Towards Understanding Mixed-Initiative in Task-Oriented Dialogues

Fan Yang, Peter A. Heeman, Kristy Hollingshead

Center for Spoken Language Understanding
OGI School of Science & Engineering
Oregon Health & Science University
{fly,heeman,hollingk}@cslu.ogi.edu

Abstract

Mixed-initiative in dialogue does not mean that either speaker is free to take the lead in the conversation at any time. Rather, who can show initiative, and when, is restricted. In this paper, we define initiative and control as two levels of dialogue phenomena, and give further evidence for our theory of restricted initiative that control belongs to the initiator of a discourse segment. We show that the initiator and the non-initiator play different roles in terms of showing initiative.

1. Introduction

Initiative involves actively contributing to a dialogue. Researchers are interested in building mixed-initiative dialogue systems that users can interact with as easily as they interact with other people. On one hand, a dialogue system should allow human users to actively participate in the reasoning procedure; for example, to ask a question or to make a proposal. On the other hand, a dialogue system also needs to actively contribute to a dialogue, efficiently and effectively helping human users to accomplish a task. Thus, a dialogue system should know when to take the lead and when to let the human user take the lead. There are conventions in human-human dialogues that govern this behavior. Such conventions help guarantee an interaction mode that is convenient to human users and supports human-style problem solving.

A number of theories have been proposed for how initiative is managed. Whittaker and Stenton [1] proposed that initiative does not simply bounce back and forth, but that once one person has it, the person tends to keep it for a while. That is, the person who shows initiative is very likely to show the next initiative. Chu-Carroll and Brown [2] proposed that there are two levels of initiative, with dialogue initiative subservient to task initiative. Guinn [3] and Allen [4] proposed that whoever is most capable of the next (sub-)task takes initiative. Cohen et al. [5] proposed four different models of initiative, the most elaborate being where a speaker has initiative in a process; and they also proposed different degrees of initiative. However, none of these theories analyzed initiative from the viewpoint of discourse structure, or tied it to the conversational goals of the participants.

In our previous work [6, 7], we analyzed a subset of the Trains dialogues [8] using Grosz and Sidner’s discourse structure theory [9] and annotated them as to which utterances demonstrate initiative. We found that initiative tends to be mainly shown by the speaker who initiated a discourse segment. When the other speaker, the non-initiator, shows initiative, either by making an utterance or a subsegment, the initiator tends to be the one who shows initiative next. We used this as evidence that initiative is subordinate to discourse structure.

This paper expands on our previous work in the following ways. First, we clarify the definitions of initiative and control, and we restate our theory in terms of these definitions. Second, we give further evidence that control is subservient to discourse structure. We re-evaluate our past results to take into account the different roles of the two participants in the Trains corpus, the system and the user. We also look at the patterns of initiative.

2. Redefining Initiative and Control

Most of the work on defining mixed-initiative has used the terms control and initiative interchangeably. These terms seem to be used at two separate levels. The first is at the utterance level. When a speaker makes a certain type of utterance, the speaker is said to be showing control or initiative. The second is at a more global level, and is something that the speaker has that lives on beyond individual utterances. We treat these two phenomena as separate, and use initiative to refer to the first, and control to the second.1

Definition 1: Initiative is the behavior of volunteering new information without being explicitly requested to, or of asking a question to establish mutual belief, or of proposing a new domain goal to achieve.

1Cohen et al. [5] used these two terms opposite from our use of them.
Initiative is short-lived. Initiative begins and ends with the utterance. In the dialogue excerpt in Figure 1, the user shows initiative in utterance u1, but nobody shows initiative in utterance u2.

<table>
<thead>
<tr>
<th>user</th>
<th>system</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td>u2</td>
</tr>
<tr>
<td>I need to get a boxcar at Avon</td>
<td>ok</td>
</tr>
</tbody>
</table>

Figure 1: An example for initiative and control

**Definition 2:** Control is the authority to drive a dialogue (or sub-dialogue) towards the achievement of its goal.

In contrast to initiative, control has lasting effect. Once a speaker has control, he or she will have it until the other speaker has it.

Our definition of control does not specify its relation to initiative. This is what a theory of mixed-initiative must do. The theory must also specify how and when control changes. According to Whittaker and Stenton’s theory [1], when a speaker shows initiative, this gives control to that speaker, and the speaker keeps control until the other speaker shows initiative. For the examples in Figure 1, since the user shows initiative in u1 and the system does not show initiative in u2, the user is said to have control in u1 and u2.

According to our theory of mixed-initiative [6, 7], control follows discourse structure. The speaker, say A, who initiated the discourse segment is said to be in control. If the non-initiator, B, makes an utterance that shows initiative, this does not give B control. The same is true when B makes an utterance that starts an embedded discourse segment. B has control for the extent of the embedded discourse segment, but once that segment finishes, control reverts to A. Grosz and Sidner [9] tied discourse segments with conversational goals; hence, when a speaker proposes a conversational goal, that speaker is in control over the extent of the conversation to achieve that goal.

In order for our theory of mixed-initiative to be sound, there must be differences in behavior between the initiator of a discourse segment and the non-initiator. In other words, the ‘authority to drive a dialogue’ must have some manifestation. In our previous work [6, 7], we showed that the initiator tends to show initiative 8.3 times more often than the non-initiator and that after the non-initiator’s utterance of initiative or subsegments, the next person to show initiative tends to be the initiator of the current discourse segment.

**3. Dialogue Annotations**

We used the same annotations as were used in [7]. We annotated hierarchical discourse segments according to Grosz and Sidner [9]. In deciding whether a group of utterances formed a discourse segment, we took into account whether there existed a shared and recognizable goal. We also followed Carletta et al. [10] by distinguishing two types of segments: transaction segments and dialogue games. Transaction segments typically correspond to sub-tasks, such as stating the goal of the plan, constructing the steps in the plan, refining the plan, summarizing the plan and evaluating the plan. Dialogue games are related to adjacency pairs, which typically consists of a first part and a second part. Each part can have one or more utterances. We explicitly mark the second part as a special type of segment. Second parts are special in that control should be viewed as explicitly passed to the non-initiator of the dialogue game. In a dialogue game, the initiator gives up control to the non-initiator after the first part. The non-initiator of the dialogue game takes over the control when initiating the second part, and relinquishes the control when completing the segment. We call such phenomena delegating control. We did not analyze what happens inside a second part. Here we just assume that the second part as a whole does not show initiative. However, sometimes the non-initiator of a dialogue game might over-answer questions; sometimes the initiator of the dialogue game might even show initiative inside a second part. These are interesting behavior that need further research.

We did not directly code which utterances show initiative. Instead, we tagged each utterance with a variation of the DAMSL scheme [11]. Utterances were tagged as forward or backward functions, or others. Forward functions include statements, questions, checks and suggestions. Backward functions include agreements, answers, acknowledgments, repetitions and completions. Others include stalls and move-ons. We used utterance tags to derive whether an utterance shows initiative. Utterances coded as forward functions show initiative while backward functions and others do not.

We annotated twelve dialogues totaling 50 minutes from the Train corpus using DialogueView [12]. Two expert annotators annotated each dialogue for utterance tags and discourse structure. They then compared their annotations to reach a consensus using ACT [13]. Most of the differences between the annotators were resolved successfully.

**4. Evidence**

In this section, we build on our previous work to give further evidence for our theory of mixed-initiative interaction.

**4.1. Frequency of Initiative**

In our previous work [6, 7], we reported on how frequently the initiator and non-initiator of a discourse segment show initiative. However, we did not distinguish between the two roles that the speakers play.

In the Trains dialogues, one speaker is playing the role of a user and the other is playing the role of the sys-
Table 1: Frequency of Initiative

<table>
<thead>
<tr>
<th>Initiative</th>
<th>System</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiator shows initiative</td>
<td>66</td>
<td>145</td>
</tr>
<tr>
<td>in first utterance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in other utterances</td>
<td>108</td>
<td>302</td>
</tr>
<tr>
<td>Non-initiator shows initiative</td>
<td>8</td>
<td>43</td>
</tr>
</tbody>
</table>

系统。用户必须制定一个计划来完成制造和运输货物以实现某个目标。系统拥有对领域知识的完全理解，比如城市之间的距离、制造产品所需的时间以及一列火车可以装载的货物重量的限制。系统还拥有笔和纸。用户需要提出一个计划来制造并运送货物以实现某个目标。系统将控制对话，并且非常可能表现出主动性。

的233个交易段和对话游戏中的被注释的对话中，69是由系统发起的，164是由用户发起的。表1报告了内部各发起人和非发起人的主动性数量。列2给出了结果，其中系统是发起人，列3给出了结果，其中用户是发起人。

我们将第一次发言视为一个段落，它的定义是发起人。在由系统发起的语句中，发起人的主动性13.5（108/8）次于非发起人（用户）。在由用户发起的语句中，发起人的主动性7（302/43）次于非发起人（系统）。尽管系统和用户在角色上有所不同，但我们发现它们的行为方式相似，当系统作为段落发起人时，它表现出主动性。这种支持我们的观点是回答人是对话的控制者，且在控制中。

4.2. After Non-initiator Shows Initiative

我们将系统模型的混合类型允许的参与者作为系统发起人的陈述，必须在一个段落中表现出主动性，并且在随后的段落中也是如此。在我们的前一工作[7]中，我们分析了在段落到段落之后发生了什么。我们发现下一句话要表现出主动性并且最有可能会是下一个人。这个结果与Whittaker和Stenton的混合类型理论一致，正如他们预测的那样，当非发起人表现出主动性时，非发起人将接棒并控制对话。因此，可以非常可能地接下来表现出主动性。

在这个子节中，我们将重复我们前一分析，即在段落到段落之后表现出主动性的人是发起人。通过检查不同角色在对话中所扮演的角色来理解这些情况，我们查看了表2中的数据结果。列2给出了结果。列2是系统为发起人且列3是用户为发起人。

在表2中，我们查看了当非发起人表现出主动性时的语句。列3给出了结果。列2是系统发起人且列3是用户发起人。

4.3. Initiative Patterns

前一子节中，我们发现当非发起人表现出主动性时，下一个表现出主动性的人是发起人。然而，在25/81的情况下，非发起人表现出主动性至少一次。要理解这些情况，我们需要查看段落之间表现出主动性的模式。

为了确定段落之间表现出主动性的模式，我们按照以下步骤进行。首先，任何嵌入的段落都被视为表现出主动性的人是发起人。其次，我们移除任何表现出主动性的人。第三，我们计算在段落中每个发言人所表现出的主动性数量。因此，模式“3-2-4-1”表示一个段落，在该段落中发起人表现出主动性，随后是2次主动性和1次主动性（或子段落）。因此，发起人表现出主动性，然后由非发起人接棒。出于上述原因，当发起人表现出主动性时，我们将其视为同一类型的模式。

我们对表现出主动性的模式进行分类，分为三种类型：好的模式、坏的模式和可疑的模式。好的模式是满足以下条件的模式。

**Majority:**发起人表现出至少80%的主动性。

例：5-1          Counterexample: 3-1

**Isolated:**非发起人表现出主动性，模式比一次。

例：4-1-3-1      Counterexample: 4-2-3-1
Non-alternating: no subpattern of 1-1-1 (starting with non-initiator).

Example: 8-1-3-1  Counterexample: 8-1-1-1

We feel that good patterns are clear examples where the initiator is driving the conversation. A pattern is bad if the initiator shows control less than 60% of the time. For bad patterns, it is not clear that the initiator is indeed driving the conversation. All other patterns are viewed as questionable. In a segment with a questionable pattern, the initiator tends to be controlling (showing most cases of initiative), yet there might be local extent where it is unclear who is the controller.

In our annotated dialogues, we focused on the 83 transaction segments. The results are reported in Table 3. In total we see that 75% of the segments have a good initiative pattern. More research work is needed to understand what is happening in segments with the other two patterns. It is also interesting to note that the majority of these segments were initiated by the user. These segments might be ones in which heavy collaboration is occurring in order to solve a difficult subproblem, and might correspond to the shared turns of Shriffin [14]. Further work is needed to investigate these segments.

5. Conclusion

In this paper, we give a clear definition of initiative and control. We show that the initiator and the non-initiator play different roles in terms of showing initiative. The initiator tends to show initiative continuously, with the non-initiator interrupting occasionally. Control belongs to the initiator. This proposal has important implications for dialogue management, as it will pave the way for building dialogue systems that can engage in mixed-initiative dialogues. Mixed-initiative dialogues do not mean that both speakers show initiative freely without any restrictions. Instead, mixed-initiative means that speakers contribute at the most appropriate timing when it is best suited. Our theory points out that the initiator is the person who mainly shows initiative, while the non-initiator mainly just responds. Human language has evolved over several generations and reached some local optimization. Perhaps people adopt such convention just because conversation in this way is more efficient. Our future work also includes understanding the benefits of using this model in human-human dialogues and in building dialogue systems.

6. References