Sentiment Adaptive End-to-End Dialog Systems

Weiyan Shi, Zhou Yu
University of California, Davis
wyshi@ucdavis.edu, joyu@ucdavis.edu

Abstract

❖ This work focuses on incorporating **sentiment** information into task-oriented dialogue systems.
❖ Current end-to-end approaches only consider user semantic inputs in learning and under-utilize other user information.
❖ But the ultimate evaluator of dialog systems is the end-users and their sentiment is a direct reflection of **user satisfaction** and should be taken into consideration.
❖ Therefore, we propose to include user sentiment obtained through multimodal information (acoustic, dialogic and textual), in the end-to-end learning framework to make systems more **user-adaptive** and effective.
❖ We incorporated user sentiment information in both **supervised** and **reinforcement learning** settings.
❖ In both settings, adding sentiment information **reduced the dialog length** and **improved the task success rate** on a bus information search task.

Multimodal Sentiment Detector

❖ We manually annotated 50 dialogs with 517 conversation turns to train this sentiment detector. The annotated set is open to public.
❖ Prediction made by the detector will be used in the supervised learning and reinforcement learning.
❖ Three sets of features: 1) Acoustic features; 2) Dialogic features; 3) Textual features.
❖ Dialogic features include: 1) Interruption; 2) Button usage; 3) Repetitions; 4) Start over. These four categories of dialog features are chosen based on the previous literature and the observed statistics in the dataset.

Model	Avg. of F-1	Std. of F-1	Max of F-1
Acoustic features only	0.635	0.027	0.686
Dialogic features only	0.596	0.001	0.596
Textual features only *	0.664	0.010	0.685
Textual + Dialogic *	0.672	0.011	0.700
Acoustic + Dialogic *	0.680	0.019	0.707
Acoustic + Textual	0.647	0.025	0.686
Acoustic + Dialogic + Text *	**0.686**	**0.028**	**0.756**

Table 1. Sentiment detector performance.

Dialogic Features	Relative Rank of importance
total interruptions so far	1
interruptions	2
total button usages so far	3
total repetitions so far	4
repetition	5
button usage	6
total start over so far	7
start over	8

Table 2. Feature importance ranking.

Supervised Learning

❖ *Hybrid Code Network (HCN)* (Williams et al. (2017)) is adopted as the baseline model.
❖ No action masks (bit vectors indicating allowed actions) are used, making our model end-to-end trainable and less labor-intensive.
❖ We added two sets of features to the baseline model: 1) eight raw dialogic features; 2) one-hot vector of the sentiment labels predicted by the sentiment detector.
❖ HCN with predicted sentiment labels performs the best, while adding raw dialogic features doesn't help because the predicted labels is more condensed than the noisy raw features..

Model	Weighted F-1	Dialog Acc.
HCN	0.4198	6.05%
HCN + raw dialogic features	0.4190	5.79%
HCN + predicted sentiment label*	**0.4261**	**6.55%**

Table 3. Supervised learning performance.

Model Architecture

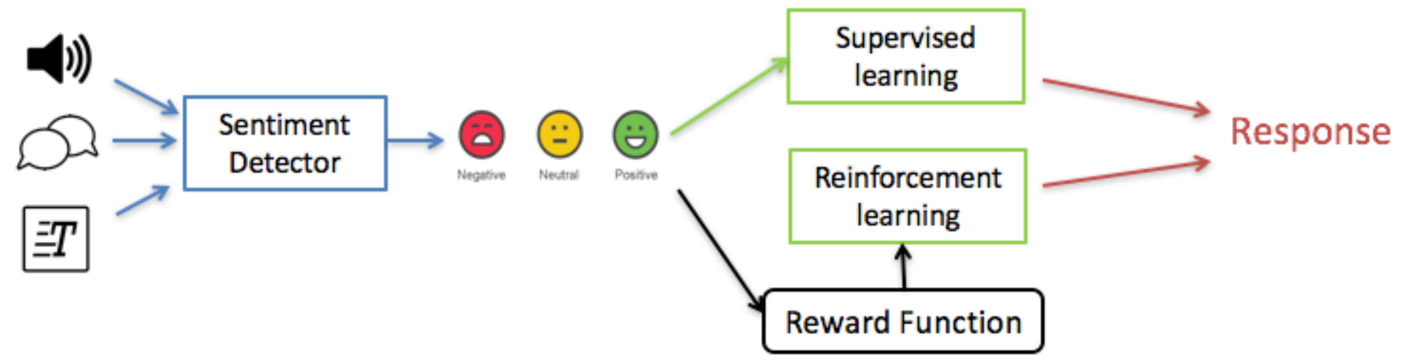

Figure 1. Proposed sentiment adaptive end-to-end dialog framework

❖ A sentiment detector is built on an annotated subset and is used to predict sentiment labels and sentiment scores for the supervised and reinforcement learning.
❖ Supervised learning uses the predicted sentiment labels from the sentiment detector as additional context features for the training.
❖ Reinforcement learning simulates the dialogs and uses the predicted sentiment scores from the sentiment detector as immediate rewards to guide the training.
❖ The whole model is end-to-end trainable and user-adaptive.

Reinforcement Learning

❖ **User simulator**

- Reinforcement learning requires feedback from the environments. So we created a user simulator and simulated user sentiment by **sampling from the real data.**
- Summary statistics, e.g. how many times one entity has been asked, are used to compare different dialogs.

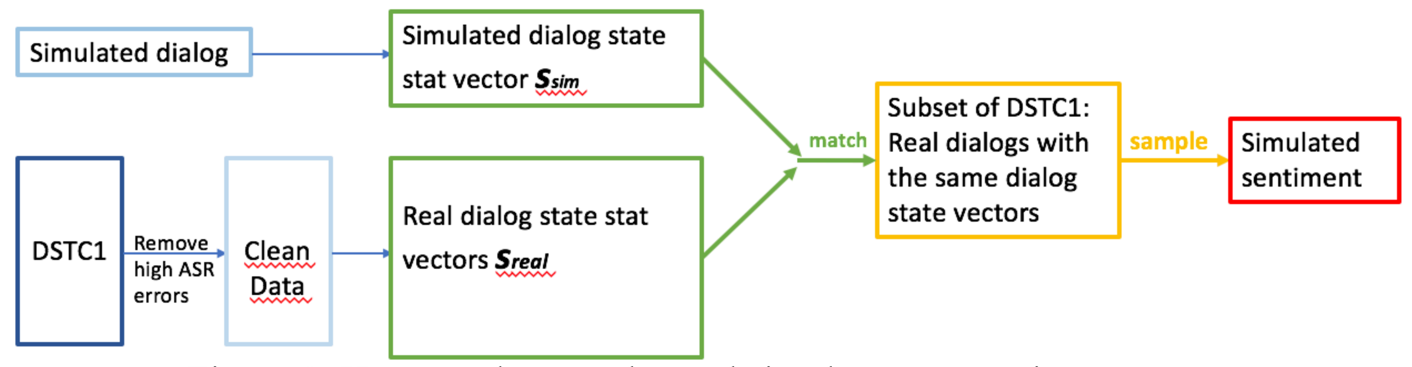

Figure 2. How to take samples and simulate user sentiment

❖ **Sentiment scores used in the reward functions**

- **Four different rewards functions with sentiment scores**. 1) baseline; 2) SRRS: baseline + sentiment score from random samples; 3) SRRP: baseline + penalty for repetitions; 4) SRRIP: baseline + penalties for both repetition and interruption.
- **Dialog length:** By adapting to user sentiment, all models with sentiment reward reduces the average dialog length.
- **Success rate:** SRRIP performs the best. By adding penalties, the model covers more data points, and improves the success rate and convergence speed.

Reward 4 SRRIP

if success **then**
  $R_4 = 20$
else if failure **then**
  $R_4 = -10$
else
  if match **then**
    if all-zero dialogic features **then**
      $R_4 = -1$
    else if non-zero dialogic features **then**
      $R_4 = -5P_{neg} - P_{neu} + 10P_{pos}$
    end if
  else if repeated question **then**
    $R_4 = -2.5$
  else
    $R_4 = -1$
  end if
  if interruption **then**
    $R_4 = R_4 - 1$
  end if
end if

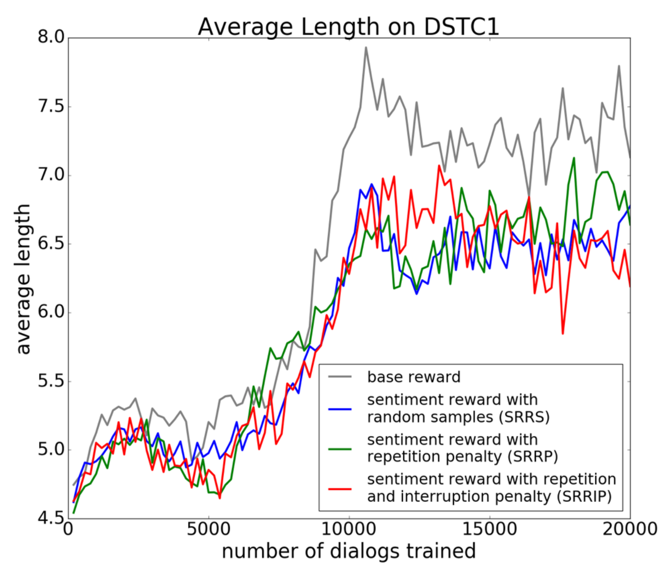

Figure 3. Dialog length in RL.

Model	Convergent success rate
Baseline	0.924
SRRS	0.938*
SRRP	0.941*
SRRIP	**0.943***

Table 4. RL convergent success rate.

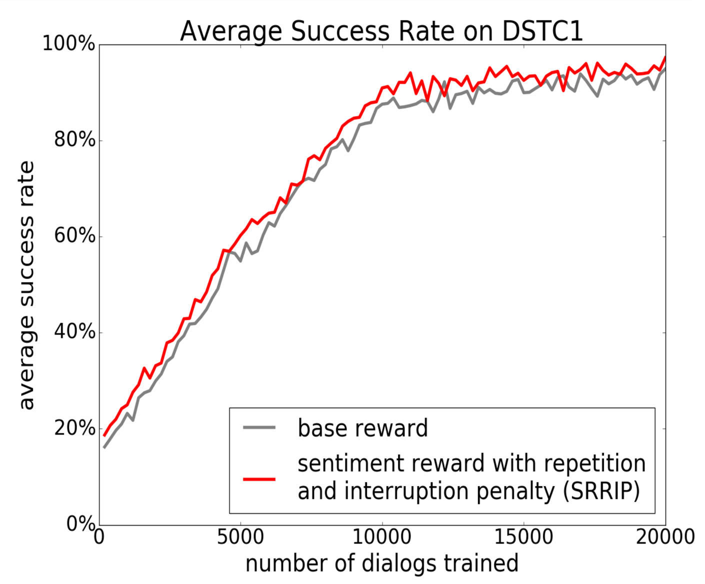

Figure 4. Success rate in RL for SRRIP.

Dataset & Discussion

❖ **Dataset: DSTC1, a bus information search task.**
1) **Sentiment Detector:** trained with a subset of 50 dialogs from DSTC1, with sentiment annotated under context.
2) **Supervised Learning:** trained and test with the entire DSTC1 set, with sentiment features predicted by the sentiment detector.
3) **Reinforcement Learning:** dialogs of the same task are simulated. The user sentiments are simulated by sampling from a subset of DSTC1.

❖ **The learned dialog policy is more sentiment adaptive**

The intuition behind the good performance of models with user sentiment is that the model learns to adapt to user sentiment.

For example, when the system detects negative sentiment, it will choose to give a more detailed error-handling strategy instead of a general one.

❖ **Further improvements**
1) Include more channels, such as vision, to improve the sentiment detector;
2) Create a similarity measure for the dialog states vectors and improve sentiment simulation
3) Reward shaping in reinforcement learning.

Sentiment Adaptive System	Baseline System without Sentiment
SYS: The <route>. Where would you like to leave from?	**SYS**: The <route>. Where would you like to leave from?
USR: *Yeah [negative sentiment]*	**USR**: *Yeah*
SYS: Where are you leaving from? For example, you can say, <place>.	**SYS**: Right. Where would you like to leave from?

Table 5. Sentiment adaptive example.

Conclusion

❖ We proposed to detect user sentiment from multimodal channels and incorporate the detected sentiment as feedback into adaptive end-to-end dialog system training.
❖ We included sentiment information directly as a context feature in the supervised learning framework and used sentiment scores as immediate rewards in the reinforcement learning setting.
❖ Experiments suggest that incorporating user sentiment is helpful in reducing the dialog length and increasing the task success rate in both SL and RL settings.
❖ We believe this approach can be easily generalized to other domains given its end-to-end training procedure and task independence.

References

▪ Jason Williams, Kavosh Asadi, and Geoffrey Zweig. 2017. Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning. In Proceedings of 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017). Association for Computational Linguistics.
▪ Dario Bertero, Farhad Bin Siddique, Chien-Sheng Wu, Yan Wan, Ricky Ho Yin Chan, and Pascale Fung. 2016. Real-time speech emotion and sentiment recognition for interactive dialogue systems. In EMNLP, pages 1042–1047.
▪ Xiujun Li, Zachary C Lipton, Bhuwan Dhingra, Lihong Li, Jianfeng Gao, and Yun-Nung Chen. 2016. A user simulator for task-completion dialogues. arXiv preprint arXiv:1612.05688.
▪ Stefan Ultes, Paweł Budzianowski, Inigo Casanueva, Nikola Mrksic, Lina Rojas-Barahona, Pei-Hao Su, ˇ Tsung-Hsien Wen, Milica Gasic, and Steve Young. ˇ 2017. Domain-independent user satisfaction reward estimation for dialogue policy learning. In Proc. Interspeech, pages 1721–1725.