Towards Personal Assistants that Can Help Users Plan

Peng Yu[1], Jiaying Shen[2], Peter Z. Yeh[2], and Brian Williams[1]

[1] CSAIL, MIT, Cambridge MA 02139, USA,
{yupeng,williams}@mit.edu
[2] NL & AI Lab, Nuance Communications Inc., Sunnyvale CA 95054, USA
{Jiaying.Shen,Peter.Yeh}@nuance.com

Abstract. In this paper, we present an intelligent personal assistant, called Uhura, that handles requests involving multiple, interrelated goals and activities by efficiently producing a coherent plan. Uhura achieves this by integrating a collaborative dialog manager, a conflict-directed planner with spatial and temporal reasoning capabilities, and a large-scale knowledge graph. We also present a user survey that assesses the usefulness of the plans produced by Uhura in urban travel planning.

1 Introduction

People are now using intelligent personal assistants for many day-to-day tasks, and many of which assistants support both verbal and visual communications to make the interaction simpler. However, their functions are limited to simple commands and information retrieval tasks, and none of them can understand user requests involving multiple goals and activities. For example, [8] reports an end-to-end personal assistant framework, but their work is focused on proactivity and task management. [7] report end-to-end personal assistants for TV program discovery, but these assistants are unable to plan.

In contrast, our goal is to develop a personal assistant, called Uhura, that supports complex user requests that require planning. Uhura can also consider semantic constraints, which occur frequently in many day to day scenarios. It achieves these benefits by uniquely utilizing three distinct components: Knowledge Base, Planner, and Collaborative Dialog Manager (CDM). First, Uhura uses a knowledge base that provides the data necessary (e.g., road condition, restaurants, movies, etc.) to formulate candidate plans that meet the users' semantic, temporal and spatial constraints. Second, Uhura integrates a dialog manager and a knowledge base with a planner that supports temporal and spatial reasoning, similar to the ones used in tourist assistant systems [6]. This allows Uhura to support semantic constraints as well as temporal and spatial ones. In addition, if there are competing requirements, the Planner will propose relaxations for them, and negotiate with the user until a resolution is reached. Finally, Uhura uses a dialog manager to: 1) serve as a mediator between the planner and knowledge base; 2) maintain the context of the interaction with the users along with their goals; 3) communicate with the users in a natural way.

2 Approach

Uhura provides the following features to simplify the planning process for the users: 1) natural language communication; 2) mixed initiative goal-directed interaction; 3) supports for multiple tasks and constraints; and 4) being robust to temporal uncertainty and over-subscription. They are supported by a coordinated system of three major components (Figure 1): Collaborative Dialog Manager (CDM), Planner and Knowledge Base. CDM handles the interactions with users, and elicits their goals and requirements as a *Qualitative State Plan* (QSP). It also takes the users' semantic, temporal and spatial constraints, and formulates them into queries for the Knowledge Base. The results of these queries ground the tasks in the QSP with additional episodes and constraints. Finally, the expanded QSP is sent to the Planner, which evaluates the alternatives of each task and produces plans that best meet the users' requirements.

The three components have different responsibilities, and *speak* very different languages. The key to an effective integration is to **disintegrate** the overall problems properly, assign the subproblem to the component that has the right reasoning capability, and supply them with the right set of data. Inside Uhura, CDM is responsible for interacting with the users and capturing their planning problems. It creates and assigns subproblems that require temporal and spatial reasoning to the Planner, and subproblems that require semantic reasoning to the Knowledge Base. The output from CDM is a set of first order logic expressions encoded in a semantic graph, while the input to the planner is a set of goals and requirements encoded as a Qualitative State Plan. A mapping is created between them such that (1) each task's temporal and spatial requirements can be extracted from the semantic graph and encoded in the QSP; and (2) the choices and relaxations in the Planner's output can be mapped back to the nodes in the semantic graph, such that CDM can present them to the user. For example, Figure 2 shows such a mapping for a movie task. The mapping between the nodes in the branch and the QSP constraint is represented by the node IDs tagged to the episodes' names, locations and durations.

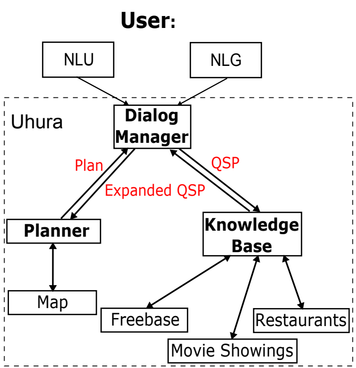

Fig. 1. The architecture graph of Uhura

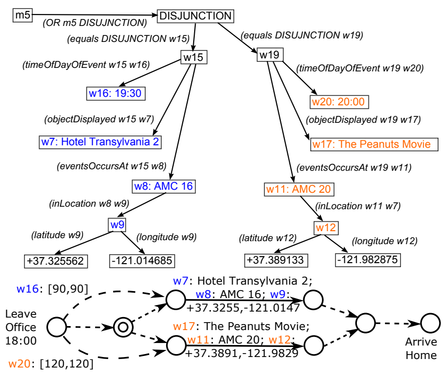

Fig. 2. The QSP generated from a semantic graph branch

Collaborative Dialog Manager CDM [3] is an extension to Disco [4], an open source dialog development framework based on Collaborative Discourse Theory [1], and Sidner's artificial language for negotiation [5]. It views a personal assistant dialogue as a process of plan augmentation, where the purpose of the dialogue is for the system and the user to collaboratively form a complete sharedPlan in order to meet the user's intention. The plan based approach of CDM puts it as the processing hub of the system, whose assembled meta task plan often includes knowledge base querying and Uhura planner invocation. On the other hand, it is also the information hub of the system, since CDM is responsible for assembling all user requests to form legitimate knowledge base queries, integrating knowledge base query results into a valid planning problem as well as interacting with the user to communicate all the information generated. In order to effectively carry out all these responsibilities, CDM uses first order logic (FOL) encoded as semantic graphs, such as the one in Figure 2, to store and process the information from various sources.

Knowledge Base Uhura uses a large-scale knowledge base to access the world knowledge, such as restaurants and movie showtimes, to properly construct a plan that satisfies the user's request. This knowledge base is constructed from a combination of open and proprietary sources of content using an ingestion pipeline [2] that transforms the raw content into RDF triples and performs entity resolution. The resulting knowledge base can be viewed as a very large knowledge graph, where the nodes represent entities and the edges represent semantic relations between these entities. The entities are typed, and a proprietary subsumption hierarchy is used to organize these types. The semantic relations have domain and range constraints, and also capture inverse relationships. Our knowledge base has over 1 billion triples, and the processing time for each query is typically less than a few hundred milliseconds.

Planner The planner fills in the details of the abstract plan generated by CDM and Knowledge Base, schedules each activity, and adds contingencies for likely delays during transit. Uhura's planner is implemented based on a constraint relaxation algorithm, Conflict-Directed Relaxation with Uncertainty (CDRU, [9]), with extensions for propagating global constraints. CDRU was first developed to solve over-constrained conditional temporal problems with uncertain durations. The algorithm uses a conflict-directed search strategy to prune infeasible candidates and find the optimal set of choices. In addition, CDRU is able to detect competing constraints in the QSPs, generate concise explanations for the cause of failure, and suggest trade-offs for the users to resolve the issues. It takes in a QSP as input, and produces a plan, as well as temporal relaxations for some constraints if necessary.

User Survey We conducted a user survey on the usefulness of Uhura in the context of a personal assistant for managing day-to-day tasks. There are six sessions in this survey, each presenting a different urban travel planning scenario. At the end of each session, the participants are asked to evaluate the last plan proposed by Uhura, and grade it on quality using a 5 point Likert scale for their satisfaction with it. The resulting QSPs have around 500 episodes with a dozen choices, and is usually solved within a few seconds. From the users' perspective, the delay in response is not longer than many popular routing applications.

Table 1. Average scores and next solution requests in each session (standard deviation)

Session	1	2	3	4	5	6
Quality	3.2 (1.40)	2.4 (1.43)	2.9 (1.58)	3.8 (1.54)	3.3 (1.35)	3.3 (1.42)
# of Solutions	4.8 (6.14)	4.8 (4.60)	4.5 (5.30)	2.9 (2.91)	2.9 (1.97)	3.0 (5.37)

We received results from ten participants, which are summarized in Table 1. In general, participants found Uhura to be useful in planning daily tasks. The plans generated are acceptable in most situations, and the average quality score is above 3. Participants also gave feedback that Uhura simplified the otherwise complicated planning tasks. Without Uhura, planning a day trip may take minutes or even hours. With Uhura, a feasible solution can be found in seconds. On the other hand, we also discovered a few issues related to Uhura's architecture and implementation in the survey. Some participants reported that Uhura is not making good trade-offs between temporal relaxations and destinations, and it occasionally produced nonsensical plans that can be avoided by applying common sense reasoning. We believe that these are the reasons for the large variance in the quality scores, and are important issues for us to address in the future.

References

1. Grosz, B.J., Sidner, C.L.: Plans for discourse. In: Cohen, P.R., Morgan, J., Pollack, M.E. (eds.) Intentions in Communication, pp. 417–444. MIT Press, Cambridge, MA (1990)
2. Noessner, J., Martin, D., Yeh, P., Patel-Schneider, P.: Cogmap: A cognitive support approach to property and instance alignment. In: ISWC (2015)
3. Ortiz, C., Shen, J.: Dynamic intention structures for dialogue processing. In: Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2014) (2014)
4. Rich, C., Sidner, C.L.: Using collaborative discourse theory to partially automate dialogue tree authoring. In: Intelligent Virtual Agents. pp. 327–340. Springer (2012)
5. Sidner, C.: An artificial discourse language for collaborative negotiation. In: In Proceedings of the Twelfth National Conference on Artificial Intelligence. pp. 814–819. MIT Press (1994)
6. Vansteenwegen, P., Souffriau, W., Berghe, G.V., Oudheusden, D.V.: The city trip planner: An expert system for tourists. Expert Systems with Applications 38, 6540–6546 (2011)
7. Yeh, P., Ramachandran, D., Douglas, B., Ratnaparkhi, A., Jarrold, W., Provine, R., Patel-Schneider, P., Laverty, S., Tikku, N., Brown, S., Mendel, J., Emfield, A.: An end-to-end conversational second screen application for tv program discovery. AI Magazine 36(3) (2015)
8. Yorke-Smith, N., Saadati, S., Myers, K., Morley, D.: The design of a proactive personal agent for task management. IJAIT 21(1) (2012)
9. Yu, P., Fang, C., Williams, B.: Resolving uncontrollable conditional temporal problems using continuous relaxations. In: Proceedings of the Twenty-fourth International Conference on Automated Planning and Scheduling (ICAPS-14) (2014)