A Multithreaded Conversational Interface for Pedestrian Navigation and Question Answering

Srinivasan Janarthanam[1], Oliver Lemon[1], Xingkun Liu[1], Phil Bartie[2],
William Mackaness[2], Tiphaine Dalmas[3]
[1]Interaction Lab, Heriot-Watt University, Edinburgh
[2] School of GeoSciences, University of Edinburgh
[3]School of Informatics, University of Edinburgh
`sc445,o.lemon,x.liu@hw.ac.uk, philbartie@gmail.com,`
`william.mackaness@ed.ac.uk, tiphaine.dalmas@aethys.com`

Abstract

We demonstrate a conversational interface that assists pedestrian users in navigating within urban environments and acquiring tourist information by combining spoken dialogue system, question-answering (QA), and geographic information system (GIS) technologies. In contrast to existing mobile applications which treat these problems independently, our Android agent addresses the problem of navigation and touristic question-answering in an integrated fashion using a shared dialogue context with multiple interleaved dialogue threads. In this paper, we present the architecture and features of our latest system, extended from an earlier version which was built and evaluated with real users (Janarthanam et al., 2013). The new features include navigation based on visible landmarks, navigation adapted to the user's previous route knowledge, and tourist information pushing based on visible and proximal points-of-interest. The system also uses social media to infer "popularity" of geographical entities.

1 Introduction

We demonstrate a conversational interface that addresses the problems of pedestrian navigation and Question Answering (QA) in urban environments, which is an extended version of the system evaluated in (Janarthanam et al., 2013). Our system brings together Spoken Dialogue Systems (SDS) and Question Answering (QA) with Geographic Information Systems (GIS) for user positioning and tracking, city modelling and urban visibility modelling. There has been little prior work that addresses these problems - navigation, visibility, and tourist information provision - in an integrated

way. By *navigation*, we refer to the problem of finding appropriate destinations to go to and the task of wayfinding to reach them (Golledge, 1992; Raubal and Winter, 2002) and by *tourist information provision* we refer to the problem of meeting the informational needs of a user about entities such as museums, statues, famous personalities, etc. With the proliferation of smartphones, there has been a number of mobile apps developed to address these problems. However, these apps have the following problems: first, they demand the user's visual attention because they predominantly present information on a mobile screen. This can be dangerous in urban environments, as well as being distracting. Second, these apps address the problems of navigation and tourist information independently and do not employ a shared interaction context. This means that users cannot switch between tourist information and navigation tasks in a natural and fluid manner.

User1: Take me to the National Museum.
System2: The National Museum is about 300 meters away ...
System3: At the KFC, turn left on to South Bridge ...
System4 : Near you is the statue of David Hume.
User2: Who is David Hume.
System5: David Hume was a Scottish philosopher....
User3: Tell me more about David Hume.
System6: He was one of the most important figures in.. ...
System7: You should be able to see the museum in front of you.
User4: Tell me more about the museum.
System8: The National Museum of Scotland is a....

Table 1: An example interaction with SpaceBook

In contrast to existing mobile apps, we present a system that has a speech-only interface and addresses navigation and QA in an integrated way.

Table 1 presents an example interaction with our system showing the integrated use of navigation and question-answering capabilities. Ut-

terance *System4-8* shows the system's capability to push information from the QA system about nearby points-of-interest (PoI) during a navigation task, which the user can then follow-up on (in utterances *User2* and *User3*). The final 3 utterances show a natural switch between navigation to an entity and QA about that entity.

2 Related work

Mobile applications such as Siri, Google Maps Navigation, Sygic, etc. address the problem of navigation while applications like Triposo, Guidepal, Wikihood, etc. address the problem of tourist information by presenting the user with descriptive information about various points of interest (PoI) in the city. While some exploratory applications present snippets of information about a precompiled list of PoI, others applications dynamically generate a list of PoI arranged based on their proximity to the users. Users can also obtain specific information about PoI using Search applications. Also, since these navigation and exploratory/search applications do not address both problems in an integrated way, users need to switch between them and therefore lose interaction context.

While most applications address these two problems independently, some like Google Now, Google Field Trip, etc, mix navigation with exploration. However, such applications present information primarily visually on the screen for the user to read. In contrast, our system has the objective of keeping the user's cognitive load low and preventing users from being distracted (perhaps dangerously so) from walking in the city (Kray et al., 2003). Also, our system allows users to interleave the two sub-tasks seamlessly and can keep entities discussed in both tasks in shared context (as shown in Table 1).

Several systems have addressed the issue of pedestrian navigation (Malaka and Zipf, 2000; Dale et al., 2003; Heinroth and Buhler, 2008). Some dialogue systems deal with presenting information concerning points of interest (Ko et al., 2005; Misu and Kawahara, 2007; Kashioka et al., 2011). In contrast to all these earlier work, we demonstrate a system that deals with both navigation and tourist information issues in an integrated fashion.
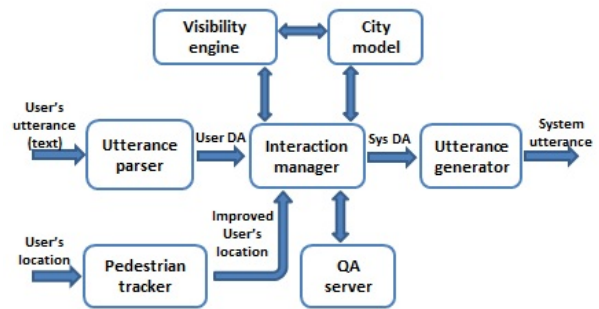

Figure 1: System Architecture

3 Multithreaded dialogue management

The architecture of the current system is shown in figure 1. The Interaction Manager (IM) is the central component of this architecture, which provides the user with navigational instructions, pushes PoI information and manages QA questions. It receives the user's input in the form of a dialogue act (DA) from the ASR module and the user's location (latitude and longitude), orientation and speed from the Pedestrian Tracker module. Based on these inputs and the dialogue context, the IM responds with a system output dialogue act. The Interaction Manager manages the conversation using five coversational threads: dialogue control, response, navigation, question answering, and PoI pushing. These different threads represent the state of different dimensions of the user-system conversation that interleave with each other. Each of these threads generates a dialogue action based on a dialogue policy. A dialogue policy is a mapping between dialogue states and dialogue actions, which are semantic representations of what the system wants to say next. Dialogue actions from the five threads are stored in five separate queues.

The queues are assigned priorities that decide the order in which items from the queues will be popped. For instance, informing the user of a PoI could be delayed if the user needs to be given an instruction to turn at the junction he is approaching. For this reason, priority is assigned to dialogue threads as follows.

Priority 1. Dialogue control (calibration phase, repeat request, clarifications etc)
Priority 2. Responding to user requests
Priority 3. System initiated navigation task actions
Priority 4. Responses to User initiated QA actions
Priority 5. PoI Push actions

Dialogue control The IM initiates the conversation with a calibration phase where the user's initial location and orientation are obtained. In this phase, the IM requests the user to walk a few yards so that the pedestrian tracker can sense the user's location and orientation. During the course of the coversation, the IM uses this thread to manage repeat requests, issues with unparsed user utterances, utterances that have low ASR confidence, and so on. The dialogue control thread is used to manage reference resolution in cases where referring expressions are underspecified.

Navigation The IM identifies the location of the destination entity and queries the City Model for a route plan. The plan provides information such as numbers of exits at junctions, the exit number the user should take, turn angle, popularity index of the street, and the slope of the road. In an attempt to adapt the route instructions to user route knowledge, the IM first picks the most popular street in the plan and asks the users if they can get to the street on their own. Also, the IM queries the Visibility Engine (VE) for highly salient visible landmarks (computed using Flickr tags) that can used to direct the user. Instructions based on visible landmarks are given whenever possible.

Question Answering The system also answers ad hoc questions from the user (e.g. "Who is David Hume?", "What is the Old College?", etc). These are sent to the QA server and answered based on responses from the QA server. The dialogue policy here is to answer the user's question with the first snippet available and ask the user to request for more if interested.

Pushing PoI Information When the user is mobile, the IM identifies points of interest on the route based on two factors: proximity and visibility. Proximity push is done by checking for PoIs near the user using high-scoring ones when there are many, based on tourist popularity ratings in the City Model. Visibility push is done by querying the VE for salient entities visible to the user that may be worth pushing. The dialogue policy is to introduce the PoI entity along with visual descriptors if available. The IM queries the QA server for snippets on entity and if available, pushes them the first snippet to the user. The user is encouraged to ask for more if interested.

4   Conclusion

We demonstrate a mobile conversational system to support pedestrian users in navigation and question-answering tasks in urban environments. The system is a speech-only interface and interleaves navigation and tourist information in an integrated way, using a shared dialogue context. For example, using the navigational context, our system can push point-of-interest information which can then initiate touristic exploration tasks using the QA module. An evaluation of an earlier version was reported in (Janarthanam et al., 2013).

References

R. Dale, S. Geldof, and J. Prost. 2003. CORAL : Using Natural Language Generation for Navigational Assistance. In *Proceedings of ACSC2003, South Australia.*

R. G. Golledge. 1992. Place recognition and wayfinding: Making sense of space. *Geoforum*, 23.

T. Heinroth and D. Buhler. 2008. Arrigator: evaluation of a speech-based pedestrian navigation system. In *Proceedings of 4th International Conference on Intelligent Environments, 2008.*

S. Janarthanam, O. Lemon, P. Bartie, T. Dalmas, A. Dickinson, X. Liu, W. Mackaness, and B. Webber. 2013. Evaluating a city exploration dialogue system combining question-answering and pedestrian navigation. In *Proc. ACL 2013.*

H. Kashioka, T. Misu, E. Mizukami, Y. Shiga, K. Kayama, C. Hori, and H. Kawai. 2011. Multimodal Dialog System for Kyoto Sightseeing Guide. In *Asia-Pacific Signal and Information Processing Association Conference.*

J. Ko, F. Murase, T. Mitamura, E. Nyberg, M. Tateishi, I. Akahori, and N. Hataoka. 2005. CAMMIA: A Context-Aware Spoken Dialog System for Mobile Environments. In *IEEE Automatic Speech Recognition and Understanding Workshop.*

C. Kray, K. Laakso, C. Elting, and V. Coors. 2003. Presenting route instructions on mobile devices. In *Proceedings of IUI 03, Florida.*

R. Malaka and A. Zipf. 2000. Deep Map - challenging IT research in the framework of a tourist information system. In *Information and Communication Technologies in Tourism 2000*, pages 15–27. Springer.

T. Misu and T. Kawahara. 2007. An Interactive Framework for Document Retrieval and Presentation with Question-Answering Function in Restricted Domain. In *Proc. of the 26th IEA/AIE conference*, pages 126–134.

M. Raubal and S. Winter. 2002. Enriching wayfinding instructions with local landmarks. In *Second International Conference GIScience. Springer, Boulder, USA.*