Mr. Clue — A Virtual Agent that Can Play Word-Guessing Games

Eli Pincus and **David DeVault** and **David Traum**
Institute for Creative Technologies, University of Southern California,
12015 Waterfront Drive, Playa Vista, CA 90094, USA

Abstract

This demonstration showcases a virtual agent, Mr. Clue, capable of acting in the role of clue-giver in a word-guessing game. The agent has the ability to automatically generate clues and update its dialogue policy dynamically based on user input.

Introduction

In this demonstration we showcase a virtual agent named Mr. Clue, who is capable of acting in the role of clue-giver in a word-guessing game. There are different forms of word-guessing games, but they have in common that a clue-giver must induce a receiver to guess a target word. The clue-giver is not allowed to mention the target (some word-guessing games have other restrictions as well). Mr. Clue is meant to emulate a talented human clue-giver – steering the receiver to a correct guess as quickly as possible. Talented clue-givers exhibit rapid interactive dialogue skills that are challenging for state of the art dialogue systems. Examples of these skills include the ability to react in real time to the other interlocutor's guesses and the ability to efficiently reformulate misunderstood clues. Mr. Clue is a research testbed for dialogue system research aimed at simulating these skills. To this end, Mr. Clue's dialogue manager's policy reflects patterns of game-play discussed in (Paetzel, Racca, and DeVault 2014), which describes the timed word-guessing game, RDG-Phrase, in the Rapid Dialog Game corpus. The targets in the RDG-Phrase game are common english nouns such as *ambulance* and *electric*.

Mr. Clue was built on top of the ICT Virtual Human Toolkit, which provides APIs and components for creating virtual humans (Hartholt et al. 2013). We created two new components to be able to play the RDG-Phrase game: a **clue generator** that queries databases and scrapes web content, and a **dialogue manager** that governs game flow. Each of these is discussed in the sections below.

The demonstration will consist of live game interactions with Mr. Clue and a human receiver. Each game instance will display Mr. Clue's ability to generate clues and direct game flow.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Clue Generator

Mr. Clue generates his clues for a given target word automatically by querying the WordNet database (Miller 1995) and scraping text from the wikipedia page and dictionary.com page that correspond to the target. The raw text from the querying and scraping is then processed in a basic filtering step to improve the quality of the final clue given. The use of the target by the giver is not allowed in the RDG-Phrase game; therefore the filtering step involves replacing instances of the target and/or any of its inflected forms with the word "*blank*".

The filtering step also includes splitting the text (returned from a WordNet query and scraping the word's dictionary.com page) into multiple clues using simple punctuation-based rules.

Mr. Clue automatically generates several different types of clues that match the types utilized by human clue-givers of the RDG-Phrase game described in (Pincus and Traum 2014). Most wikipedia derived clues result in a **DescriptionDef** clue which defines the target. On the other hand, WordNet and dictionary.com derived clues result in several different types including: **DescriptionDef**, **PartialPhrase**, **AssocAction**, and **Synonym** clue types. **PartialPhrase** clues are clues composed of words that are commonly used with the target, **AssocAction** clues describe what the target does, what it is used for, or what uses it, and **Synonym** clues are synonyms of the target.

Table 1 gives examples of automatically generated clues, their type, and their source. Currently, Mr. Clue uses a list of 2,334 common english nouns as possible targets, taken from a freely available noun-list (Quintans 2013). For each target on the noun list, an average of 11 clues are obtained from WordNet, an average of 27 clues are obtained from each word's dictionary.com page, and up to 1 clue is obtained from a word's wikipedia page.

Empirical Analysis of Clues

We have initial results from a Mechanical Turk experiment designed to compare the effectiveness of automatically generated clues and clues output by a human clue-giver at eliciting a correct guess from a human receiver.[1] In this exper-

[1] Human clues chosen were filtered so that the clues did not depend on the context of previous utterances.

Clue	Target	Type	Source
to collect one's *blanks*.	Thought	PartialPhrase	Dictionary.com
a *blank* is a solid block of wax with an embedded wick which is ignited to provide light, and sometimes heat, and historically was used as a method of keeping time	Candle	DescriptionDef	Wikipedia
he always rode the *blank* to work	Bus	AssocAction	WordNet

Table 1: Automatically Generated Clue Type Examples

Target	Human Clue (# of Turkers Providing a Correct Guess)	Auto. Generated Clue (# of Turkers Providing a Correct Guess)
Ambulance	car for an emergency (8/10)	a vehicle that takes people to and from hospitals (9/10)
Video	um you a cinematographer shoots this (0/10)	the visible part of a television transmission (3/10)
Convertible	the roof comes down on a car its called a (2/10)	a car that has top that can be folded or removed (8/10)

Table 2: Automatically Generated Clue Efficacy vs. Human Clue Efficacy Examples

iment Turkers were presented with a range of clues for different targets in written form and asked to submit a written list of guesses of what the target might be. The experiment required 10 different Turkers to provide guesses for a given clue. 275 automatically generated clues and 20 human clues were given to Turkers. The results indicate that some of the automatically generated clues can be more effective at eliciting a correct guess from a human receiver than clues output by a human clue-giver. Table 2 supports this claim by providing examples of automatically generated clues that elicited more correct guesses for the same targets than clues output by a human clue-giver. 58 out of 275 (21%) of the automatically generated clues elicited a correct guess from at least half of the Turkers (at least 5/10 Turkers) that were presented with the clue. The corresponding statistic for human generated clues is 5 out of 20 clues (25%). On-going work includes continuing to analyze the features of clues that result in the highest percentage of correct guesses.

Dialogue Policy

Using the Virtual human toolkit architecture, Mr. Clue is capable of accepting speech or text input. Mr. Clue classifies user input into 1 of 3 categories: *Incorrect*, *Empty*, *Correct*. *Incorrect* user input does not contain the target, *Empty* user input does not contain any words, and *Correct* user input contains the current target. *Incorrect* user input prompts Mr. Clue to inform the receiver that their prior guess(es) were not correct and provide a new clue or recycle an already used clue if he does not possess unused clues for the current target. Mr. Clue responds to *Empty* user input by informing the user that their input was empty and repeating the last used clue. Finally, if the user input is *Correct* Mr. Clue indicates that the user made a correct guess and provides a clue for a new target. Figure 1 shows Mr. Clue in front of a woman avatar, who plays the game judge, keeping time and monitoring the rules.

Future Work

We will continue our analysis of the results from our Mechanical Turk experiment in order to learn an automatic

Figure 1: Mr. Clue (right) and game judge (left)

method for predicting which generated clues are most effective. We plan to provide Mr. Clue the ability to incrementally process human utterances to facilitate dynamic updates to clue-giving strategy and game-flow. For instance, if a correct guess is given while Mr. Clue is speaking, Mr. Clue should be able to interrupt himself and respond by requesting a synonym (of the correct guess). This type of response would be in keeping with common patterns of human clue-giver game-play discussed in (Pincus and Traum 2014).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1219253. We also want to thank Ed Fast for help with virtual human toolkit integration, and Anton Leuski, for the speech recognizer interface.

References

Hartholt, A.; Traum, D.; Marsella, S. C.; Shapiro, A.; Stratou, G.; Leuski, A.; Morency, L.-P.; and Gratch, J. 2013. All together now. In *Intelligent Virtual Agents*, 368–381. Springer.

Miller, G. A. 1995. Wordnet: A lexical database for english. *Communications of the ACM* 38:39–41.

Paetzel, M.; Racca, D. N.; and DeVault, D. 2014. A multimodal corpus of rapid dialogue games. In *Language Resources and Evaluation Conference (LREC)*.

Pincus, E., and Traum, D. 2014. Towards a multimodal taxonomy of dialogue moves for word-guessing games. In *10th Workshop on Multimodal Corpora (MMC)*.

Quintans, D. 2013. The great noun list. http://www.desiquintans.com/downloads/nounlist.txt. accessed July 1, 2014.