Dynamic User Modelling for Personalised Report Generation of Time-series Data

Dimitra Gkatzia
Heriot-Watt University
EH14 4AS
Edinburgh, UK
dg106@hw.ac.uk

Helen Hastie
Heriot-Watt University
EH14 4AS
Edinburgh, UK
H.Hastie@hw.ac.uk

ABSTRACT

The proposed work focuses on demonstrating how personal interests and preferences can be used to improve report generation from time-series data. We apply these ideas in the weather and health informatics domains. In this extended abstract, we propose a dynamic User Model approach to data-driven generation that adds personalisation to the generated reports. Our approach proposes that a User Model should be updated and influenced from both user's traits and socially similar users taking into account the fact that these attributes may change over time. This is currently work in progress as part of a PhD by Dimitra Gkatzia and in connection with the Help4Mood project[1].

1. INTRODUCTION

It is widely accepted that NLG systems should generate realizations that are tailored to a particular user in order to improve the quality of the system. User-Model techniques have been previously implemented in NLG systems to tackle the challenge of personalisation and assist people in decision making [13]. The issue with current approaches is that they require either input from the user through questionnaires [12] or from experts who decide what different types of users may want to have generated [16]. The main issue with these approaches is that, as well as being subjective, the User Models can quickly become out of date. Using a data-driven approach, we hope to be able to handle a certain level of uncertainty with regards the accuracy of the User Model and generate *dynamic* User Models which can be continually updated.

Current work uses the Weather Forecasts data [11, 1] with a view to transitioning to the Help4Mood domain once data is available. The challenge we address for both domains is to generate a summary based on the interaction history, activity sensor data and a User Model in order to optimize relevance and pursue personalisation. In addition, we will investigate the idea of approximating User Models with information based on socially similar users. The idea being that a user is unique on the one hand, but also shares common qualities with the other users.

The weather forecasts intend to assist users in making the most suitable decisions according to certain user traits e.g. interests, occupation, preferences, because different users have different decisions to make with regards the weather conditions. For example, a farmer is more interested in temperature whereas a mariner is more interested in the wind speed and direction. Table 1 gives concrete examples of how User Models may influence generation. From the Weather Forecast records for one day the context that has been selected is the mode of sky cover, the minimum temperature and the mean value of windspeed. The first row of the table shows how the initial realization would be without a particular user in mind. The second row shows the added information of the wind direction and maximum wind speed that are valuable for a mariner as dictated by the mariner's User Model. Finally, the third row shows a possible generated weather report based on a farmer's User Model and thus the report is enhanced with the maximum temperature.

The Help4Mood project is intended to support the treatment of people with major unipolar depression in the community. Patients interact with a virtual agent and their level of depression is monitored through clinical questionnaires and sensors. After each session a report is generated such as the example given in Figure 1. This domain lends itself well to user modelling as there is quite a large amount of information available for each user including interaction and medical history, location, gender, language profile and social context as well as varying levels of depression as determined by the system and clinician. We aim to investigate what it means for users suffering from depression to be socially similar in terms of their traits as captured in the User Model and sensor data. For example, the fact that sensor data indicates less sleep than normal may affect different users in different ways. Finally, we will look at whether the User Model can be enhanced by the above findings in a collaborative manner by both user and clinician. The reader is referred to [18] for a full description of the project.

2. RELATED WORK

Summarization from time-series data has been researched widely and existing methods have been used in several domains; weather forecasts [2, 1, 15], clinical data summariza-

[1] http://www.help4Mood.info

Selected User Model	Realization	Extra fields or records used
None	Mostly clear, with a low around 41. Calm wind	None
Mariner's User Model	Mostly clear, with a low around 41. *South wind between 1 and 5.*	windDir, max windSpeed
Farmer's User Model	Mostly clear, with a low around 41 and *a high around 60.* Calm wind.	Temperature(max)

Table 1: Example of the realization from the original dataset, and two User Models: the mariner User Model and the farmer User Model.

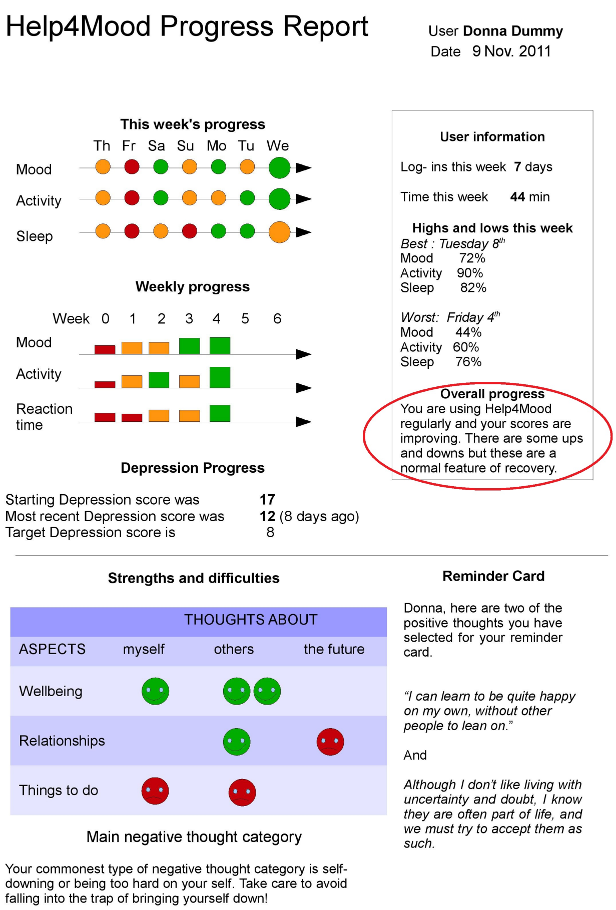

Figure 1: Example of report from a session with Help4Mood system

tion [8, 6], narrative from sensor data to assist children with communication needs [3], audiovisual debriefs from sensor data from Autonomous Underwater Vehicles missions [10].

Recent work on report generation has started to move away from hand-written rules to data-driven techniques including: statistical techniques from Machine Translation [2] and supervised machine learning [1]. NLG systems tend to be very domain-specific and data-driven systems that seek to simultaneously optimize content selection and surface realisation have the potential to be more domain-independent, automatically optimized and lend themselves to automatic generalization [1, 14, 4].

With regards personalisation, NLG systems have used a number of different approaches. In the STOP project, content personalisation was conducted using a questionnaire to acquire users' smoking behaviour and produce letters to assist them in quitting [12]. However, letters were only produced once based on the content of the user's response to the questionnaire and there is no dynamic User Modeling. More importantly, the questionnaire could not reflect users' change of habits and needs. An example of dynamic user modelling that adapts to users' background knowledge (expert vs novice users) with the notion that this can be changed over time is described by [9]. Other previous systems require the users to fill out questionnaires or tests in order to get personalised results [12, 17]. This approach is not ideal, particularly for users who get easily stressed or cranky such as with the users of Help4Mood. Other approaches involve the construction of a User Model influenced by testing the user's skill performance and imperfect User Models that exploit whatever knowledge is known for the user and then the user can ask additional questions to request information or clarifications [13].

User Models have been used in NLG systems and Spoken Dialogue Systems in order to add personalisation to the system's outcome (realization) by adapting to the user's background knowledge and preferences. [7] construct a Bayesian network that is based on experimental results that define individual differences, level of experience and domain task characteristics. Bayesian networks can assist in predicting user preferences regarding particular characteristics, thus it seems a promising way of modelling the user. [5] use a User Model that consults short and long term interests of a user in a newspaper domain in order to generate personalised summarization.

3. OUR APPROACH

We adopt a data-driven approach to generation similar to that described in [1]. Using crowd-sourcing, we aim to look at whether personalisation does in fact improve summarization and which attributes of a User Model are important for summarization, for example location, interests, previous interaction etc. The initial User Models will be based on a questionnaire to derive user's details, as used in the STOP project [12]. The proposed User Model technique associates user's preferences with particular time-series records (in our case weather records) and learns the relationship between these two elements in order to specify the content.

The User Model could also be estimated using statistical models from similar user's profiles, according to the techniques used in recommendation and collaborative systems, in order to be able to predict future changes in user profiles as well as to predict a new user's needs. Another approach to users' similarity can be Linear Function Approximation where each user can be regarded as a state with variable feature values.

Finally, user characteristics and preferences can change over time. Consequently, the system should sense that user preferences have been changed, requiring the User Model to be dynamic, updating certain attributes over time. For example, a farmer that changed location may not be interested in work-related weather for a particular time interval. The generation in that case may be influenced by similar users, e.g. users from that location. We will look at this phenomenon in relation to interactions with a Spoken Dialogue System, such as one for getting weather forecasts and the Help4Mood virtual agent.

4. ACKNOWLEDGEMENTS

The research leading to this work has received funding from the EC's FP7 programmes: (FP7/2011-14) under grant agreement no. 248765 (Help4Mood); (FP7/2011-14) under grant agreement no. 287615 (PARLANCE).

5. REFERENCES

[1] G. Angeli, P. Liang, and D. Klein. A simple domain-independent probabilistic approach to generation. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2010.

[2] A. Belz and E. Kow. Extracting parallel fragments from comparable corpora for data-to-text generation. In *Proceedings of the 6th International Natural Language Generation Conference (INLG'10)*, 2010.

[3] R. Black, J. Reddington, E. Reiter, N. Tintarev, and A. Waller. Using NLG and Sensors to Support Personal Narrative for Children with Complex Communication Needs. In *Proceedings of NAACL-10 Workshop on Speech and Language Processing for Assistive Technology*, 2010.

[4] N. Dethlefs and H. Cuayahuitl. Combining hierarchical reinforcement learning and bayesian networks for natural language generation in situated dialogue. In *Proceedings of ENLG*, 2011.

[5] A. Diaz and P. Gervas. User-model based personalized summarization. In *Journal Information Processing and Management: an International Journal, Volume 43 Issue 6*, 2007.

[6] A. Gatt, F. Portet, E. Reiter, J. Hunter, S. Mahamood, W. Moncur, and S. Sripada. From Data to Text in the Neonatal Intensive Care Unit: Using NLG Technology for Decision Support and Information Management. In *AI Communications 22:153-186*, 2009.

[7] B. Grawemeyer and R. Cox. A bayesian approach to modelling users' information display preferences. In *Proceeding UM'05 Proceedings of the 10th international conference on User Modeling*, 2005.

[8] J. Hunter, Y. Freer, A. Gatt, Y. Sripada, C. Sykes, and D. Westwater. BT-Nurse: Computer Generation of Natural Language Shift Summaries from Complex Heterogeneous Medical Data. *Journal of the American Medical Informatics Association 18:621-624*, 2011.

[9] S. Janarthanam and O. Lemon. Adaptive referring expression generation in spoken dialogue systems: Evaluation with real users. In *SigDial 2010 (Tokyo, Japan)*, 2010.

[10] N. A. R. Johnson and D. M. Lane. Narrative Monologue as a First Step Towards Advanced Mission Debrief for AUV Operator Situational Awareness. In *ICAR 2011: The 15th International Conference on Advanced Robotics, Tallinn June 20 - 23 / 2011*, 2011.

[11] P. Liang, M. I. Jordan, and D. Klein. Learning semantic correspondences with less supervision. In *Association for Computational Linguistics and International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, 2009.

[12] E. Reiter, R. Robertson, and L. Osman. Types of knowledge required to personalise smoking cessation letters. In *W.Horn et al (Eds.): Artificial Intelligence in Medicine: Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99)*, 1999.

[13] E. Reiter, S. Sripada, and S. Williams. Acquiring and using limited user models in nlg. In *European Natural Language Generation Workshop, pages 87-94*, 2003.

[14] V. Rieser, O. Lemon, and X. Liu. Optimising Information Presentation for Spoken Dialogue Systems. In *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, Uppsala, Sweden, 2010.

[15] S. Sripada, E. Reiter, I. Davy, and K. Nilssen. Lessons from Deploying NLG Technology for Marine Weather Forecast Text Generation. In *Proceedings of PAIS session of ECAI-2004, pp760-764*, 2004.

[16] S. Sripada, E. Reiter, J. Hunter, and J. Yu. Segmenting time series for weather forecasting. In *In Applications and Innovations in Intelligent Systems X*, pages 193–206. Springer-Verlag, 2002.

[17] S. Williams and E. Reiter. Generating readable texts for readers with low basic skills. In *10th European Workshop on Natural Language Generatione*, 2005.

[18] M. Wolters, J. Martinez-Miranda, H. Hastie, and C. Matheson. Managing data in Help4mood. In *2nd International Workshop on Computing Paradigms for Mental Health - MindaCare*, 2012.