol{\gamma}}_t}{\widetilde{\boldsymbol{\psi}}_t^2} = \frac{\boldsymbol{\mu}_{t-1}}{\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2}, \tag{34}$$

which results in:

$$\widetilde{\boldsymbol{\gamma}}_t = \widetilde{\boldsymbol{\psi}}_t^2\left(\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_{t-1}, \tag{35}$$

$$\widetilde{\boldsymbol{\psi}}_t^2 = \left[\left(\boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\alpha}_{t-1}^2\mathbf{I}\right)^{-1} + (1/\overline{\boldsymbol{\alpha}}_0^2)\mathbf{I}\right]^{-1}. \tag{36}$$

REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *WSDM*, pages 5–14, 2009.
[2] K. Balog and M. de Rijke. Determining expert profiles (with and application to expert finding). In *IJCAI*, pages 2657–2662, 2007.
[3] K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch. Broad expertise retrieval in sparse data environments. In *SIGIR*, pages 551–558, 2007.
[4] R. Bamler and S. Mandt. Dynamic word embeddings. In *ICML*, pages 380–389, 2017.
[5] C. Bishop. Pattern recognition and machine learning (information science and statistics). *Springer, New York*, 2007.
[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.
[7] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. 2016.
[8] P. Bromiley. Products and convolutions of gaussian probability density functions. *Tina-Vision Memo*, 3(4), 2014.
[9] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR*, pages 659–666, 2008.
[10] R. Collobert, J. Weston, L. Bottou, and et al. Natural language processing (almost) from scratch. *JMLR*, 12:2493–2537, 2011.
[11] N. Craswell, A. P. de Vries, and I. Soboroff. Overview of the TREC 2005 enterprise track. In *TREC'05*, pages 1–7, 2005.
[12] W. B. Croft, D. Metzler, and T. Strohman. *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2015.
[13] V. Dang and W. B. Croft. Diversity by proportionality: An election-based approach to search result diversification. In *SIGIR*, pages 65–74, 2012.
[14] Y. Fang and A. Godavarthy. Modeling the dynamics of personal expertise. In *SIGIR*, pages 1107–1110, 2014.
[15] A. Graves and N. Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In *ICML*, pages 1764–1772, 2014.
[16] W. L. Hamilton, J. Leskovec, and D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *ACL*, pages 1489–1501, 2016.
[17] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic tracking model for analyzing consumer purchase behavior. In *IJCAI*, volume 9, pages 1427–1432, 2009.
[18] Y. Kim, Y.-I. Chiu, K. Hanaki, D. Hegde, and S. Petrov. Temporal analysis of language through neural language models. *arXiv preprint arXiv:1405.3515*, 2014.
[19] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
[20] V. Kulkarni, R. Al-Rfou, B. Perozzi, and S. Skiena. Statistically significant detection of linguistic change. In *WWW*, pages 625–635, 2015.
[21] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW'10*, pages 591–600, 2010.
[22] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *ICML*, pages 1188–1196, 2014.
[23] S. Liang and M. de Rijke. Formal language models for finding groups of experts. *Information Processing & Management*, 52(4):529–549, 2016.
[24] S. Liang, E. Yilmaz, and E. Kanoulas. Dynamic clustering of streaming short documents. In *KDD*, pages 995–1004, 2016.
[25] S. Liang, Z. Ren, Y. Zhao, J. Ma, E. Yilmaz, and M. D. Rijke. Inferring dynamic user interests in streams of short texts for user clustering. *ACM Trans. Inf. Syst.*, 36(1):10:1–10:37, 2017.
[26] S. Liang, E. Yilmaz, H. Shen, M. D. Rijke, and W. B. Croft. Search result diversification in short text streams. *ACM Trans. Inf. Syst.*, 36(1):8:1–8:35, 2017.
[27] S. Liang, I. Markov, Z. Ren, and M. de Rijke. Manifold learning for rank aggregation. In *WWW*, pages 1735–1744, 2018.
[28] S. Liang, E. Yilmaz, and E. Kanoulas. Collaboratively tracking interests for user clustering in streams of short texts. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–16, 2018.
[29] R. Mihalcea and V. Nastase. An introduction to the kalman filter. In *SIGGRAPH*, pages 27599–23175, 2001.
[30] R. Mihalcea and V. Nastase. Word epoch disambiguation: Finding how words change over time. In *ACL*, pages 259–263, 2012.
[31] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
[32] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119, 2013.
[33] S. Okura, Y. Tagami, S. Ono, and A. Tajima. Embedding-based news recommendation for millions of users. In *KDD*, pages 1933–1942, 2017.
[34] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *EMNLP*, pages 1532–1543, 2014.
[35] R. Ranganath, S. Gerrish, and D. Blei. Black box variational inference. In *Artificial Intelligence and Statistics*, pages 814–822, 2014.
[36] J. Rybak, K. Balog, and K. Nørvåg. Temporal expertise profiling. In *ECIR*, pages 540–546, 2014.
[37] T. Salimans and D. P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *NIPS*, pages 901–909, 2016.
[38] X. Wang and A. McCallum. Topics over time: a non-markov continuous-time model of topical trends. In *KDD*, pages 424–433, 2006.
[39] X. Wei, J. Sun, and X. Wang. Dynamic mixture models for multiple time-series. In *IJCAI*, pages 2909–2914, 2007.
[40] J. Wieting, M. Bansal, K. Gimpel, and K. Livescu. Towards universal paraphrastic sentence embeddings. 2015.
[41] C. Zhai and J. D. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *SIGIR*, 2001.
[42] Y. Zhao, S. Liang, Z. Ren, J. Ma, E. Yilmaz, and M. de Rijke. Explainable user clustering in short text streams. In *SIGIR*, pages 155–164, 2016.