A game-based setup for data collection and task-based evaluation of uncertain information presentation

Dimitra Gkatzia, Amanda Cercas Curry, Verena Rieser and Oliver Lemon
Interaction Lab
Heriot-Watt University
EH14 4AS, Edinburgh
{d.gkatzia,ac293,v.t.rieser,o.lemon}@hw.ac.uk

Abstract

Decision-making is often dependent on uncertain data, e.g. data associated with confidence scores, such as probabilities. A concrete example of such data is weather data. We will demo a game-based setup for exploring the effectiveness of different approaches (graphics vs NLG) to communicating uncertainty in rainfall and temperature predictions (www.macs.hw.ac.uk/InteractionLab/weathergame/). The game incorporates a natural language extension of the MetOffice Weather game[1]. The extended version of the game can be used in three ways: (1) to compare the effectiveness of different information presentations of uncertain data; (2) to collect data for the development of effective data-driven approaches; and (3) to serve as a task-based evaluation setup for Natural Language Generation (NLG).

1 Introduction

NLG technology achieves comparable results to commonly used data visualisation techniques in terms of supporting accurate human decision-making (Gatt et al., 2009). In this paper, we present a task-based setup to explore whether NLG technology can also be used to support decision-making when the underlying data is uncertain. The Intergovernmental Panel on Climate Change (IPCC) (Manning et al., 2004) and the World Meteorological Organisation (WMO) (Kootval, 2008) list the following advantages of communicating risk and uncertainty: information on uncertainty has been shown to improve decision making; helps to manage user expectations; promotes user confidence; and is reflective of the state of science. Results by Stephens (2011) further show that, although people prefer reports us-
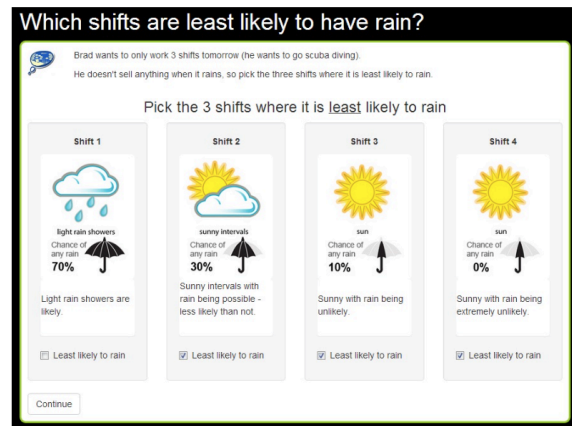

Figure 1: Screenshot of the Extended Weather Game (graphs and text version).

ing percentages (e.g. 10% chance of rain), this does not necessarily equate with *understanding*, i.e. making the right decision based on this information. One possible explanation is low "risk literacy" (Cokely et al., 2012), i.e. a reduced ability to accurately interpret and act on numerical information about risk and uncertainty.

In this research, we aim for a better understanding of how to effectively translate numerical risk and uncertainty measures into "laymen's" terms using natural language, following the recommendations of the WMO (Kootval, 2008). For example, the relative risk of 1 in 1000 could be described as *exceptionally unlikely*. We expect that through the use of language we will improve understanding and thus decision-making for users with low risk literacy (as measured by the Berlin literacy test[2]).

2 The Weather Game

Recruiting users to perform evaluations is a laborious task and many studies suffer from underpowered evaluations. Therefore, we use a crowdsourcing technique known as "game with a purpose", which has been shown to assist in recruit-

---

[1] http://www.metoffice.gov.uk/news/releases/archive/2011/weather-game

[2] http://www.riskliteracy.org/researchers/

ing more participants and collecting more accurate results (Venhuizen et al., 2013).

We build upon a previous study by Stephens (2011) called the Weather Game, which was conducted in collaboration with the MetOffice. The game starts by asking demographic questions such as age, gender and educational level. Then, the game introduces the "ice-cream seller" scenario, where given the temperature and rainfall forecasts for four weeks for two locations, users have to choose where to send the ice-cream seller in order to maximise sales. These forecasts describe predicted rainfall and temperature levels in three ways: (a) through graphical representations (original game), (b) through textual forecasts and (c) through combined graphical and textual forecasts. The textual format is generated with NLG technology as described in the next section. Users are asked to initially choose the location to send the seller and then they are asked to state how confident they are with their decision. Based on their decisions, the participants are finally presented with their "monetary gains", i.e. the higher likelihood of sunshine, the higher the monetary gains.

3 NLG Extension for the Weather Game

We developed two NLG systems (WMO-based and NATURAL) using SimpleNLG (Gatt and Reiter, 2009), which generate textual descriptions of rainfall and temperature data addressing the uncertain nature of forecasts in two ways:

1. WMO-based: uses the guidelines recommended by the WMO (Kootval, 2008) for reporting uncertainty. Consider for instance a forecast of sunny intervals with 30% probability of rain. This WMO-based system will generate the following forecast: "Sunny intervals with rain being possible - less likely than not." (Figure 1).

2. NATURAL: imitates forecasters and their natural way of reporting weather. For the previous example, this system will generate the following forecast: "Mainly dry with sunny spells".

4 Future Work

The Extended Weather Game is used in two ways:

- Firstly, to explore what type of information presentation can assist in decision making under uncertainty. The participants are presented with three main categories of information presentation: (1) graphical representa-

tions, (2) textual representations and (3) both.

- Secondly, we plan to use the information derived from the previous step, to develop an optimisation system, that is able to choose the right format of uncertain information presentation dependent on the data. We will then use the same setup to evaluate our optimisation approach.

5 Conclusions

This demo paper describes an NLG extension of the MetOffice Weather Game to be used for task-based evaluation and data collection for uncertain information presentation. At ENLG, we will demo the Extended Weather Game and we will discuss initial findings.

References

Edward T. Cokely, Mirta Galesic, Eric Schulz, Saima Ghazal, and Rocio Garcia-Retamero. 2012. Measuring risk literacy: The berlin numeracy test. *Judgment and Decision Making*, 7(1):25–47.

Albert Gatt and Ehud Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *ENLG*.

Albert Gatt, Francois Portet, Ehud Reiter, James Hunter, Saad Mahamood, Wendy Moncur, and Somayajulu Sripada. 2009. From Data to Text in the Neonatal Intensive Care Unit: Using NLG Technology for Decision Support and Information Management. *AI Communications*, 22: 153-186.

Haleh Kootval, editor. 2008. *Guidelines on Communicating Forecast Uncertainty*. World Meteorological Organsation.

Martin Manning, Michel Petit, David Easterling, James Murphy, Anand Patwardhan, Hans-Holger Rogner, Rob Swart, and Gary Yohe. 2004. IPCC Workshop on Describing Scientific Uncertainties in Climate Change to Support Analysis of Risk and of Options.

Liz Stephens, Ken Mylne, and David Spiegelhalter. 2011. Using an online game to evaluate effective methods of communicating esemble model output to different audiences. In *American Geophysical Union, Fall Meeting*.

Noortje Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. 2013. Gamification for word sense labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, pages 397–403.