Proceedings of Machine Learning Research vol 245:1–10, 2024 2024 International Conference on Machine Learning and Intelligent Computing

Conversational Recommendation System Based on Utterance Act and Emotion

Jiahao An* AJH_CQUT@163.COM
Department of Computer Science and Engineering, Chongqing University of Technology, Chongqing, 400054, China; Department of Transportation Information, Shaanxi College of Communications Technology, Xi'an, 710018, China

Huibo Dang
Department of Transportation Information, Shaanxi College of Communications Technology, Xi'an, 710018, China

Editors: Nianyin Zeng and Ram Bilas Pachori

Abstract

Conversational recommendation system (CRS) aims to acquire user preferences of conversation and then make recommendations to users. Existing CRS enhance the characterization of items by introducing external information to improve the recommendation effect, but they ignore the most essential utterance semantic attributes in the dialogue, and do not fully con-sider the different emotion or act feedback of users. Based on this, this paper fully explores the utterance semantics, makes the construction of user characteristics fuller by extracting the act and emotion of the utterance, and predicts the act of the conversation process, and then enhances the response through the emotional vocabulary, to improve the interaction experience between the user and the system. A large number of experiments on public datasets show that the model proposed in this paper outperforms the most advanced methods.

Keywords: Conversational recommendation, Utterance act, Semantic fusion, Personal recommendation

1. Introduction

Conversational recommender systems (CRS) (Gao et al., 2021) aim to use conversations to provide recommendations to users. This approach of natural language interaction improves the user experience and is better applicable to cold-start users or users whose preferences change frequently. As shown in Figure 1, CRS also needs to provide explanations when providing appropriate recommendations (e.g., movies, books) or utilize natural language to further acquire user preferences it make accurate recommendations (Zhang et al., 2019). Therefore, most researchers (Chen et al., 2019; Zhou et al., 2020; Lu et al., 2021) have divided CRS into two modules: conversation module, and recommendation module. The conversation module focuses on high-quality dialog responses, incorporates the recommendations generated by the recommendation module into the conversation, and adds explanatory language so that users can learn the reasons for the recommendations or item introductions; the recommendation module needs to make full use of the information involved in the conversation to construct user preferences, to select the appropriate items in the candidate set for recommendation.

In the field of CRS, existing research (Chen et al., 2019; Zhou et al., 2020; Lu et al., 2021) focuses on how to obtain the links more accurately between items and projects, to recommend the correct items, so some researchers have introduced external knowledge graphs (Chen et al., 2019;

© 2024 J. An* & H. Dang.

Zhou et al., 2020), or item reviews (Lu et al., 2021), introduction (Zhang et al., 2023), etc., but it ignores the deeper semantic information expressed in the user's words. To be more detailed, the previous model only roughly characterizes the entity mentioned by the user, and unifies it as the user's preference, which does not take into account the user's feedback on the entity, such as rejection, acceptance, etc., and the negative feedback from the user on the entity will affect the final recommendation item, and the direct inclusion of the entity in the user's preference will make the user's real preference shift, which will affect the recommendation effect (Deng et al., 2022). As shown in Table 1, the user implicitly expresses that he/she does not like the movie about aliens very much, the system does not capture his/her utterance feedback, interprets it as a favorite, and then incorporates it into the user preference to recommend a movie about aliens to the user, which he/she recommends incorrectly.

Table 1: A Sample of a Bad Conversation Recommendation.

User	...
System	Do you like stories with aliens?
User	Well, actually I'm not really interested in stories with aliens.
System	I recommend a movie called E.T. the Extra-Terrestrial. It was directed by Steven Spielberg.
User	I love Spielberg, but I don't like movie about aliens.
System	...

In response to the above problems, we revisited CRS, we construct a novel model for improving conversation recommender systems through utterance act and emotion. Specifically, we first extract semantic vocabulary from utterance and construct utterance subgraphs via emotion-level knowledge graphs and emotion values to capture utterance emotion features. Further, we use historical conversations above to extract utterance act features through deep subspace clustering (Ji et al., 2017), and then fuse the utterance emotional features and act features with entity preferences through the fusion module, to capture user preferences. During the conversation process, according to the past utterance act through the distance-aware attention allows the system to produce a more suitable next act, so that the conversation process proceeds more smoothly. And empathetic replies are generated through emotion-enhanced decoders to improve the user's interaction experience with the system. By considering utterance acts and emotions, we construct an updated hierarchy of user preferences that fully analyzes and exploits the conversational context, thus improving the accuracy of recommendations and the user experience. Our contributions are summarized as follows:

- We constructed utterance subgraphs using emotion knowledge graphs and emotion values to capture sentiment representations of utterance.
- We design a semantic fusion module that learns utterance acts and then augments user preferences by uniting utterance acts and emotions.
- The emotion-enhanced act-directed conversation module generates more fitting responses, thus improving the user interaction experience.

2. Problem Formulation

Formally, u represents a user in a set U of users, i represents an item in a set I of items, and w represents a word in a vocabulary V. A set of conversations C should contain several sentences s in which the system engages in a dialog with a user, then $C = \{s_t\}_{t=1}^{n}$, a sentence contains m words. In round t, the recommendation module selects a set of candidate items I_t from the set of items I according to some strategy. The conversation module needs to generate the next statement s_{t+1} to reply until the user accept the recommendation or ends the dialog voluntarily.

3. Proposed UAE

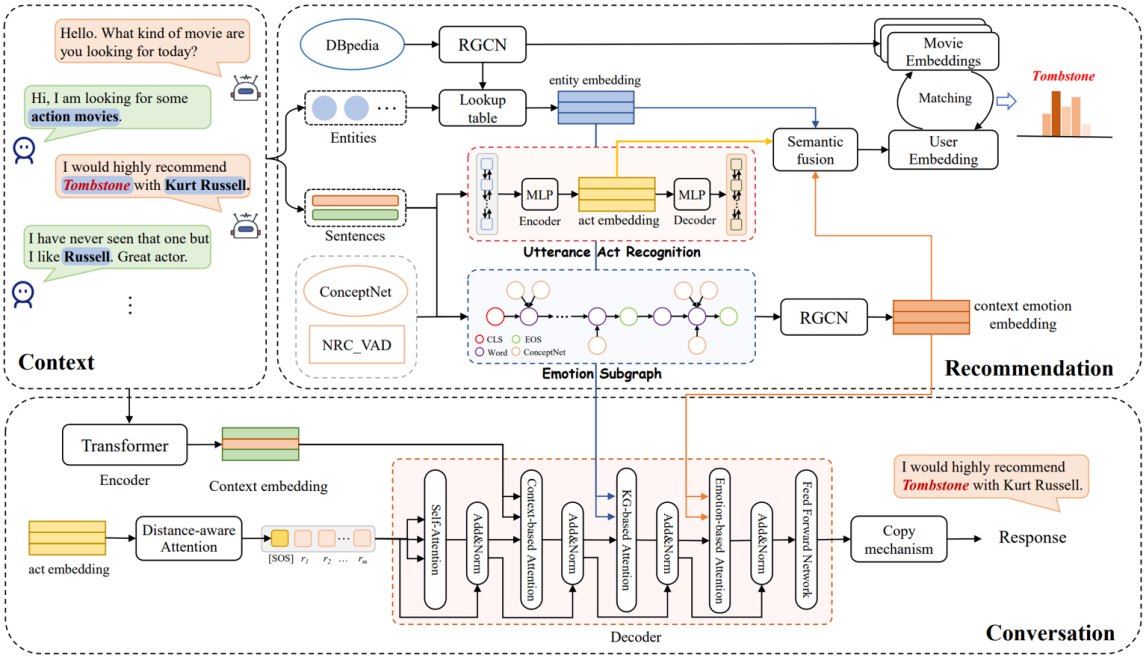

Figure 1: Overall architecture of the proposed UAE.

3.1. Utterance Feature Extraction

3.1.1. UTTERANCE ACT RECOGNITION.

Since REDIAL (Zhong et al., 2019), a commonly used dataset in the CRS domain, is an unlabeled dataset of utterance acts, an unsupervised deep subspace clustering method (Lei et al., 2020) is used to learn the latent features in the conversations to capture the utterance act information. Specifically, first initialize K potential vectors of conversation acts, and then use autoencoder to encode historical conversations C, to extract the potential features in the utterance, the autoencoder can map the input vectors to a deep and nonlinear feature space, which has a better fitting ability to complex high-

dimensional data, as shown in Equation 1.

$$\begin{aligned} \mathbf{H}&=\mathrm{MLP}_{enc}(C) \\ \mathbf{A}&=\mathbf{H}\cdot\mathbf{Z}^{\top} \\ \mathbf{H}^{*}&=\mathrm{softmax}(\mathbf{A})\cdot\mathbf{Z} \\ C^{*}&=\mathrm{MLP}_{dec}(\mathbf{H}^{*}) \end{aligned} \tag{1}$$

where A is the potential act vector. Its loss function is shown in Equation 2 and consists of three components: information exchange loss, self-representation loss, and regularization term loss.

$$\mathcal{L}_{\mathrm{DAR}}=|C^{*}-C|_{F}^{2}+\lambda_{1}\left|\mathbf{H}^{*}-\mathbf{H}\right|_{F}^{2}+\lambda_{2}\left|\mathbf{Z}\mathbf{Z}^{\top}-\mathbf{I}\right| \tag{2}$$

3.1.2. UTTERANCE SENTIMENT RECOGNITION.

Our model constructs sentiment subgraphs based on utterance and semantically related words, and the specific process and rules are as follows:

a. Initialize Nodes: Four types of nodes are set up, which are start node, separation node, vocabulary node, and sentiment word node. The start node and separation node are mainly for distinguishing different utterances, while the vocabulary node represents each word in the utterance, and the emotion word node is the emotion-related words obtained through ConcepetNet (Poria et al., 2019).

b. Sentiment Word Selection: Introduces NRC_VAD (Ren et al., 2023) as the external sentiment value, which is a word list that contains three features V, A, and D, which denote pleasantness, arousal, and dominance, e.g. unhappy: [0.12, 0.50, 0.16].

$$\eta(w)=\min-\max(|V_{a}(w)-\frac{1}{2},\frac{A_{r}(w)}{2}|_{2}) \tag{3}$$

For a word w, first retrieve its related words in ConceptNet: $w_r^1, w_r^2, ..., w_r^n$ and then calculate the sentiment value of the word using Equation 3.

c. Node Connection: For utterance words link them in order, with separators in the middle to distinguish different utterances, and for sentiment words link them bi-directionally and utterance word nodes.

d. Weight Assignment: The sentiment values of sentiment word obtained through the computation are assigned to the edges where the sentiment words are linked to the utterance words, where the weight between the utterance words is assigned to 1.

After obtaining the sentiment subgraph through the above steps, feature extraction is performed using GCN as shown in Equation 4:

$$\mathbf{E}^{(l)}=\sigma(\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^{(l-1)}\mathbf{W}^{(l)}) \tag{4}$$

Finally, the sentiment-enhanced representations of all words of a sentence are merged as shown in Equation 5, which is then used as the final sentiment feature representation corresponding to the utterance entity E.

$$\mathbf{E}=\sum_{i=1}^{m}\frac{e^{n_i}}{\sum_{j=1}^{n}e^{n_j}}w_e^i \tag{5}$$

3.2. Recommendation Module for Act Emotion Enhancement

3.2.1. ENTITY-UTTERANCE FEATURE FUSION.

Since connectivity relationships are also important in knowledge graphs, a Relational Graph Convolutional Network (R-GCN) (Lehmann et al., 2014) is used to learn all the entity embeddings N.

This paper fuses utterance act and sentiment features for user preferences, as shown in Equation 6.

$$\begin{aligned} Act &= \mathrm{SF}(\mathbf{C}_e, \mathbf{A}, \mathbf{A}) \\ Emo &= \mathrm{SF}(\mathbf{C}_e, \mathbf{E}, \mathbf{E}) \\ \mathrm{SF}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \mathrm{softmax}(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}})\mathbf{V} \end{aligned} \tag{6}$$

They are fused with each other using the gating mechanism, as shown in Equation 7.

$$\begin{aligned} \mathbf{U} &= \beta \cdot Act + (1-\beta) \cdot Emo \\ \beta &= \sigma(\mathbf{W}_{\mathrm{gate}}\left[Act; Emo\right]) \end{aligned} \tag{7}$$

3.2.2. ACT-EMOTION ENHANCED RECOMMENDATION.

With the user representation U, a multilayer perceptual machine and softmax are used to obtain the predicted recommendation probability as shown in Equation 8:

$$\mathrm{P}_{\mathrm{rec}}(i) = \mathrm{softmax}(\mathbf{U} \cdot \mathbf{I}) \tag{8}$$

Finally, optimization is performed using cross entropy as a loss function:

$$\mathcal{L}_{\mathrm{rec}} = -\sum_{j=1}^{M}\sum_{i=1}^{N} y_{ij} \cdot \log(\mathrm{P}_{\mathrm{rec}}^{j}(i)) \tag{9}$$

3.3. Conversation Module for Act Emotion Enhancement

3.3.1. NEXT ACT PREDICTION.

We proposes a distance-aware attention to infer the next utterance potential act a_{t+1} based on the potential act sequences of the past utterance $\mathrm{A} = \{a_1, a_2, ..., a_i, ...a_t\}$.

$$a_{t+1} = \sum_{i=1}^{t} \frac{\lambda_{i-1}}{\sum_{i=1}^{t}\lambda_{i-1}} a_i \tag{10}$$

which means that the recently appeared utterance act will have a greater influence on the next utterance act.

3.3.2. ACT-EMOTION ENHANCED UTTERANCE GENERATION.

The context is first encoded using a standard Transformer (Sun and Zhang, 2018). At the decoding step, in order to generate empathic utterance replies, [CLS] in the generated token is replaced with the next utterance act at+1. As shown in Equation 11:

$$R^{n-1} = \left[a_{t+1}; \left[1 : R^{n-1}\right]\right] \tag{11}$$

In the specific utterance generation process, based on Transformer, external information is gradually fused through the transformation chain.

In the final generation part, replies with relevant recommendation items are generated through a copying mechanism Li et al. (2019), which is computed as follows:

$$\Pr\left(y_i|\left\{y_{i-1}\right\}\right) = \underset{1}{\Pr}\left(y_i|\mathbf{Y}_i\right) + \underset{2}{\Pr}\left(y_i|\mathbf{Y}_i, KG\right) + \underset{3}{\Pr}\left(y_i|\mathbf{Y}_i, \mathbf{E}\right) \tag{12}$$

where is the output of the decoder, is the probability function of generating common words from the vocabulary. Cross entropy loss is set to optimize the generation of responses in the conversation module:

$$\mathcal{L}_{gen} = -\frac{1}{N}\log \Pr(s_t|s_1, ..., s_{t-1}) \tag{13}$$

4. Experiments

4.1. Dataset

We used the dataset REDIAL to evaluation. This dataset contains 10,006 complete conversations mentioning 51,699 movies with a total of 182,150 words. The training set, validation set, and test set in the ratio of 8:1:1.

4.2. Evaluation Metrics

For the recommendation task, the recall ReCall@k (k=1,10,50) is used to measure the accuracy of the recommendation results. For the conversation task, using Distinct n-grams (n=2,3,4), defined as the number of words with different n-grams divided by the total number of words.

4.3. Comparison Experiments

There are two sub-tasks to be evaluated: the recommendation task, and the conversation task, and representative models are selected for comparison experiments to validate the effectiveness of the model proposed in this paper.

- TextCNN (Ji et al., 2017): this model applies a CNN-based model to extract user features from the conversation context to rank the items.
- Transformer (Ren et al., 2023): an Encoder-Decoder based approach to generate conversations.
- ReDial (Sun and Zhang, 2018): the dialog module of this model is based on HRED (Li et al., 2019) and the recommendation module is built based on RNN.
- KBRD (Chen et al., 2019): this model introduces an external knowledge graph DBpedia to enhance the semantic information of context items.
- KGSF (Zhou et al., 2020): this model using mutual information maximization to align the semantic space.
- CRFR (Ren et al., 2023): this model uses reinforcement learning to perform explicit multi-hop inference on the knowledge graph, flexibly learning multiple inference segments.

- RevCore (Lu et al., 2021): this model introduces unstructured external data: item reviews, which generates well-explained responses.

4.3.1. RECOMMENDED TASK EVALUATION.

For the recommendation task, the results are shown in Table 2. The method proposed in this paper outperforms all baselines. Firstly, utterance act and utterance sentiment are extracted, and the semantic information expressed by the user is fully considered; previous models only considered the entities mentioned by the user, and did not deeply consider the semantic information.

Table 2: Results on the recommendation task. Best results are in bold.

Models	Recall@1	Recall@10	Recall@50
TextCNN	0.013	0.068	0.191
ReDial	0.024	0.140	0.320
KBRD	0.031	0.150	0.336
KGSF	0.039	0.183	0.378
CRFR	0.040	0.202	0.399
RevCore	0.046	0.220	0.396
Ours	**0.053**	**0.238**	**0.439**

4.3.2. CONVERSATION TASK EVALUATION.

The results of the conversation task evaluation are shown in Table 3. Compared to these baselines, our model performs better on all evaluation metrics. The model proposed in this paper further enhances the item representation by incorporating utterance act and sentiment features, and improves the diversity of system responses in the conversation context. In terms of automatic evaluation, as shown in Table 3.

Table 3: Results on the conversation task. Best results are in bold.

Models	Dist-2	Dist-3	Dist-4
Transformer	0.148	0.151	0.137
ReDial	0.225	0.236	0.228
KBRD	0.263	0.368	0.423
KGSF	0.289	0.434	0.519
CRFR	0.415	0.562	0.605
RevCore	0.424	0.558	0.612
Ours	**0.446**	**0.593**	**0.645**

4.4. Discussion

This subsection analyzes the effect of the number of emotion words selected on the model, as shown in Figure 2 and Figure 3, it can be seen that when the number of emotion words increases, with the increase in the number of emotion words, there is a decrease in the performance of the model, which

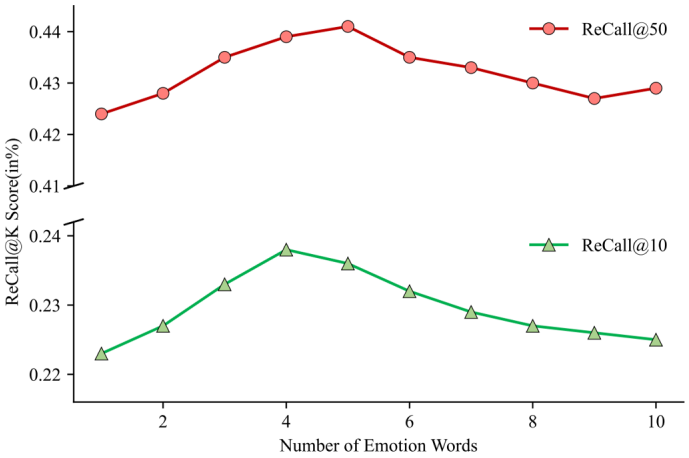

Figure 2: The impact of the number of emotional words on recommendation task.

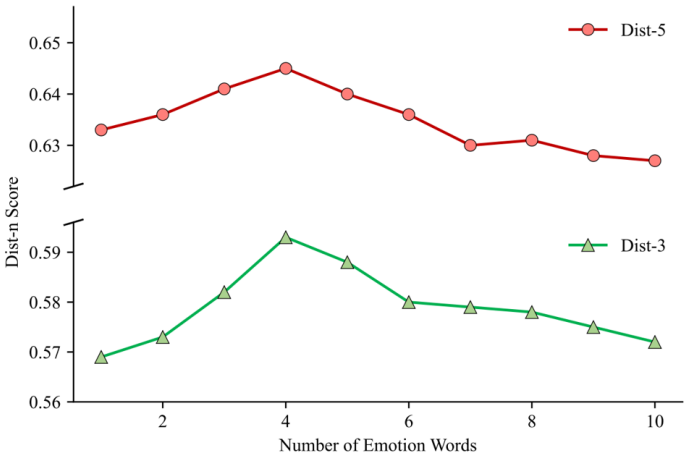

Figure 3: The impact of the number of emotional words on conversation task.

we believe is the noise caused by too many external words, so the number of emotion-related words is finally selected as 4, which balances the effect of the model.

5. Conclusion

In this paper, we propose a novel conversational recommender system through utterance acts and emotions. Firstly, we learn the utterance latent acts by clustering the utterance, and construct sentiment subgraph based on external knowledge. Finally, the proposed fusion module fuses utterance acts and emotions with entity preferences to obtain user preferences for recommendation. For conversations, the distance-aware attention is used to predict the next utterance act, thus improving the smoothness of user interaction with the system and the experience of using it.

References

Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. Towards knowledge-based recommender dialog system. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1803–1813, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1189.

Yang Deng, Wenxuan Zhang, Wai Lam, Hong Cheng, and Helen Meng. User satisfaction estimation with sequential dialogue act modeling in goal-oriented conversational systems. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 2998–3008, New York, NY, USA, 2022. Association for Computing Machinery. doi: 10.1145/3485447.3512020.

Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. Advances and challenges in conversational recommender systems: A survey. *AI Open*, 2:100–126, 2021. doi: 10.1016/j.aiopen.2021.06.002.

Pan Ji, Tong Zhang, Hongdong Li, Mathieu Salzmann, and Ian Reid. Deep subspace clustering networks, 2017.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, and Christian Bizer. Dbpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 6, 01 2014. doi: 10.3233/SW-140134.

Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. Conversational recommendation: Formulation, methods, and evaluation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 2425–2428, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3397271.3401419.

Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. Towards deep conversational recommendations, 2019.

Yu Lu, Junwei Bao, Yan Song, Zichen Ma, Shuguang Cui, Youzheng Wu, and Xiaodong He. RevCore: Review-augmented conversational recommendation. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1161–1173, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-acl.99.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, 7:100943–100953, 2019. doi: 10.1109/ACCESS.2019.2929050.

Xuhui Ren, Tong Chen, Quoc Viet Hung Nguyen, Lizhen Cui, Zi Huang, and Hongzhi Yin. Explicit knowledge graph reasoning for conversational recommendation, 2023. URL , .

Yueming Sun and Yi Zhang. Conversational recommender system. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, page 235–244, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3209978.3210002.

Chengyang Zhang, Xianying Huang, and Jiahao An. Macr: Multi-information augmented conversational recommender. *Expert Systems with Applications*, 213:118981, 2023. doi: 10.1016/j.eswa.2022.118981.

Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.*, 52(1), feb 2019. doi: 10.1145/3285029.

Peixiang Zhong, Di Wang, and Chunyan Miao. Knowledge-enriched transformer for emotion detection in textual conversations. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 165–176, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1016.

Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 1006–1014, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3394486.3403143.