Correction

Correction: Huang et al. An Attention-Based Recommender System to Predict Contextual Intent Based on Choice Histories across and within Sessions. *Appl. Sci.* 2018, 8, 2426

Ruo Huang ¹, Shelby McIntyre ², Meina Song ^{1,*}, Haihong E ¹ and Zhonghong Ou ¹

¹ School of Computer Science, Beijing University of Posts & Telecommunications, Beijing 100876, China; charleshuangruo@bupt.edu.cn (R.H.); ehaihong@bupt.edu.cn (H.E.); zhonghong.ou@bupt.edu.cn (Z.O.)

² Leavey School of Business, Santa Clara University, Santa Clara, CA 95053, USA; smcintyre@scu.edu

* Correspondence: mnsong@bupt.edu.cn; Tel.: +86-139-1102-3433

Citation: Huang, R.; McIntyre, S.; Song, M.; E, H.; Ou, Z. Correction: Huang et al. An Attention-Based Recommender System to Predict Contextual Intent Based on Choice Histories across and within Sessions. *Appl. Sci.* 2018, 8, 2426. *Appl. Sci.* 2021, 11, 6633. <https://doi.org/10.3390/app11146633>

Received: 10 June 2021

Accepted: 25 June 2021

Published: 19 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

We, the authors, wish to make the following corrections to our paper [1]. We found that in our experiment the attention mechanism applied in our proposed models ASLM-LHS and ASLM-AP mistakenly utilized the user's future events including the ground-truth event and the ones with which the user would interact after as the prior knowledge to predict the user's next event. This issue resulted from the overlapped input data $\{e_{t,i}^u \in \mathbb{R}^m | i = 1, 2, \dots, |S_t^u| - 1\}$ to the encoder and the ground-truth events $\{e_{t,j}^u \in \mathbb{R}^m | j = 2, 3, \dots, |S_t^u|\}$ as the labels which were fed into the models based on Section 3.1. In the RNN's architecture without the attention mechanism, the future events will not be used, while, in our proposed model with the attention-based encoder-decoder architecture, the encoder will first process all the events in the input data to extract user's sequential patterns and then generate the final output vector $[\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2 \dots \bar{\mathbf{h}}_i]$, which contains the context information from all events. Afterwards, the attention mechanism will use the entire information provided by $[\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2 \dots \bar{\mathbf{h}}_i]$ to calculate the relevance scores, and therefore the information from future events will be included during the calculation. This issue led to incorrect experimental results. Thus, we resolved this issue by only providing the events with which the user interacted before the ground-truth event to the models as the input data (e.g., $\{e_{t,1}^u, e_{t,2}^u\}$) and using their subsequent ground-truth event as the label to predict (e.g., $e_{t,3}^u$). We conducted the experiment again and would like to correct the values in Tables 5–9, Figures 8 and 9, and their corresponding sentences in the main text. This correction does not affect the scientific conclusions.

1. Changes in Tables

The values in Tables 5–9 in the original paper would be changed to those in Tables 1–5 in this correction, respectively.

Table 1. Group A: Recall and MRR scores for the ASLM models and the baselines for the test sets in Reddit and Last.fm datasets.

	R@5	R@10	R@20	MRR@5	MRR@10	MRR@20
Most Popular	0.1241	0.1799	0.2499	0.0817	0.0890	0.0940
II-RNN-AP	0.2943	0.3717	0.4527	0.1920	0.2024	0.2080
II-RNN-LHS	0.3112	0.3904	0.4709	0.2025	0.2131	0.2186
ASLM-AP	0.3162 (+1.62%)	0.3879 (−0.63%)	0.4643 (−1.40%)	0.2277 (+12.44%)	0.2373 (+11.36%)	0.2426 (+10.95%)
ASLM-LHS	0.3396 (+9.12%)	0.4100 (+5.04%)	0.4843 (+2.85%)	0.2473 (+22.10%)	0.2567 (+20.48%)	0.2619 (+19.79%)
(a) Reddit Dataset (Test).						

Table 1. *Cont.*

	R@5	R@10	R@20	MRR@5	MRR@10	MRR@20
Most Popular	0.0569	0.0689	0.0866	0.0470	0.0487	0.0499
II-RNN-AP	<u>0.1009</u>	<u>0.1341</u>	<u>0.1784</u>	<u>0.0682</u>	<u>0.0725</u>	<u>0.0755</u>
II-RNN-LHS	0.0855	0.1142	0.1555	0.0588	0.0625	0.0654
ASLM-AP	0.1053	0.1387	0.1837	0.0732	0.0776	0.0806
	(+4.36%)	(+3.46%)	(+3.01%)	(+7.33%)	(+6.99%)	(+6.75%)
ASLM-LHS	0.1042	0.1415	0.1914	0.0693	0.0742	0.0776
	(+3.34%)	(+5.52%)	(+7.31%)	(+1.66%)	(+2.34%)	(+2.78%)

(b) Last.fm Dataset (Test).**Table 2.** Group B: Recall and MRR scores for the ASLM models and the baselines for the test sets in 1/10 Reddit and 1/80 Last.fm datasets.

	R@5	R@10	R@20	MRR@5	MRR@10	MRR@20
Most Popular	0.1451	0.2155	0.2882	0.0966	0.1060	0.1111
II-RNN-AP	0.1484	0.2161	0.2928	0.0969	0.1058	0.1112
II-RNN-LHS	<u>0.1614</u>	<u>0.2292</u>	<u>0.3062</u>	<u>0.1061</u>	<u>0.1151</u>	<u>0.1204</u>
ASLM-AP	0.1904	0.2507	0.3245	0.1256	0.1336	0.1387
	(+17.96%)	(+9.38%)	(+5.98%)	(+18.37%)	(+16.07%)	(+15.17%)
ASLM-LHS	0.1886	0.2488	0.3251	0.1234	0.1314	0.1366
	(+16.83%)	(+8.55%)	(+6.19%)	(+16.27%)	(+14.13%)	(+13.48%)

(a) 1/10 Reddit Dataset (Test).

	R@5	R@10	R@20	MRR@5	MRR@10	MRR@20
Most Popular	<u>0.0371</u>	<u>0.0493</u>	0.0596	<u>0.0290</u>	<u>0.0308</u>	<u>0.0314</u>
II-RNN-AP	0.0349	0.0448	0.0604	0.0282	0.0294	0.0305
II-RNN-LHS	0.0361	0.0459	<u>0.0608</u>	0.0285	0.0298	0.0308
ASLM-AP	0.0388	0.0496	0.0656	0.0308	0.0323	0.0334
	(+4.38%)	(+0.57%)	(+7.95%)	(+6.18%)	(+4.93%)	(+6.13%)
ASLM-LHS	0.0388	0.0503	0.0674	0.0309	0.0324	0.0336
	(+4.38%)	(+1.92%)	(+10.97%)	(+6.52%)	(+5.37%)	(+6.98%)

(b) 1/80 Last.fm Dataset (Test).**Table 3.** Group B: Recall and MRR scores for the ASLM models and the baselines in the Tmall dataset.

	R@10	R@20	MRR@20
Most Popular	0.0234	0.0420	0.0123
SWIWO-I	<u>0.3177</u>	<u>0.3810</u>	<u>0.1903</u>
SWIWO	0.3082	0.3703	0.1885
ASLM-AP	0.5412	0.5661	0.4037
	(+70.34%)	(+48.57%)	(+112.12%)
ASLM-LHS	0.5428	0.5679	0.4082
	(+70.86%)	(+49.06%)	(+114.50%)

Table 4. Group A: Recall and MRR scores for the ASLM models and the baselines for the validation sets in Reddit and Last.fm datasets.

[illegible]

Table 5. Group B: Recall and MRR scores for the ASLM models and the baselines for the validation sets in 1/10 Reddit and 1/80 Last.fm datasets.

[illegible]

2. Changes in the Statements Corresponding to the Tables

The statements of Section 5.1 corresponding to these tables would be changed to:

(The statements of Tables 1–3 remain the same as those in the original paper.) For all datasets, ASLM-LHS outperforms all baselines with moderate margins. Table 4 shows the validation set performances of three different methods under the metric of Recall@K and MRR@K in Reddit and Last.fm datasets. Please note that, for the Reddit dataset, the validation set performance of ASLM-LHS is better than that of ASLM-AP, and, for Reddit and Last.fm datasets, the validation set performances of ASLM-AP and ASLM-LHS are all better than the test set performances with moderate margins, respectively. Table 5 shows the validation set performances of three different methods in 1/10 Reddit and 1/80 Last.fm datasets. The validation set performances of the two ASLM models in the 1/10 Reddit dataset are all better than the test set performances, while, in the 1/80 Last.fm dataset, the majority of the validation set performances of the two ASLM models are better than the test set performances.

The statements of Section 6.1 corresponding to these tables would be changed to:

For Group A, in Table 1, we observe that ASLM-LHS consistently outperforms all the baselines under all measurements for testing cases in both Reddit and Last.fm datasets with moderate margins; In most cases, ASLM-AP outperforms all the baselines in both datasets, except in Recall@10 and Recall@20 scores in the Reddit test set. Specifically, ASLM-LHS improves 9.12% and 22.10% in Recall@5 and MRR@5 scores compared with the II-RNN-LHS method for test cases in the Reddit dataset, respectively. ASLM-AP improves 4.36% and 7.33% in Recall@5 and MRR@5 compared with the II-RNN-AP method for test cases in the Last.fm dataset, respectively. The reason ASLM-AP's Recall@10 and Recall@20 scores are lower than those of II-RNN-LHS in the Reddit dataset is that the average pooling method only utilizes the average of the embedding of each event $e_{t,i}^u \in S_t^u$ as the representation of the current session S_t^u , and the important context information, e.g., the sequential patterns and the user's intent captured by the attention mechanism, will be lost when this representation is stored in the user's long-term session representations. It results in the fact that the abundant context information cannot be fully utilized by ASLM-AP, and thus its performance declined. However, the performance of ASLM-AP is still better than II-RNN-AP, which also utilizes the average pooling method. The good performance of ASLM-LHS is attributed to the attention mechanism and the bidirectional-LSTM we employed in the attention-based layer and the last hidden state method. One of the main characteristics of the attention mechanism is that it calculates the importance of each given event and captures the user's short-term intent with it. The bidirectional-LSTM can extract the sequential patterns from each given event in both the forward and the backward directions. Both ASLM-LHS and II-RNN-LHS utilize the last hidden state method which retains the context information from the current session and regards it as one part of the user's long-term session representations, compared with the average pooling method. Therefore, the model's subsequent training process benefits from this method by receiving the valuable context information from history sessions. In addition, in Table 1 in the original paper, we observe that, in the Reddit dataset, the number of sessions per user (namely, 62.15) and the average number of events in a session (namely, 3.00) are much smaller than those in the Last.fm dataset (namely, 645.62 and 8.10), which shows that ASLM-LHS can perform better when the user's history information is less adequate. Tables 1 and 4 show the results that the performance of ASLM for validation cases and that for testing cases in Reddit and Last.fm datasets are at the same level, which assure us the validity of our model.

For Group B, similar results can be seen in Tables 2 and 3 that ASLM-LHS and ASLM-AP outperform all the baselines under all measurements in 1/10 Reddit, 1/80 Last.fm, and Tmall datasets. For the Tmall dataset, the evaluation result of ASLM-LHS shown in Table 3 improves 70.86%, 49.06%, and 114.50% in R@10, R@20, and MRR@20 scores compared with the SWIWO-I method, respectively. Please note that, for the Tmall dataset, as shown in Table 1 in the original paper, there is scarcely enough sessions for each user

(namely, 3.99 per user) and number of average session length (2.99, per user) compared with those in Reddit and Last.fm. For 1/10 Reddit and 1/80 Last.fm, since the user history information is reduced dramatically compared with the full Reddit and Last.fm datasets, most of the models' performances significantly decrease accordingly. However, both ASLM models still outperform the strongest baselines, as shown in Table 2. It demonstrates that ASLM can reach better performance even when it is severely short of both long-term and short-term user behavior.

When K changes from 5 to 20, for the Reddit and 1/10 Reddit datasets, there is a downward trend in the relative scores of ASLM-LHS and ASLM-AP. For the Last.fm dataset, there is an upward trend in those of ASLM-LHS. The reason for the upward trend is that there is more abundant user history information in the Last.fm dataset, and, with the increase of K , ASLM-LHS still has the potential to further capture the context information implied by the Last.fm dataset, which results in some space to improve its performance.

In terms of the two versions of ASLM (ASLM-AP using average pooling and ASLM-LHS using the last hidden state), Tables 1–3 show that, for the Reddit dataset, the test performance of ASLM-LHS has an obvious advantage compared to that of ASLM-AP, while they are at the same level for the other four datasets. As mentioned above, the model's subsequent training process benefits from the last hidden state method by receiving the valuable context information from history sessions.

3. Changes in Figures

Figures 8 and 9 in the original paper would be changed to Figures 1 and 2 in this correction, respectively.

Figure 1. Comparison of cold-start effects at the start of a new session for Reddit and Last.fm datasets between ASLM and II-RNN.

Figure 2. Cold-start effect at the start of a new session for the Tmall dataset. Please note that no data are provided for the cold-start problem by the SWIWO method.

4. Changes in the Statements Corresponding to the Figures

The statements of Section 6.2 corresponding to these figures would be changed to:

In addition to the overall recommendations we reported in Sections 5 and 6.1 in the original paper, we evaluate the effectiveness of ASLM-LHS on recommendations for the first n time steps, for $n = 1, \dots, 5, L$, where L is the maximum session length and $L = 19$ in our experiment.

Figure 1 shows the comparison of performances of Recall@5 at the first n recommendations of each session for Reddit and Last.fm datasets between ASLM and II-RNN.

Notice that, for Reddit and Last.fm, where abundant user history behavior is given, our attention-based ASLM-LHS model achieves $R@5 > 0.3$ and $R@5 > 0.099$ when $n = 1$, respectively, and, under most circumstances, each $R@5$ score of the ASLM-LHS model increases throughout the session. ASLM-LHS has an advantage over II-RNN-LHS and II-RNN-AP at the start of a new session in Reddit and Last.fm datasets, with the improvements of 7.00% and 3.12%, respectively. When more events have been investigated in the current session, for the Reddit dataset, the advantage becomes more obvious, while, for the Last.fm dataset, since the performances of ASLM-LHS and II-RNN-AP are identical when $n = 2$, the advantage becomes greater when n increases from 3.

Similar results are shown in Figure 2 for the Tmall dataset. In terms of the Tmall dataset, which is severely short of both long-term and short-term user behavior, ASLM-LHS still provides 0.5063 in $R@5$ with just a slight difference compared with the overall $R@5$ (i.e., 0.5074). Due to the lack of history information from each user in the Tmall dataset, its performance is not stable at the start of a session when n is between 1 and 3, however it gets better when n increases from 3. As a result, the attention-based short-term intent layer can thoroughly reveal the variation in the user's short-term intent between events from the current session S_t^u . To sum up, ASLM significantly alleviates the cold-start problem.

The authors would like to apologize for any inconvenience caused. The change does not affect the scientific conclusions. The manuscript will be updated, and the original version will remain online on the article webpage.

5. Change in the Description of the Last.fm Dataset

Since the link to the Last.fm dataset described in Section 4.1 in the original paper is not valid any more, the statement of Section 4.1 corresponding to the link to the Last.fm dataset would be changed to: Last.fm dataset: <https://blog.csdn.net/hopygreat/article/details/96444827>.

Author Contributions: Conceptualization, R.H., S.M., M.S. and H.E.; Methodology, R.H.; Software, R.H.; Validation, R.H., S.M., M.S., H.E. and Z.O.; Formal Analysis, R.H.; Investigation, R.H.; Resources, R.H., S.M., M.S. and H.E.; Data Curation, R.H.; Writing—Original Draft Preparation, R.H.; Writing—Review & Editing, R.H., S.M., M.S., H.E. and Z.O.; Visualization, R.H.; Supervision, S.M., M.S., H.E. and Z.O.; Project Administration, Z.O.; Funding Acquisition, Z.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (Grant No. 61702046).

Data Availability Statement: Please refer to the following links to the datasets utilized in both the original paper and this correction. Reddit dataset: <https://www.kaggle.com/colemaclean/subreddit-interactions>, Last.fm dataset: <https://blog.csdn.net/hopygreat/article/details/96444827>, and Tmall dataset: <https://tianchi.aliyun.com/datalab/dataSet.html?spm=5176.100073.0.0.40236fc1klb4f8&dataId=42>.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

1. Huang, R.; McIntyre, S.; Song, M.; E, H.; Ou, Z. An Attention-Based Recommender System to Predict Contextual Intent Based on Choice Histories across and within Sessions. *Appl. Sci.* **2018**, *8*, 2426. [CrossRef]