# On System-Initiated Transitions in a Unified Natural Language Generation Model for Dialogue Systems

**Ye Liu**[1], **Yung-Ching Yang**[2], **Wolfgang Maier**[3], **Wolfgang Minker**[4] and **Stefan Ultes**[5]

[1,3,5]Mercedes-Benz AG, Sindelfingen, Germany

[1,4]Ulm University, Ulm, Germany, [2]University of Stuttgart, Stuttgart, Germany

[1,3,5]{ye.y.liu,wolfgang.mw.maier,stefan.ultes}@mercedes-benz.com
[2]yangching18@gmail.com, [4]wolfgang.minker@uni-ulm.de

## 1 Introduction and Motivation

Spoken Dialogue Systems (SDS) have been separately developed under two different categories: task-oriented and chit-chat. The former focuses on achieving functional goals and the latter aims at creating engaging social conversations without special goals. Creating unified models (Lin et al., 2021; Zhao et al., 2021; Young et al., 2022) that can reply to both task-oriented and chit-chat requests is a promising research topic in recent years. It is a desired property of unified models to be sensitive to dialogue mode transitions (switch from chit-chat to task-oriented or from task-oriented to chit-chat) and domain transitions in a task-oriented multi-domain setting (switch from one domain to another domain). The dialogue agent can then proactively guide the transition through generating a transition sentence (red parts in Figure 1).

We will discuss the *system-initiated transitions* of a unified dialogue agent. First, we elaborate the motivation behind this work. If the dialogue agent can track the preceding chit-chat interaction and recognize the potential user intention for requiring some task-oriented service, then the agent can proactively ask if the user needs this task-oriented service. As an example, consider the first part of Figure 1, where the agent realizes that the user wants to visit a "college" and actively guides to task-oriented interaction by saying "Do you want to visit some colleges?". It is beneficial for commercial dialogue systems to offer or sell their service (Chiu et al., 2022) at a right moment. This is also possible in a multi-domain scenario. If the agent is aware that user has the possibility to order other services after completing a task request, the agent can continually promote those services. Like in the third part of Figure 1, the user potentially needs the taxi service after booking a restaurant. This is also a good moment to actively offer this service in a commercial setting. When the human-machine

interaction starts with a task-oriented request, the users might have the feeling that they are talking to their friends if the system naturally switches to chit-chat interaction after providing all task-related information (see second example in Figure 1). This can highly improve the user interaction experience.
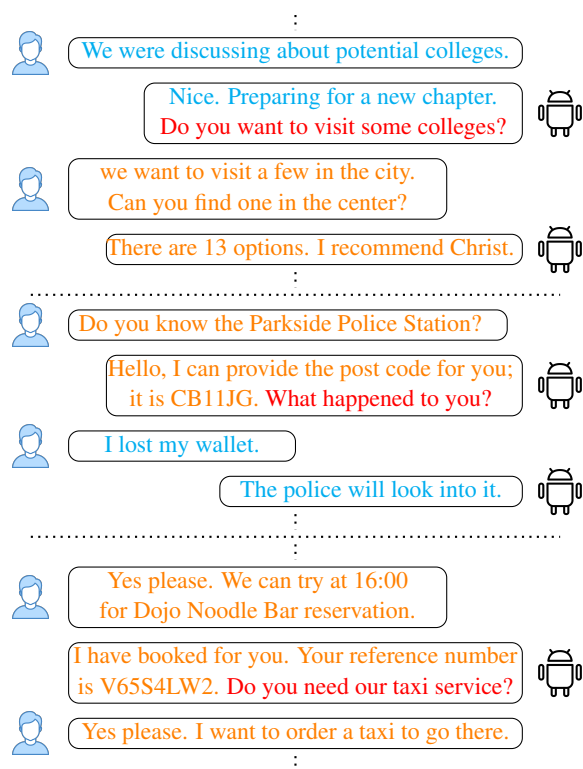


Figure 1: The system-initiated dialogue examples with chit-chat to task-oriented transition, task-oriented to chit-chat transition, and domain transition. The blue parts represent the chit-chat interaction, while the orange parts are task-oriented communication. The newly annotated transition sentences (red parts) demonstrate the system-initiated transitions are controlled by dialogue agent rather than user.

## 2 Initiative Discussion

(Walker and Whittaker, 1990) describe *initiative* as occasionally "taking the conversational lead".

For task-oriented dialogue, the initiative tends to represent "driving the task" (Smith, 1994; Smith and Gordon, 1997). Novick and Sutton (1997) introduce mixed-initiative interaction and describe initiative as a multi-factor concept, which includes choice of task, choice of speaker and choice of outcome. However, the formal definition of the term initiative is still missing from the literature. With the surge in interest in unified (Lin et al., 2021; Zhao et al., 2021; Young et al., 2022) models that can respond to both chit-chat and task-oriented user requests, we explore system-initiated transitions based on a unified model from **three** perspectives as follows (the first two are *dialogue mode* transitions, the third one is a *domain* transition):

- The system-initiated transition from chit-chat to task-oriented as in the first dialogue example shown in Figure 1, where the initiative agent captures the potential task-related information and proactively guides the switch at a proper moment.

- The system-initiated transition from task-oriented to chit-chat as in the second dialogue example shown in Figure 1, in which the system is aware of the completion of a task request and smoothly switches to chit-chat.

- The system-initiated transition from one domain to another in task-oriented interaction, as in the third example shown in Figure 1. Here, the dialogue agent is sensitive to the completion of current task request and proactively switch to another potential task domain.

## 3 Potential Challenges

Concerning the potential challenges of this work, we have the following questions that need to be precisely discussed and our opinion on these challenges is also elaborated here:

- **When is it a good moment for initiative transition?**[1] If the interaction starts with chit-chat, a good transition moment to switch to task-oriented mode could be when the agent captures some potential task-related information, which could be a task domain, an intent or a slot (such as "college" in the Figure 1). If the interaction starts from task-oriented, a

good transition moment could be the completion of the current request, followed by a decision to switch to another domain or to chit-chat interaction.

- **How to guide the generation of a transition sentence?** Transition sentences to chit-chat are hard to control, because they can be diverse and free in style. However, no matter whether it is switching from chit-chat to task-oriented, or from one task domain to another task domain, the system can generate the transition sentence based on relevant information it captures, such as a task domain or a slot.

- **How to evaluate the transition sentence generation?** Firstly, the evaluation on generation tasks is still a challenge in general (Chaganty et al., 2018). Additionally, we argue that the evaluation of transition sentences is even more difficult. One reason is that current publicly available datasets rarely have human annotated transition sentences as a reference to compute automatic metrics, such as BLEU (Papineni et al., 2002) or Meteor (Banerjee and Lavie, 2005). Another reason is that transition sentence generation is different in the three cases mentioned in Section 2, so the evaluation emphasis might be also different.

## 4 Future Work

In our future work, we will utilize the FusedChat dataset (Young et al., 2022), where human annotated open-domain sentences were prepended and appended to the dialogues of the task-oriented dataset MultiWOZ (Budzianowski et al., 2018; Ye et al., 2021). Hence, every dialogue in FusedChat includes two dialogue modes with the chit-chat and task-oriented parts being interdependent. In addition, many task-oriented dialogues in MultiWOZ include multi-domain interaction. We will first build a unified model that can reply to chit-chat and task-oriented requests with the FusedChat dataset. After that, we will apply the efficient prompt learning (Liu et al., 2021; Li et al., 2022) method to activate the initiative transition of the unified model so it can be sensitive to the timing of transitions and proactively guide them by generating transition sentences.

---

[1]The question of "when" to switch has also been addressed for pro-activity, which is similar to our initiative switch, by (Nothdurft et al., 2015).

# References

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. 2018. Multiwoz-a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026.

Arun Chaganty, Stephen Mussmann, and Percy Liang. 2018. The price of debiasing automatic metrics in natural language evalaution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 643–653.

Ssu Chiu, Maolin Li, Yen-Ting Lin, and Yun-Nung Chen. 2022. Salesbot: Transitioning from chit-chat to task-oriented dialogues. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6143–6158.

Lei Li, Yongfeng Zhang, and Li Chen. 2022. Personalized prompt learning for explainable recommendation. *arXiv preprint arXiv:2202.07371*.

Zhaojiang Lin, Andrea Madotto, Yejin Bang, and Pascale Fung. 2021. The adapter-bot: All-in-one controllable conversational model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 16081–16083.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.

Florian Nothdurft, Stefan Ultes, and Wolfgang Minker. 2015. Finding appropriate interaction strategies for proactive dialogue systems—an open quest. In *Proc. of the 2nd European and the 5th Nordic Symposium on Multimodal Communication 2014*, pages 73–80. LiU Electronic Press.

David G Novick and Stephen Sutton. 1997. What is mixed-initiative interaction. In *Proceedings of the AAAI spring symposium on computational models for mixed initiative interaction*, volume 2, page 12.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ronnie W Smith. 1994. Spoken variable initiative dialog: An adaptable natural-language interface. *IEEE Expert*, 9(1):45–50.

Ronnie W Smith and Steven A Gordon. 1997. Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialogue. *Computational Linguistics*, 23(1):141–168.

Marilyn Walker and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78.

Fanghua Ye, Jarana Manotumruksa, and Emine Yilmaz. 2021. Multiwoz 2.4: A multi-domain task-oriented dialogue dataset with essential annotation corrections to improve state tracking evaluation. *arXiv preprint arXiv:2104.00773*.

Tom Young, Frank Xing, Vlad Pandelea, Jinjie Ni, and Erik Cambria. 2022. Fusing task-oriented and open-domain dialogues in conversational agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11622–11629.

Xinyan Zhao, Bin He, Yasheng Wang, Yitong Li, Fei Mi, Yajiao Liu, Xin Jiang, Qun Liu, and Huanhuan Chen. 2021. Unids: A unified dialogue system for chit-chat and task-oriented dialogues. *arXiv e-prints*, pages arXiv–2110.