

Mining human interactions to construct a virtual guide for a virtual fair

Andrés Luna

LIIS Group, FaMAF

Universidad Nacional de Córdoba

Córdoba, Argentina

andres.ignacio.luna@gmail.com

Luciana Benotti

LIIS Group, FaMAF

Universidad Nacional de Córdoba

Córdoba, Argentina

luciana.benotti@gmail.com

Abstract

In this paper we describe how we mine interactions between a human guide and a human visitor to build a virtual guide. A virtual guide is an agent capable of fulfilling the role of a human guide. Its goal is to guide visitors to each booth of a virtual fair and to provide information about the company or organization through interactive objects located at the fair.

The guide decides what to say, using a graph search algorithm, and decides how to say using generation by selection based on contextual features. The guide decides where to speak at the virtual fair by creating clusters using a data classification algorithm to learn in what positions the human guide decided to talk.

1 Introduction and previous work

Fairs are spaces where companies that offer similar products and services meet to promote them. A virtual fair emulates a real fair and can be available before the real fair happens in order to promote it to its potential visitors.

The virtual fair used in this work is a tourism fair that took place in Mexico, where visitors could find in each company's booth interactive video and links to tourist companies' websites promoting particular products. The goal of the virtual guide is to walk the user through the virtual fair, providing information about the companies' booths and inviting them to click on interactive objects to obtain more information.

In (Jan et al., 2009) the authors describe a virtual guide used to promote an island in the online game *Second Life* whose goal was to provide information to US army veterans. Our approach differs to that of (Jan et al., 2009) in that the virtual guide learns where to speak and how to realize

its contributions from an automatically annotated corpus, rather than by using manually designed rules. However, our guide is not able to interpret utterances from the visitor, its decisions are only based on the visitor behavior. Natural language generation is achieved by adapting the *generation by selection* method described in (Benotti and Denis, 2011a; Benotti and Denis, 2011b).

The generation by selection method affords the use of complex and human-like sentences, and it does not need rule writing by a dialogue expert or manual annotations, among other of their many advantages. The disadvantage of corpus based generation is that the resulting dialogue may not be fully coherent. Shawar and Atwell (2003; 2005) present a method for learning pattern matching rules from corpora in order to obtain the dialogue manager for a chatbot. Gandhe and Traum (2007a; 2007b) investigate several dialogue models for negotiating virtual agents that are trained on an unannotated human-human corpus. Both approaches report that the dialogues obtained by these methods are still to be improved because the lack of dialogue history management results in incoherence. Since in task-based systems, the dialogue history is restricted by the structure of the task, the absence of dialogue history management is alleviated by tracking the current state of the task.

In Section 2 we introduce the corpus used by this work. We discuss the clustering method used on the corpus in Section 3; the clustering is used to decide where to speak. After that, we describe in Section 4 the mechanisms for instruction generation and graph search used to guide the visitors. Later, in Section 5 we show the results obtained in the evaluation process and compare our system's performance with other virtual instructors. Finally, in Section 6 we elaborate a conclusion about the virtual guide performance and capabilities, as well as discuss the possible improvements.

2 Virtual guide human-human corpus

We collected a corpus using a human guide in a wizard of Oz setup (Kelley, 1983). The corpus is comprised by 5 correct sessions in total performed by the same virtual tour guide, and according to the desired behavior and actions as specified for both participants. We recorded 2 hours and 2 minutes of virtual fair guided visits which produced a total of 136 utterances, having employed 18.02 words and 89.29 characters in average per utterance. 9 different interactive objects were clicked located in 4 different booths in average per session. In Figure 1 we show an aerial view of the virtual fair and the occurrence of utterances, marked in blue.

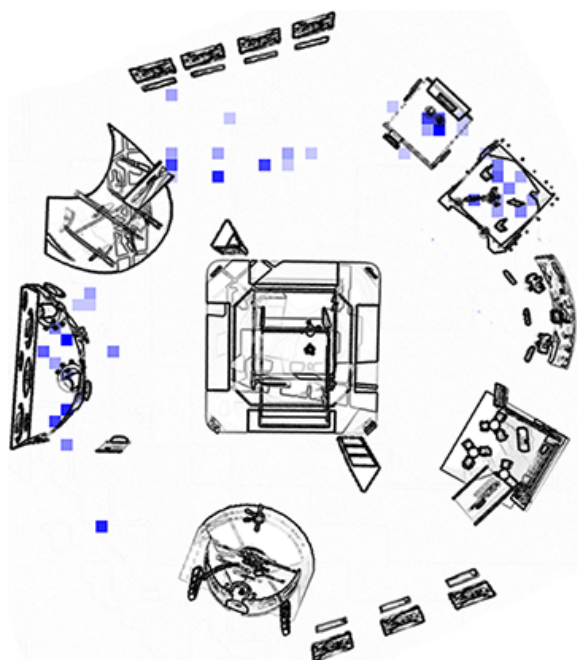


Figure 1: Map of registered utterances in corpus. A higher color intensity denotes a higher utterance density in the area.

3 Behavior-based utterance clustering

The generation by selection method that we use in this work is based on contextual features, in particular it is based on the position of the visitor inside the virtual fair and the actions that are affordable from that region in the fair. Deciding whether two positions in the fair have the same affordances, or, as we call it, fall into the same region is critical to select appropriate utterances from the corpus depending on the guide’s location and task progress.

The discretization employed in (Benotti and Denis, 2011a) was geometrical discretization, di-

viding the world in regions based on the area visible to the guide. Instead of doing a geometrical discretization our virtual fair discretization was behavior-oriented which means that regions are delimited by clustering utterances that were uttered in a close position from each other. In the corpus utterances tend to cluster around *decision points*, locations there is more than one affordable and salient action available to the user and when the help and direction of the guide is required.

Geometrical region identification based on visibility normally requires a larger corpus in order to get a correct utterance generation, because the chance of having a region without any utterance occurrence inside is higher. In such discretization, different regions may contain a very different number of utterances while using behavior-oriented discretization results in regions with a similar number of utterances each. That is why the behavior-oriented discretization is an advantage for our virtual guide, since our corpus is considerably smaller to that used in (Benotti and Denis, 2011a).

We ran a modified version of the *k-means clustering* algorithm (Pakhira, 2009) that avoids empty clusters over our corpus to group instructions. As paraphrase instructions, while performing a task, occur in a same decision point, then we wanted close instructions to be in the same cluster, and therefore our criteria of “similarity” between them was euclidean distance. Ideally, different decision points should be in different clusters to guarantee selected utterances are appropriate in every situation.

Let us visualize virtual fair as a directed graph (V, E) where $V = regions$, and if $a, b \in V$ then $(a, b) \in E$ if and only if there is at least one utterance in the corpus whose immediate reaction was moving from region a to the region b . If we choose a low number of clusters the k-means clustering algorithm would cluster instructions of different nature, and conversely a too high value would make the virtual fair disconnected. Then, to obtain an optimal clustering -and therefore an optimal discretization- we maximize the k parameter such that the virtual fair’s graph is still connected.

Discretization is finally obtained by matching every position (x, y) in the environment to the nearest cluster’s *centroid*. We show in Figure 2 the virtual fair discretized in $k = 22$ regions, as that number was the maximum number of clusters we

could reach without breaking the graph connectivity. Regions are delimited by lines and centroids are represented by white squares.

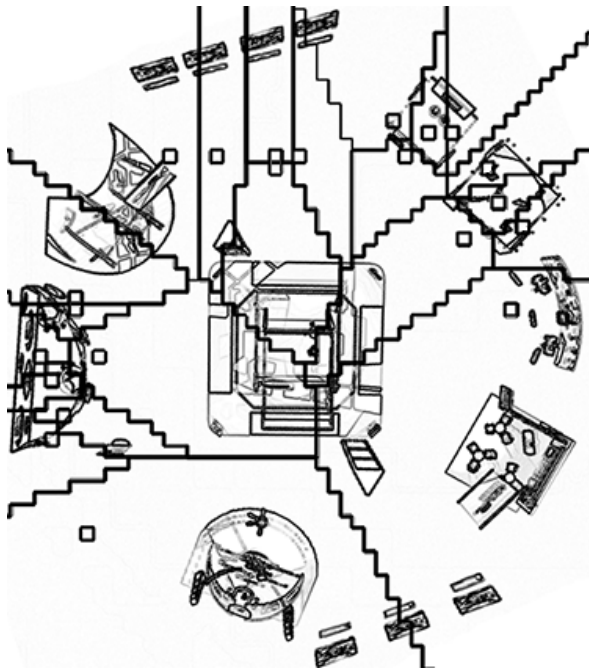


Figure 2: Virtual fair divided in $k = 22$ regions

4 The virtual guide

The virtual guide must direct visitors through the fair to interactive objects in order to complete its promotion duty in each visit session. We show in Figure 3 a situation in which a visitor is near an interactive object and the virtual guide encourages him/her to click it generating an utterance whose translation is “If you click on the green cube you will access Lawson’s website where you can learn more about them and the communication services they offer”.

We can see the use of a referring expression, a negative politeness strategy (Brown and Levinson, 1987) to suggest an action but not impose it while some information about the Lawson firm is given.

In subsection 4.1 we discuss about the corpus automatic annotation. Then we describe how utterances are selected in subsection 4.2.

4.1 Corpus annotation

Our annotation process was simpler and more straightforward than (Benotti and Denis, 2011a), where artificial intelligence planning is used to normalize reactions, mainly due to the fact that users can not change the virtual fair state during their visit, they can only change their own posi-



Figure 3: The virtual guide took the visitor to an interactive object and encourages him/her to manipulate it

tion and visibility area (defined by the orientation in the virtual fair) and manipulate interactive objects.

In a virtual fair visit, the set of user’s relevant actions are:

- Move from one region to another
- Change orientation to left or right
- Click on an interactive object

Consequently, the set of atoms representing a virtual fair’s state was simplified to

- $user-region(region)$
- $user-orientation(x,y,z,w)$ ¹
- $clicked(anInteractiveObject)$

In short, to do automatic annotation on the virtual guide’s corpus, it was sufficient to observe the subsequent action to each utterance by looking for a change on any of the atoms shown above, and annotating and associating the corresponding reaction to the utterance and the valid atoms set when it was said.

4.2 Selecting what to say

The virtual guide’s goal is to make the visitor visit a number of given objectives, namely a set of stands and interactive objects. Using the virtual fairs discretization and taking the directed graph representation we presented in Section 3, the virtual guide uses the A* algorithm to obtain a path, that is a sequence of actions, from its current position to the region where the next objective is located. In case the visitor got lost or simply took an alternative path, the virtual guide recalculates the shortest path and proceeds to guide the visitor through it.

¹In quaternion representation

Clearly, in order to do this calculation it is critical that every objective is reachable from any node in the graph, so choosing a k parameter in the discretization process must be done taking care of that.

The virtual guide gives the visitor a new instruction depending on next actions to perform using the selection algorithm taken from (Benotti and Denis, 2011a), shown in Algorithm 1. The algorithm obtains set of utterances C , all of which have a reaction that corresponds to the sequence of actions that the virtual guide wants the visitor to perform next.

Algorithm 1 Virtual guide’s selection algorithm

```

 $C \leftarrow \emptyset$ 
 $action \leftarrow nextAction(currentObjective)$ 
for all  $Utterance U \in Corpus$  do
  if  $action \approx U.Reaction$  then
     $C \leftarrow C \cup U$ 
  end if
end for

```

5 Evaluation results

In the evaluation process 11 evaluators participated, completing the proposed visit to the virtual fair, each manipulating 9 interactive objects. Evaluators were also asked to complete a questionnaire after the tour, in which we wanted to obtain several subjective metrics. We were particularly interested in the questions

- **S1:** *I had difficulties identifying the objects that the system described for me*
- **S2:** *The Utterances sounded robotic*
- **S3:** *The system was repetitive*

where we previously supposed the virtual guide would have better results than other virtual instructors, if we consider the results showed in (Benotti and Denis, 2011a).

We compared our virtual guide results with the two best symbolic systems built for another virtual environment, the GIVE-2 Challenge. Those systems were NA from INRIA and SAAR from University of Saarland (see (Koller et al., 2010)). Furthermore, we checked if the virtual guide results were similar to another virtual instructor, also built for GIVE-2, called OUR, in which generation

by selection was applied to make natural language generation possible.

In Table 1 we show the results for each virtual instructor in the three categories we are interested. We can see that the virtual guide obtained significantly better results than the SAAR and NA and in questions **S1**, **S2** and **S3**, as we had supposed. All three questions range from 1 (one) to 9 (nine), the lower the number the better the system (since questions are negative).

Table 1: Results comparison between virtual guide and three GIVE-2 systems

Question	NA	SAAR	OUR	VP
S1	4.1	4	3	1.81
S2	5.2	4.75	3.6	1.82
S3	6.55	6.3	5.4	2

6 Conclusions and future work

In this paper we described the construction of a virtual guide for a virtual fair with the purpose of guiding visitors through the stands and to interactive objects located inside the fair. Immersive virtual fairs and expositions constitute a promising way to promote such events.

On our evaluation, the virtual guide had comparable results than the virtual instructor GIVE-2 implemented using generation by selection, using a much smaller corpus. Our guide got better results than the two best performing symbolic systems. These results are preliminary, but also encouraging.

A possible extension of this work could be that virtual guide can continue to improve its behavior by learning online when input from a human guide of the fair is available. If more corpus is available in this way the virtual guide could discard those utterances that do not lead most visitors to perform the intended reaction.

As a result of this work we conclude that virtual guide met the basic functions of navigation and natural language generation that we expected and that the resulting prototype is ready to be deployed at the virtualization of events website <http://www.inixiavf.com/>.

References

- Luciana Benotti and Alexandre Denis. 2011a. Giving instructions in virtual environments by corpus based selection. In *Proceedings of the SIGDIAL 2011*

- Conference, SIGDIAL '11, pages 68–77, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Luciana Benotti and Alexandre Denis. 2011b. Prototyping virtual instructors from human-human corpora. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 62–67, Portland, Oregon, June. Association for Computational Linguistics.
- Penelope Brown and Stephen Levinson. 1987. *Politeness: Some Universals in Language Usage*. Studies in Interactional Sociolinguistics. Cambridge University Press.
- Sudeep Gandhe and David Traum. 2007a. Creating spoken dialogue characters from corpora without annotations. In *Proceedings of 8th Conference in the Annual Series of Interspeech Events*, pages 2201–2204, Belgium.
- Sudeep Gandhe and David Traum. 2007b. First steps toward dialogue modelling from an un-annotated human-human corpus. In *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Hyderabad, India.
- Dusan Jan, Antonio Roque, Anton Leuski, Jacki Morie, and David Traum. 2009. A virtual tour guide for virtual worlds. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents, IVA '09*, pages 372–378, Berlin, Heidelberg. Springer-Verlag.
- John F. Kelley. 1983. An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '83*, pages 193–196, New York, NY, USA. ACM.
- Alexander Koller, Kristina Striegnitz, Andrew Gargett, Donna Byron, Justine Cassell, Robert Dale, Johanna Moore, and Jon Oberlander. 2010. Report on the second nlg challenge on generating instructions in virtual environments (give-2). In *Proceedings of the 6th International Natural Language Generation Conference, INLG '10*, pages 243–250, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Malay K. Pakhira. 2009. A modified k-means algorithm to avoid empty clusters. *International Journal of Recent Trends in Engineering*, 1(1):220–226.
- Bayan Abu Shawar and Eric Atwell. 2003. Using dialogue corpora to retrain a chatbot system. In *Proceedings of the Corpus Linguistics Conference*, pages 681–690, United Kingdom.
- Bayan Abu Shawar and Eric Atwell. 2005. Using corpora in machine-learning chatbot systems. volume 10, pages 489–516.