Automatically Generating Stories from Sensor Data

Joseph Reddington
Dept. of Computing Science
University of Aberdeen
j.reddington@abdn.ac.uk

Nava Tintarev
Dept. of Computing Science
University of Aberdeen
n.tintarev@abdn.ac.uk

ABSTRACT

Recent research in Augmented and Alternative Communication (AAC) has begun to make use of Natural Language Generation (NLG) techniques. This creates an opportunity for constructing stories from sensor data, akin to existing work in life-logging. This paper examines the potential of using NLG to merge the AAC and life-logging domains. It proposes a four stage hierarchy that categorises levels of complexity of output text. It formulates a key subproblem of clustering sensor data into narrative events and describes three potential approaches for resolving this subproblem.

Author Keywords

NLG, AAC, event generation, narrative, story, sensors, life-logging

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User-centered Design,Natural Language,Input Devices and Strategies

General Terms

Experimentation, Design, Human Factors

INTRODUCTION

Advances in text-to-speech technology and mobile computing have made a large range of Augmented and Alternative Communication (AAC) devices available to the public [1] [2] [3]. Such devices are intended to support users who have cognitive or motor limitations on producing speech. The communication allowed by these advances can enable users to live more complete and satisfying lives than would otherwise be possible.

[1] http://uk.dynavoxtech.com/products/xpress/, last retrieved November 2010.

[2] http://www.techcess.co.uk/3_3_smart.php, last retrieved November 2010.

[3] http://www.liberator.co.uk/index.php/vantage-lite-1.html, last retrieved November 2010.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IUI'11, February 13–16, 2011, Palo Alto, California, USA.
Copyright 2011 ACM 978-1-4503-0419-1/11/02...$10.00.

Despite advances in AAC devices, creating sentences 'on the fly' for spontaneous conversation is still slow and difficult (typically from 8 to 10 words per minute, or up to 12 to 15 per minute when techniques such as word prediction are used [8]). In the case of severe difficulties, typing may not be possible at all. This results in a situation where new utterances must be prepared in advance either by the user or a carer, with a large time and energy cost. Recent or single use events, such as talking about one's day or talking about yesterday's television, are expensive to prepare in advance relative to the limited potential for future (re-)use. As a result, AAC users tend to be passive, responding to questions with single words or short sentences. When personal stories are told, they tend to be as a monolog or a sequence of pre-stored utterances [15].

This work challenges this difficulty by examining the potential of using automatically collected information to generate new phrases, allowing AAC users to increase social interaction and reduce the overhead involved in maintaining a communications device. It also attacks the Natural Language Generation (NLG) task of document planning: structuring narrative from unstructured data.

This paper tackles the problem of generation of narrative summaries of daily events. This problem can be split into two separate phases: 'clustering' the data into events; and deciding which events would be of interest to the user, or the listener. This paper focuses primarily on the first: algorithms for setting event boundaries of stories based on sensor data.

RELATED WORK

This section examines work in AAC, NLG, and life-logging that relates directly to automatic content creation.

AAC Movement towards automatic generation

Although NLG has featured in AAC research before (e.g. [10]), more recently research has moved toward supporting automatic generation of content. [3, 14] use external data to populate a communication device, and [4] uses a NLG engine to generate text from a database of personal facts. Such research complements a growing range of AAC devices that support internet connectivity [4] [5].

[4] http://www.proloquo2go.com/, last retrieved November 2010

[5] http://www.taptotalk.com/, last retrieved November 2010

It is the position of the authors that future AAC devices can and should support the automatic inclusion of new text and that there is a clear and present research need to provide methodologies to enable this.

Automatic Life-logging

Research focusing on collecting personal data over longer periods has been referred to as life-logging. Microsoft's MyLifeBits [7] introduced a system for collecting data such as images, GPS location data, text and audio annotations. The EyeTap project designed a head mounted camera for the collection of images as seen by the eye [9]. Media such as audio has also been collected, analysed [6, 11], and tagged with contextual information. In particular, [11] has been used for the MIT based Reality Mining project, which analyses longitudinal data for large samples of people [5].

Natural Language Generation

Natural Language Generation (NLG) is a subcategory of natural language processing that examines the creation of English text from nonlinguistic data such as sensor readings. NLG techniques can dynamically combine and change the output depending on the changing internal state of the system [12]. NLG has been developed for a number of areas including weather forecasting [16], and in hospitals [1]. Much research in NLG has focused on summarising technical data for expert users, with the goal of effectively communicating key information. In contrast, this work focuses on everyday events for the purpose of easing social communication.

AAC is an interesting domain for a NLG system connected to a logging system. This supports and scaffolds conversation for users with speech impairments, and creates a context in which this type of logging is reasonably not intrusive. It also creates a setting for studying the NLG challenge of creating a structured narrative from unstructured data.

PROBLEM FORMULATION

This section categorises the potential outputs of automatically generated content into a four-tiered hierarchy of meta-conversation, network-based input, sensor-based input, and the creation of narratives from sensor input. It then identifies the context in which the clustering problem occurs and goes on to give details of three approaches that could be used.

Hierarchy of automatically generated content

The potential uses of automatically generated content for AAC devices divide into four broad categories, based on ascending levels of complexity. Here 'input' is used in the sense of 'input to the device as new content' - it can also be output directly from the user.

Inferred input

Inferred input is defined as utterances that can be generated from examination of previous user utterances. Thus, if a device registered the phrase "Hello Mary" and later "Thank you Mary" it would be reasonable to deduce that the user had spent some time with an individual called Mary and so the phrase "Today I spent time with Mary" could be added to the list of available phrases (later, of course, becoming "Yesterday I spent time with Mary"). An advanced version of this type of work is reported in [4], where suggested topics of conversation are based on previous ones. An interesting example is [2] where content is generated by a log of a computer-based conversation with an intubated patient's friends and family.

Network-based input

Network-based input is defined as new utterances that can be determined by access to information over the Internet, or some other information portal. An example is talking about the weather - phrases such as "It's very warm today", "It's going to rain tomorrow", and "It snowed on Sunday!". Also included are observations about recent media: "On YouTube I watched a video called 'the Four Chord Song' ".

Sensor-based input

Sensor-based input is defined as the use of single facts about the user provided by sensor data. Examples might include "I went to the supermarket" - provided by GPS data, or "I saw a book in the supermarket called 'Breaking Dawn' that I really like" - provided by use of a barcode scanner and an online database. Although this sort of data collection can affect both privacy and also the workload required to maintain it, utterances can be better adapted: "I got a text message from Jamie this morning, he said 'looking forward to tomorrow' ". Voice messages recorded by care staff, for example, are also included in this category: recorded in the first person, they can include information that would never be picked up by a sensor - an example from this study is: "I had spag bol today, it was thumbs up but wasn't as good as my Dad's!".

Creation of narratives from sensor data

This category contains groups of messages, created from sensor data, that together relate an experience or tell a story. This adds the problem of creating a narrative structure and consistent style to the sensor-based input (for NLG work on narrative importance see e.g. [13]). An example might be:

> I had my breakfast quickly because I was excited to go to the arcade. I got on the bus, I went to the arcade, I played the games at the arcade and won a cuddly bear.

which can be formed from location data, voice recordings and RFID data to identify objects and people. The creation of multi-fact, multi-sentence messages with a structured narrative is a big move forward in NLG terms, requiring more sophisticated techniques than previous steps in the hierarchy.

This paper focuses particularly on this analysis of sensor-based data, defining one of these multi-fact and multi-sentence messages as an 'event'. While the NLG techniques outlined in [12] can be used to combine facts into plain English, the challenge is in defining boundaries between groups of sensor data to separate events. The goal is to arrange the sensor-based input into solid narrative structure that accurately relates events that happened and is similar to the type of story that equivalent typically developing adults and children might tell.

09:34, Voice Recording, A man came to talk to me in gym. I signed an important document.
09:36, Voice Recording, My dad was in school today.
09:40, Location, Gym.
09:40, Object, Skittles
09:40, Object, Baton
09:40, Person, Mrs Table
10:48, Voice Recording, I was doing relay racing at school and I joined in really well.
10:48, Location, Changing
10:50, Voice Recording, I did some high jump and every time I jumped I had a lay down on the mat.
11:08, Location, Classroom
11:12, Object, Blackboard
11:31, Person, Mary
11:36, Location, Tutorial Room
11:36, Object, Money
11:39, Object, Monkey Game
11:58, Location, Classroom

Figure 1. Example data from the HWST Project

Telling Stories in School

The dataset for our examples was provided by the ' "How was School Today?"...' (HWST) project, which logged sensor data for students with complex communication needs at a special needs school. Figure 1 gives an example of the anony-mised data used in this work.

The HWST Project [3, 14] logs object and person interactions, voice recordings, and location information (at the room level). It also records positive and negative evaluations (e.g."It was not a good day."). This framework is operational and is used in the context of generating stories for children at the school.

For this particular domain, the types of data recorded for each user are:

- Location data - each time the user enters a new room, this information was recorded (Pre-processing removed rooms entered for less than three minutes).
- Object interaction - each time the user interacted with an object that had an RFID tag that interaction was recorded.
- Person interaction - each time the user interacted with a person that had an RFID tag that interaction was recorded.
- Voice messages - staff and teachers were encouraged to record voice messages, as if the user was speaking in the first person, that described the user's recent activities.

Clustering Algorithms

This section describes three potential algorithms for grouping elements into events. Each algorithm is motivated by a different mental model of how children may structure a narrative about an important event in their day.

Location based clustering

Location based clustering assumes that a change of location is likely to denote a change of context, and can thus be used as an event barrier. A potential disadvantage with this method is that some target users conduct much of their activities in the same location, e.g. in a home classroom. To compensate, this algorithm also considers an upper bound for time, creating a new event once a certain amount of time has passed. Once a new event with location is identified, the additional sensor data (who was there, what objects were interacted with, any voice recordings) is tied to the event.

For example, a cluster generated from Figure 1 would be:

11:36, Location, Tutorial Room
11:36, Object, Money
11:39, Object, Monkey Game

When converted into English text, the above cluster gives the story:

> I played with Money and Monkey Game. This happened at a Tutorial Room.

Time based clustering algorithm

The time based clustering is a hierarchical clustering based on temporal proximity. The motivation behind this algorithm is to create a 'clean' algorithm for clustering that prioritizes temporal closeness over the supremacy of any given data type. This algorithm groups data elements strictly according to the temporal proximity of the data. That is, things that happened around the same time are likely to belong to the same event.

This algorithm first considers each element to be its own unique cluster, and then seeks the two elements that are closest together in time without being in the same cluster. Once found, the two clusters are merged in classic greedy algorithm fashion. This process continues until the desired number of clusters are obtained. Thus, the number of events can be specified in advance and no data is left out of the clusters.

Voice recording based clustering

Clustering based on voice recordings was designed to capitalise on the richness of the data in voice recordings, and replicate the sentence structure used by the target age group.

This algorithm takes the set of recordings as the set of events, and uses the other data to establish where the recording took place, and what people and objects had been involved. For example, given the recording "We did the high jump today, and every time I jumped I had a little lay down on the mat." the system would deduce the location ('gym'), the people present ('Mrs Table, PE teacher'), and the objects that had been interacted with ('baton','skittles') and produce the cluster: ('gym', 'Mrs Table, PE teacher', 'baton', 'skittles', 'We did the high jump today, and every time I jumped I had a little lay down on the mat.'). The NLG engine used would convert such a cluster into:

> I went to the gym today, Mrs Table was there, I played with skittles and the baton, we did the high jump and every time I jumped I had a little lay down on the mat.

This algorithm works on the premise that, if a voice recording has been made, this should be a reportable event. During preprocessing, voice recordings within 15 minutes of each other are merged. Note that the example story used for location based and time based clustering is discarded by this algorithm because it contains no voice recording. This potentially removes 'uninteresting' stories early on but may remove too much information.

CONCLUSION AND ONGOING RESEARCH

This paper presented three algorithms for defining boundaries between events. This paper also presented a four stage hierarchy for the classification of automatically generated input. The next natural step for this work is to evaluate the narratives generated with these algorithms. Another goal is to examine the second subproblem of story generation; that of grading relevance of generated clusters. The intended usage of an AAC system is as a *dialogue* between a child and a second party, and future evaluations can focus an evaluation on a dialog, or its transcript. While it may be harder to control such an experiment, this type of study is likely to result in a more in depth understanding of dialogue in personal narrative, such as when and how communication breakdowns may occur, or when usages of repetition may be applicable.

ACKNOWLEDGEMENTS

The authors are grateful to the members of the original ' "How was school today?"...' team. The authors acknowledge the help of Margaret Mitchell for evaluation, the Aberdeen NLG group for insightful comments, and Steven Knox for code development. Thanks also to our reviewers for their help in reducing this paper.

REFERENCES

1. 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. *AI Commun.*, 22(3):153–186, 2009.
2. S. Ashraf, A. Judson, I. W. Ricketts, A. Waller, N. Alm, B. Gordon, F. MacAulay, J. K. Brodie, M. Etchels, A. Warden, and A. J. Shearer. Capturing phrases for ICU-Talk, a communication aid for intubated intensive care patients. In *ACM Conference on Assistive technologies*, pp. 213–217, New York, NY, USA, 2002.
3. R. Black, R. Turner, E. Reiter, and A. Waller. A new tool to support interactive narrative experience for children with communication disorders. *Conference of the International Society for Augmentative and Alternative Communication*, Barcelona, July 2010.
4. M. Dempster, N. Alm, and E. Reiter. Automatic generation of conversational utterances and narrative for autmentative and alternative communication: a prototype system. *NAACL HLT SLPAT Workshop*, pp. 1–9, Los Angeles, California, June 2010.
5. N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63:7:1057–1066, 2009.
6. D. Ellis and K. Lee. Minimal-impact audio-based personal archives.*Workshop on continuous archival and retrieval of personal experiences in Multimedia*, 2004.
7. J. Gemmell, L. Williams, K. Wood, R. Lueder, and G. Bell. Passive capture and ensuing issues for a personal lifetime store. In *CARPE Workshop in Assoc. with the ACM Multimedia Con*, pp. 48–55. 2004.
8. D. J. Higginbotham, H. Shane, S. Russell, and K. Caves. Access to AAC: Present, past, and future. *Augmentative and Alternative Communication*, 23(3):243–257, January 2007.
9. S. Mann. Continuous lifelong capture of personal experience with eyetap. In *CARPE Workshop in Assoc. with the ACM Multimedia Conference*, 2004.
10. K. F. Mccoy and P. Demasco. Some applications of natural language processing to the field of augmentative and alternative communication. In *Proceedings of the IJCAI-95 Workshop on Developing AI Applications for People with Disabilities*, pp. 97–112. Press, 1995.
11. M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4:51–59, 2005.
12. E. Reiter and R. Dale. Building natural language generation systems, 2000.
13. E. Reiter, A. Gatt, F. Portet, and M. van der Meulen. The importance of narrative and other lessons from an evaluation of an NLG system that summarises clinical data. *INLG '08*, pp. 147–156, Morristown, NJ, USA, 2008. Association for Computational Linguistics.
14. E. Reiter, R. Turner, N. Alm, R. Black, M. Dempster, and A. Waller. Using NLG to help language-impaired users tell stories and participate in social dialogues. *European Workshop on Natural Language Generation (ENLG-09)*, 2009.
15. G. Soto, E. Hartmann, and D. Wilkins. Exploring the elements of narrative that emerge in the interactions between an 8-year-old child who uses an AAC device and her teacher. *Augmentative and Alternative Communication 22(4)*, 2006.
16. E. Goldberg, N. Driedger, and R. I. Kittredge. Using natural-language processing to produce weather forecasts. *IEEE Expert: Intelligent Systems and Their Applications*, 9(2):45–53, 1994.