Intention Recognition for Mixed-Initiative Recommender Systems *

Lorraine McGinty and Barry Smyth
Smart Media Institute, Department of Computer Science,
University College Dublin, Belfield, Dublin 4, Ireland.
Email: {firstname.lastname}@ucd.ie

Abstract

Recently ideas from mixed-initiative systems have been explored in the context of conversational recommender systems as a way to improve the interaction between the user and system. In this paper we examine some of the shortcomings of existing conversational recommender systems. In particular, we highlight how a more flexible recommendation strategy, one that responds to intermediate recommendation success and failures, can lead to significant improvements in both the efficiency and quality of recommendation dialogs. We argue that such techniques have a role to play in future mixed-initiative recommender systems.

1 Introduction

Conversational recommender systems are good examples of interactive artificial intelligence (IAI) systems [Aha and Muoz-Avila, 2001] as the end-user is engaged in an extended dialog with the recommendation engine. The user participates in a sequence of interactions with the recommender system providing feedback and additional information about their needs and preferences during each interaction. The form of feedback is important and distinguishes two important classes of recommendation strategy: those that employ *navigation by asking* and those that employ *navigation by proposing* [Shimazu, 2001]. In the former the user is asked specific questions about specific features within the product space. For instance, in a PC recommender: "How much are you willing to pay?". In contrast, navigation by proposing avoids asking direct questions and instead presents the user with a set of *best guesses* based on their current query, inviting them to provide feedback by either expressing a simple preference (eg. "I prefer PC number 2") [McGinty and Smyth, 2002] or a specific feature critique (eg. "I like PC number 2 but want more memory.")[Burke *et al.*, 1997].

In the past, conversational recommender systems have adopted a very rigid conversational structure with the user and recommender playing fixed roles within a turn-taking conversation [Aha and Muoz-Avila, 2001; Allen, 1999]. However, recently, research from the mixed-initiative systems field (e.g. [Allen, 1999; Bridge, 2002; McSherry, 2002a]) has led to the development of more flexible conversational strategies capable of taking better advantage of the relative strengths of the human and machine reasoning styles that are brought together within a recommendation dialog. For instance, mixed-initiative recommenders allow for a more flexible division of responsibility between user and recommender breaking away from rigid turn-taking [Bridge, 2002]. In addition, one of the important features of a mixed-initiative system is the maintenance of a *shared awareness* with respect to the current state of the human and machine agent(s). In the context of a recommender system this involves helping the user and system to better understand the current recommendation state, both in terms of the user's current and evolving requirements and the system's interpretation of how these requirements map on to the product-space. For example, [McSherry, 2002b] proposes the use of explanations as a way to help the user to better understand the intermediate suggestions of the system.

In this paper we are interested in a related issue, that of *intention recognition* [Allen, 1999; Horvitz and Paek, 1999; Louwerse *et al.*, 2002]. The ability of a recommender system to accurately recognise and respond to the intentions of a user obviously plays an important role in the success of any recommender system. We describe a flexible recommendation technique that uses feedback from the user to adapt their current query and assess whether or not recent recommendations were on-target. Two different recommendation strategies are proposed and a switching mechanism that allows the recommender to switch between them, depending on whether recent recommendations have been on target or not. This approach has been shown to deliver significant improvements in recommendation efficiency, resulting in shorter recommendation dialogs under a variety of experimental conditions [McGinty and Smyth, 2003; Smyth and McGinty, 2003].

In should be pointed out at this stage that the system described in this paper is *not* a mixed-initiative recommender system. Nevertheless, the approaches that we describe, implement and evaluate address relevant challenges in this area and the resulting techniques are potentially useful for true mixed-initiative recommenders.

*The support of the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

2 Hill-Climbing and False-Leads

One assumption that is often made in conversational recommender systems might be called the "hill-climbing assumption". In simple terms this refers to the assumption that each iteraction/cycle of a conversation recommender system tends to bring us closer to the appropriate target case/item [Shimazu, 2001]. However, this assumption does not always hold, especially in recommender systems that operate over complex data sets and that rely on critiquing [Burke *et al.*, 1997] or preference-based feedback [McGinty and Smyth, 2002; Smyth and McGinty, 2003]. For example, Figure 1 show the results obtained from graphing the *similarity profiles* of two different conversational recommender systems, one that employs critiquing and one that employs preference-based feedback, over two different domains (Travel and Whiskey - see Section 4). Each similarity profile is a graph of the similarity of the closest case in a recommendation cycle to the known target case against the number of recommendation cycles. And in each graph we see a similarity profile that breaks the hill-climbing assumption in the sense that many cycles lead to recommendations that are farther from the target case than earlier recommendations.

Three types of cycle transitions can be distinguished, based on the change in similarity between the best case in a given cycle and the target case; where "best case" refers to the case that is most similar to the target. An *ascent* occurs when the best case in a new cycle is more similar to the target than the best case from the previous cycle. A *plateau* occurs when the best case in the new cycle has the same similarity to the target as that from the previous cycle. Finally, a *descent* occurs when the best case from the new cycle is less similar to the target than the best case from the previous cycle.

In Figure 1(a) we see the similarity profile of a particular query from the Whiskey domain presented to a preference-based recommender system (PBF). Instead of an increasing similarity profile, we find that almost 50% of the cycles are descents for the PBF recommender. For instance, between cycle 7 and 11, target similarity falls from 0.66 to as low as 0.43 as the user is forced to accept poor recommendations. The PBF strategy fluctuates erratically between cycles 8 through to 54. In fact, for the duration of these 46 cycles the recommender conducts an exhaustive search for the target case in what is clearly the wrong region of the recommendation space, given that the user is not presented with any case during this period with a higher similarity to the target than the preference they indicated much earlier on in cycle 7! Obviously, it is unreasonable to expect that any user would interact with a recommender system for this many cycles. This is especially true when there is no evidence of *positive progress* (i.e. the system fails to consistently retrieve cases in each cycle that bring the user closer to the target). In reality, the user will abandon their search unless they see positive progress early on in the recommendation dialogue.

Similar profile trends are characteristic for the recommenders using other forms of feedback. For example, Figure 1(b) shows the similarity profile for the same query from the Whiskey domain presented to a recommender system that uses critiquing as it's feedback strategy; with accents (e.g. cycles 1 to 3), descents (e.g. cycles 3 to 6) and plateaus (e.g. cycles 22 to 24) for critiquing. See also Figure 1(c&d) for Travel domain results.

3 Intention Recognition in Comparison-Based Recommendation

Comparison-based recommendation is a generic framework for conversational recommender systems that emphasises the roles of case selection, user feedback, and query modification during navigation by proposing [McGinty and Smyth, 2002]. It is an iterative recommendation algorithm (Figure 2) that presents the user with a selection of k items as recommendations during each of a sequence of n recommendation cycles. Although initially comparison-based recommendation was proposed as a framework for investigating similarity-based recommenders utilising pure preference-based feedback, it is in fact sufficiently generic to accommodate a range of different recommendation strategies and feedback types [McGinty and Smyth, 2003].

In this work we are particularly interested in the interaction between user and system. Both parties contribute to each recommendation cycle. The recommender system contributes a set of recommendations and the user may provide some form of feedback as a response to these recommendations. Ultimately the job of the recommender system is to present the user with a recommendation that is satisfactory and this will only be achieved if a *shared awareness* is developed between system and user with respect to the goals and intentions of each. We are interested in how the feedback provided by the user can be effectively interpreted by the recommender system, not just as a way to update the user's query [Burke *et al.*, 1997; McGinty and Smyth, 2002], but also as a way to evaluate the relative success of the recent recommendation cycle.

The essential point is that determining whether or not the recommender is currently focusing on the correct portion of the recommendation space is important when it comes to deciding on the right recommendation strategy to use in the next recommendation cycle. While traditional similarity-based approaches to recommendation have been used extensively in the past – retrieving the k most similar cases to the query – we believe that such approaches are applicable only when the recommender system is correctly focused. If it is poorly focused then the current query is unlikely to be an accurate or complete representation of the user's true needs.

A pure similarity-based approach will have a tendency to over-fit to the assumptions that are made by the recommender about user needs, and it will retrieve cases that match these potentially flawed assumptions. These cases will also have a tendency to be very similar to each other, as well as being similar to the elaborated query, and as such will cover a very limited region of the underlying recommendation space. If incorrect assumptions have been made by the recommender then none of the recommended cases may be suitable, and indeed they may represent poorer recommendations than many of those that have been made previously in the current recommendation session. This is one of the basic reasons that the hill-climbing assumption breaks down in practice.

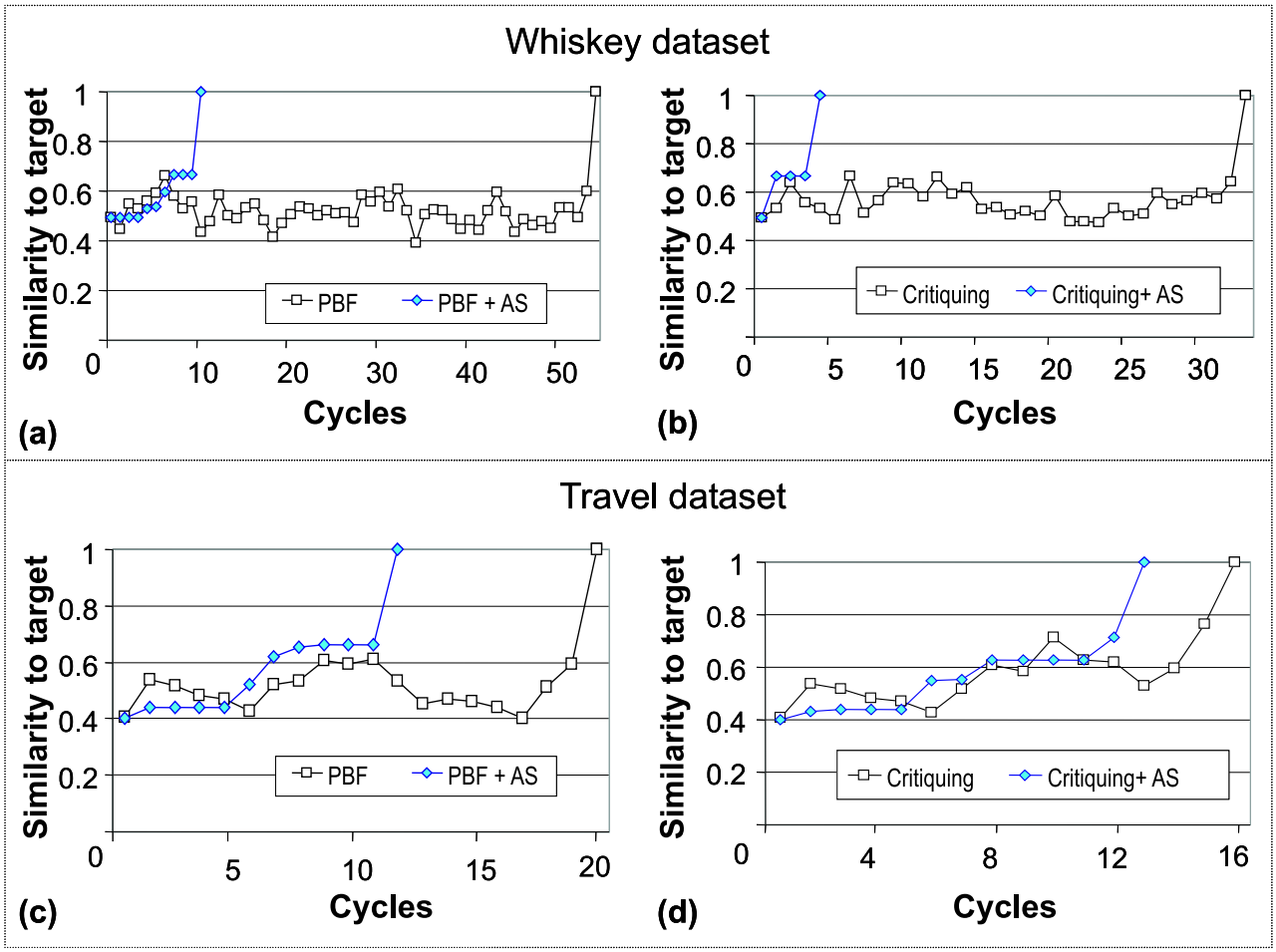

Figure 1: Similarity profiles for preference-based feedback(a&c) and critiquing (b&d) over Whiskey (a&b) and Travel (c&d) data-sets.

```
define Comparison-Based-Recommend(q, CB, k)
  i_{p-1} , i_p ← null
  do
   R ← ItemRecommend(q, CB, k, i_p, i_{p-1})
   i_p ← UserReview(R, CB)
   Q ← QueryRevise(q, i_p, R)
   i_{p-1} ← i_p
  until UserAccepts(i_p)

define QueryRevise(q, i_p, R)
  R' ← R - {i_p}
  q ← i_p
return q

define UserReview(R, CB)
  i_p ← user's preferred case from R
  CB ← CB - R
  return i_p

define ItemRecommend(q, CB, k, i_p, i_{p-1})
  if(i_p != null) && (i_p == i_{p-1})
     R ← ReFocus(q, CB, k)
  else
     R ← ReFine(q, CB, k)
  return R

define ReFine(q, CB, k)
  CB' ← sort CB in decreasing order of their sim to q
  R ← top k items in CB'
  return R

define ReFocus(q, CB, k, i_p, i_{p-1})
  return BoundedGreedySelection(q, CB, k, b)

define BoundedGreedySelection (q, CB, k, b)
  CB' := bk cases in CB that are most similar to q
  R := {}
  For j := 1 to k
    Sort CB' by Quality(q,i,R) for each case i in CB'
    R  := R + First(CB')
    CB' := CB' - First(CB')
  EndFor
  return R
```

Figure 2: The comparison-based recommendation algorithm with adaptive selection.

By presenting a more diverse set of cases the recommender can cover a number of different points in the recommendation space in the hope that one will be a fair match for the user's needs. Thus, in situations where there is evidence that the recommender system is not properly focused the importance of similarity in recommendation becomes less critical and recommendation diversity becomes more critical. These observations motivate the need for a more sophisticated recommendation strategy, one that adapts its response (i.e. it's use of similarity and diversity) depending on whether the recommender has focused in on the right region of the recommendation space. This is a significant departure for conventional conversational recommender systems and one that it vitally important in the context of mixed-initiative recommender systems. The bottom-line is that an efficient collaboration between user and recommender system can only be achieved if the system is capable of recognising when its own suggestions are suitable or unsuitable, adjusting its recommendation strategy accordingly.

Adaptive Selection (AS) is such a technique and is described in detail in [Smyth and McGinty, 2003]. Briefly, if takes advantage of the key idea that it is possible to determine whether or not the recommender is correctly focused by determining whether the recent recommendations represent an improvement on those made in the previous cycle. This is achieved by making two further modifications to the basic comparison-based recommendation technique. First, instead of making k new recommendations in each new cycle, the current preference case (or the critiqued case) is added to $k-1$ new recommendations. On its own this modification introduces redundancy, in the sense that a previously seen case is repeated in one or more future cycles. However, including the previous preference makes it possible to avoid the problems that ordinarily occur when none of the newly recommended cases are relevant to the user; the user can simply reselect the carried preference case instead of being forced to follow a less relevant recommendation.

The essential point is that if the user prefers (or critiques) a case other than the carried preference, then it must be because it is closer to the target, and thus positive progress has been made. In this situation diversity is not warranted and the emphasis should be on similarity in the next recommendation cycle. This corresponds to the **ReFine** method in Figure 2. If, however, the user prefers the carried preference case then it suggests that the other $k-1$ cases are less relevant than the carried case, and thus that the recommender has failed to make positive progress towards the target. In this situation two things happen (see **ReFocus** method in Figure 2). First, diversity is introduced into the next recommendation cycle. And secondly, during the selection of the new cases for the next recommendation cycle, the dissimilarity of these candidate cases to the rejected cases is taken into account. The basic idea is to prioritise cases that are not only similar to the query, but also dissimilar to the rejected cases. The algorithm components in Figure 2 are the modifications needed to implement adaptive selection for use in comparison-based recommendation with preference-based feedback. Similar modifications can be made when critiquing is the form of feedback being used [McGinty and Smyth, 2003].

4 Evaluation

In previous work we have extensively evaluated the impact of adaptive selection on recommendation efficiency, and reported significant decreases in recommendation cycles across a range of data-sets and for a variety of feedback strategies [McGinty and Smyth, 2003]. In this section we will re-evaluate the adaptive selection technique, this time looking at how it influences the type of recommendation cycles that are produced. In particular we are interested in whether or not it produces recommendation cycles that are consistent with the hill-climbing assumption. For example, the similarity profile graphs (Figure 1) in Section 2 also contain the similarity profiles generated by the adaptive selection technique, and it should be clear that these profiles (PBF+AS and Critiquing+AS) are characteristically different from those produced by the standard comparison-based recommendation techniques. They are shorter, requiring fewer interactions from the end-user, than those produced by the standard techniques. In Figure 1(a) the AS technique locates the correct target case in cycle 11, compared to cycle 55 for the standard preference-based feedback technique. We also see that none of the AS cycles are descents. All of the cycles are either ascents (where the user selects a recommended case that is closer to the target case) or plateaus (where the user reselects the preference case that has been carried).

The *Travel* case-base contains 1024 cases, each describing a vacation in terms of features such as *location*, *duration*, *accommodation*, *price* etc. The *Whiskey* case-base ([McGinty and Smyth, 2003]) contains a set of 552 cases, each describing a particular Scotch whiskey in terms of features such as *distillery*, *age*, *proof*, *sweetness*, *flavour*, *finish* etc. We wish to test four different recommender systems divided in to two groups of two. One group uses preference-based feedback and one uses critiquing. In each group one of the recommenders employs a standard similarity-approach, in which the k most similar cases to the current query are returned in each cycle, and one employs the AS approach which uses different recommendation strategies depending on whether or not the recommender is currently focused in an appropriate part of the recommendation space. Each recommender is implemented using the comparison-based recommendation framework with $k = 3$.

Using a leave-one-out methodology, each case (*base*) in a case-base is temporarily removed and used in two ways. First it serves as the basis for a set of queries constructed by taking random subsets of item features. Second, we select the case that is most similar to the original base. These cases serve as the recommendation *targets* for the experiments. Thus, the base represents the ideal query for a user, the generated query is the initial query that the user provides to the recommender, and the target is the best available case for the user based on their ideal. Each generated query is a test problem for the recommender, and in each recommendation cycle the users preference is assumed to be the case that is most similar to the known target case. Preference-based feedback or critiquing is applied to this preference case as appropriate; in the case of the latter, a random critique is applied to the preferred case in each cycle.

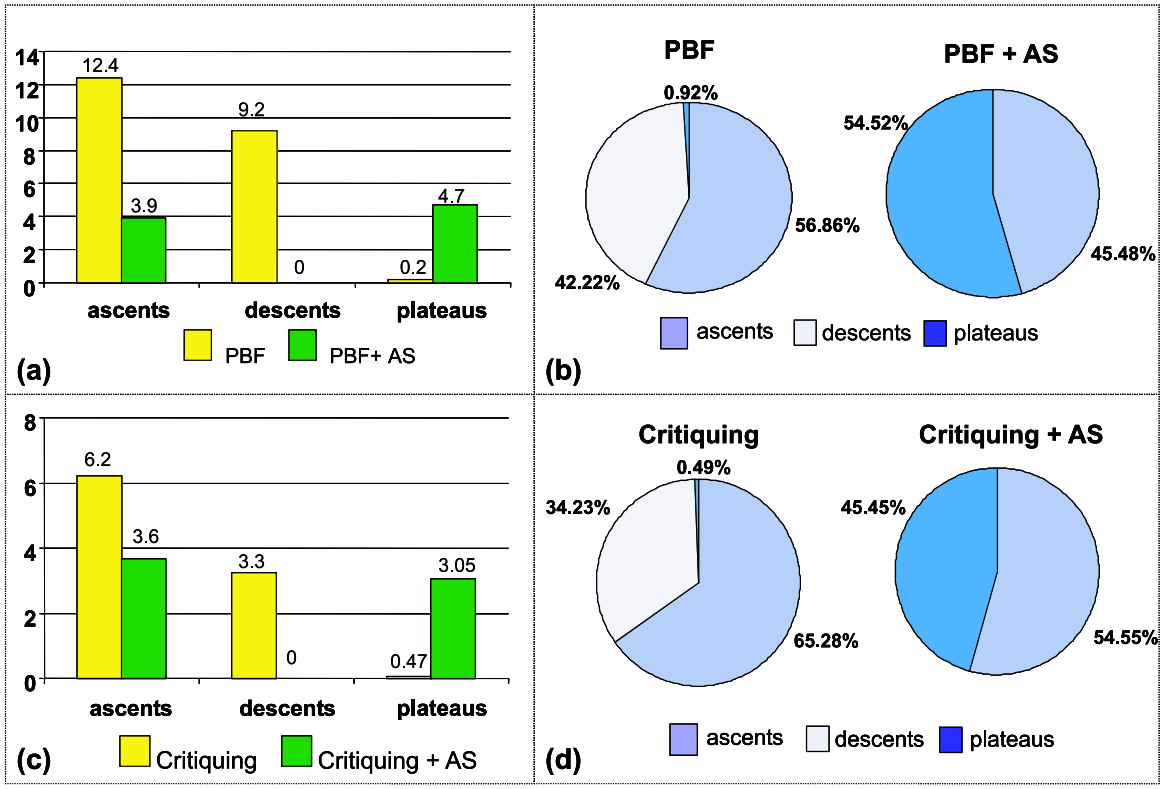

Figure 3: Evaluation results over the Travel data-set.

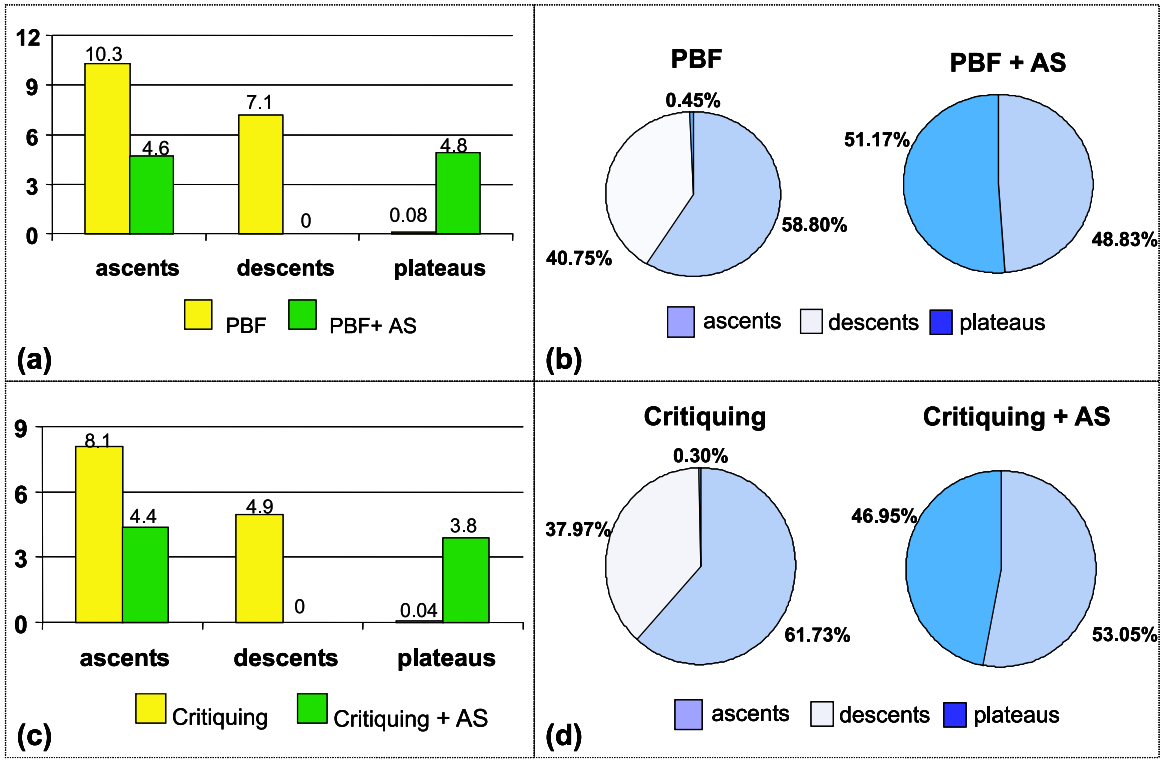

Figure 4: Evaluation results over the Whiskey data-set.

4.1 Results

The results are presented in Figures 3 and 4. For each test query and recommendation technique we measure the number of cycle transitions that are ascents, descents and plateaus. Figures 3(a&b) focus on the recommendation sessions in the Travel domain, comparing pure preference-based and critiquing forms of feedback to their AS counterparts. In Figure 3(a), which corresponds to preference-based feedback, we see that the average recommendation session is made up of 12.4 ascents, 9.2 descents, and 0.2 plateaus; an average session length of 21.8. With AS the average session length is reduced to 8.6, and descents are eliminated entirely. Similar results are found for the Travel domain with critiquing as the feedback mechanism (Figure 3(c)) and for the Whiskey domain (Figure 4(a&c)).

Overall the percentage of ascents is higher in critiquing-based recommenders. In the Travel data-set, 57% of cycles are ascents when preference-based feedback is used, but only 65% of cycles are ascents when critiquing is used (see Figures 3(b&d)). One view of this result concerns changes in the seat of control within the system. In a traditional conversational recommender system the user is very much in control selecting a *new* case during each cycle as the basis for the next. However, with AS the carrying of the previous preference case into the current cycle gives the user the opportunity to relinquish control in a given cycle by reselecting this carried case, effectively indicating that no improved recommendations have been presented by the recommender. This form of feedback is conceptually very different from the user selecting a new preference case, and indeed signals a change in recommendation strategy as indicated in the previous section. The AS results in Figure 3 and 4(b&d) indicate a relatively even sharing of control between the user and system with plateaus occurring between 45% and 55% of the time.

5 Conclusions

In this paper we have looked at a problem facing conversational recommender systems, namely that intermediate recommendations may be less relevant than earlier recommendations, which is likely to frustrate the user and lead to recommendation failures. The problem occurs when conversational recommenders follow false-leads and we propose that one solution is to help the recommender system to better understand user intentions and recognise inappropriate recommendations. Adaptive selection achieves this by determining whether or not the recommender is correctly focused on the user's needs and switches it's recommendation strategy accordingly. Evaluations show that this can improve recommendation efficiency [Smyth and McGinty, 2003] and in this paper we have shown that it improves the quality of recommendation dialogues. Overall recommendation dialogs are shorter and of higher quality.

In closing, we believe that the above ideas are relevant to the field of recommender systems in general, and particularly important for the development of mixed-initiative recommender systems where an improved understanding of the user, and the ability to monitor the success or failure of intermediate recommendations is likely to be important.

References

[Aha and Muoz-Avila, 2001] D.W. Aha and H. Muoz-Avila. Interactive case-based reasoning. *Applied Intelligence*, 14(1):7–8, 2001.

[Allen, 1999] J. Allen. Mixed-Initiative Interaction. *IEEE Intelligent Systems, Trends and Controversies*, 14(5):14–16, 1999.

[Bridge, 2002] D. Bridge. Towards Conversational Recommender Systems: A Dialogue Grammar Approach. In D. Aha, editor, *Proceedings of the Mixed-Initiative Workshop on CBR at the European Conference on Case-Based Reasoning (ECCBR-02).*, pages 8–22. The Robert Gordon University, 2002. Aberdeen, Scotland.

[Burke *et al.*, 1997] R. Burke, K. Hammond, and B.C. Young. The FindMe Approach to Assisted Browsing. *Journal of IEEE Expert*, 12(4):32–40, 1997.

[Horvitz and Paek, 1999] E. Horvitz and T. Paek. A Computational Architecture for Conversation. In *Proceedings of the Seventh International Conference on User Modeling.*, pages 201–210. Springer Wien, 1999. Banff, Canada.

[Louwerse *et al.*, 2002] M.M. Louwerse, A.C. Graesser, and A. Olney. Good computational manners: Mixed-initiative dialog in conversational agents. In C. Miller, editor, *Etiquette for Human-Computer Work. Papers from the 2002 Fall Symposium.*, pages 71–76, 2002.

[McGinty and Smyth, 2002] L. McGinty and B. Smyth. Comparison-Based Recommendation. In Susan Craw, editor, *Proceedings of the Sixth European Conference on Case-Based Reasoning (ECCBR-02)*, pages 575–589. Springer, 2002. Aberdeen, Scotland.

[McGinty and Smyth, 2003] L. McGinty and B. Smyth. Tweaking Critiquing. In *Proceedings of the Workshop on Personalization and Web Techniques at the International Joint Conference on Artificial Intelligence (IJCAI-03)*. Morgan-Kaufmann, 2003. Acapulco, Mexico.

[McSherry, 2002a] D. McSherry. Mixed Initiative Dialogue in Case-Based Reasoning. In D. Aha, editor, *Proceedings of the Mixed-Initiative Workshop on CBR at the European Conference on Case-Based Reasoning (ECCBR-02).*, pages 1–8. The Robert Gordon University, 2002. Aberdeen, Scotland.

[McSherry, 2002b] D. McSherry. Recommendation Engineering. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI-02).*, pages 86–90, 2002. Lyon, France.

[Shimazu, 2001] H. Shimazu. ExpertClerk : Navigating Shoppers' Buying Process with the Combination of Asking and Proposing. In Bernhard Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, pages 1443–1448. Morgan Kaufmann, 2001. Seattle, Washington, USA.

[Smyth and McGinty, 2003] B. Smyth and L. McGinty. The Power of Suggestion. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-03)*. Morgan-Kaufmann, 2003. Acapulco, Mexico.