Towards a Diagnostic Toolbox for Medical Communicaton

William Billingsley[1], Cindy Gallois[1], Andrew Smith[1], Timothy Marks[1], Fernando Bernal[1], Marcus Watson[1,2]

[1] NICTA, Level 5, Axon Building, University of Queensland, Queensland 4072, Australia
[2] Skills Development Centre, Queensland Health, PO Box 470, Herston, Queensland 4029, Australia
{William.Billingsley, Cindy.Gallois, Andrew.Smith, Timothy.Marks, Fernando.Bernal, Marcus.Watson}@nicta.com.au

Abstract. Poor communication is a major cause of adverse patient events in hospitals. Although sophisticated simulators are in use for performing medical operations, there is comparatively little technology support being used for improving communication skills including patient history taking. Artificial Intelligence and Natural Language Processing researchers have developed sophisticated algorithms for analysing conversations. We are experimentally developing software that can visualise the combined output of these algorithms, as a diagnostic toolkit for medical communication.

Keywords: Communication, Visualisation, Patient History Taking, Natural Language Processing, Artificial Intelligence.

1 Introduction

Adverse patient events occur in between 3.7% [1] and 16.6% [2] of hospitalisations. Even the studies that have found the lower rates have concluded that a substantial amount of patient harm happens through medical management. Failures of communication have been found to be a major contributor in up to forty per cent of these events [3], and seventy per cent of those that cause serious harm [4]. One Australian study attributed communication as a major cause to approximately 14,000 preventable deaths each year [5]. If failures of communication are such a large contributor to patients being harmed or killed within the health service, then this suggests that improving medical professionals' skills and habits in communication would make a significant improvement to patient outcomes.

Technology now plays a pivotal role in diagnosing patient health issues. Particularly, visualisations of technical inspections of the body – such as ultrasound, electrocardiograms, and magnetic resonance imagery – have become a day-to-day part of diagnostic practice. We propose that technical inspections of medical communication could similarly help to diagnose and understand how it fails.

There are, of course, many different modes of communication in the health service. A failure of communication could include, for instance, that something was written incorrectly on a patient chart, that a nurse was not informed of relevant information that was known to a doctor, or that a misunderstanding took place during a medical conversation. It is very unlikely that a single technology could be designed that could encompass all of these different kinds of communication and the ways they can fail. Instead, we propose starting with a kind of communication event that every patient undergoes, and that precedes every other medical interaction: patient history taking.

2 Background

Many research projects have analysed medical communication [6, 7, 8, 9]. Most studies have involved manually coding and classifying particular utterances or events in a conversation. There are a number of software tools, such as Transana[1] and Noldus's products[2], that assist observers to mark up these events manually on a transcript or video, and then analyse them, for example through connections to statistics packages. A recent meta-analysis of the analysis techniques that are commonly used [10] found that most studies, for practical reasons, focused on very short-term effects within a conversation – such as the influence of one participant's utterance on what the other participant says next. It also recommended that multiple analysis methods should be used, to better cope with the complexity of clinical communication. Besides the practical limitations of utterance coding studies, it is also clear that manual coding of a conversation is very laborious to perform

Communications researchers are not the only academics who have an interest in analysing conversations. Within computer science, Artificial Intelligence and Natural Language Processing groups have developed a number of algorithmic techniques for analysing a conversation. For example, Leximancer [11], Latent Semantic Analysis [12], and Latent Dirichlet Analysis [13] all analyse how words co-occur within a text to discover the flow of topics. It can be difficult to prove objectively that word co-occurrence is the same thing as a topic, but comparisons against human analysis have found that the algorithms' results correlate well with human judgement [14, 15]. We are also aware of communications researchers who use these tools in their research, so the algorithmic approaches are gaining acceptance.

Applying an algorithmic analysis would make it feasible to analyse doctors' clinical conversations as part of their training or in professional development. However, we can go further. Just as Connor, *et al*, found in their review of manual analysis that it would be better to use multiple techniques to capture the full complexity of clinical communication, we believe the same is also true for automatic analysis. By combining different analyses, we can uncover more about the conversation. For example, colleagues of ours have developed algorithms for estimating the cognitive load of a speaker by analysing the sound of his or her speech [16]. By combining that with the text analysis, we would be able to see both what

[1] http://www.transana.org
[2] http://www.noldus.com

participants are saying and how much thought they are having to put into their words. Each additional algorithmic tool can add an extra dimension to the analysis.

3 Concept of Operation

The analysis system we are developing has a visualisation client and a data server. Both the client and the server accept plug-ins to extend their functionality. This extensibility is necessary as it is not certain which kinds of computer analysis will prove to be informative about a medical conversation, and which will not.

The server has to make as few assumptions as possible about the data it will be handling. Different analysis components use different kinds of data. For this reason, we often depict the data repository as a bucket, as shown in Figure 1.

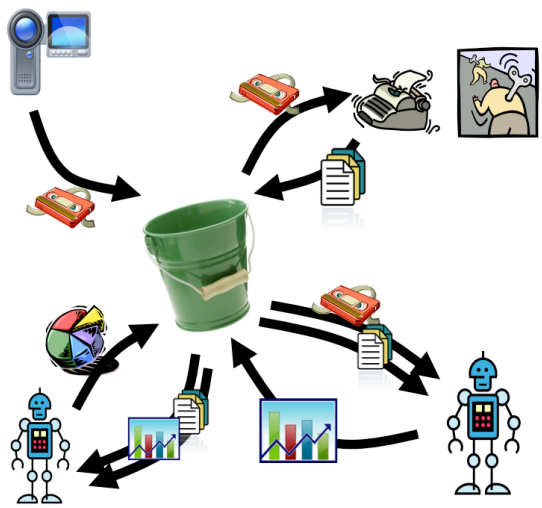

Fig. 1. The server repository acts as a generic bucket for data. When a video is uploaded, it begins to be processed through the various installed plug-ins automatically. In the figure, processing starts at the top-left and proceeds clockwise. Each plug-in takes data from the repository, passes it to an external processing component, and puts the resulting output into the repository as data that other plug-ins can then use in their processing.

When a video is uploaded into the repository, the server begins to process it automatically. At each step in the processing, a plug-in takes some form of data out of the repository, processes it, and adds a new form of data back into the repository. One of the first steps in the workflow (one of the first plug-ins invoked) is likely to be to transcribe the video. This may be a semi-automatic step, in which a speech-to-text engine produces a draft transcript, which is then edited manually using common

transcription editing software. A second plug-in would take the transcript and process it through an NLP engine. A third plug-in would process the sound through a cognitive load analysis tool. Another plug-in might correlate the cognitive load data with the topic data for the video, and identify common stress patterns.

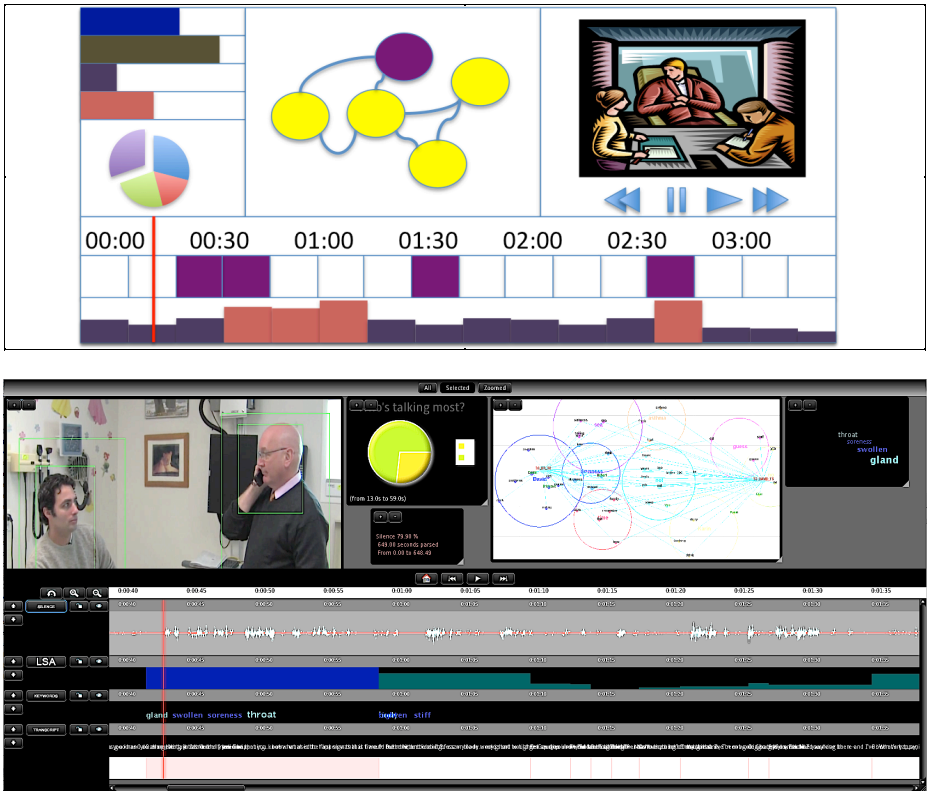

Fig. 2. Stylised and real images of the client. The client plays back visualisations of the server data alongside the video. The visualisations are linked to the video, and so can also be used to navigate the video. The client supports a plug-in architecture, both so that it can be extended for new kinds of data, and also so that different visualisations can be used for different kinds of user.

The final result of the server processing is that there is a large amount of analysis data in the repository. The data can be played back as visualisations alongside the video, as depicted in Figure 2. Just as the server supports a plug-in architecture so that we can continually extend it with new kinds of analysis, the client also supports plug-ins so we can extend it with new kinds of visualisation. The mapping of server plug-ins to client plug-ins is not one-to-one. As we describe later in the paper, the presentation of the data has to be varied for a number of different kinds of user, so we expect there to be more visualisation plug-ins than analysis plug-ins. It is also worth

mentioning that advice is a kind of visualisation. By providing an appropriate plug-in, we can turn the visualisation plug-in into an explicit teaching tool.

3 Patient History Taking

One of the ways trainee doctors are taught patient history taking and communication skills is through simulated patients. An actor is trained to portray a patient scenario. The actor is trained to portray not only the history and symptoms of the patient, but also the emotional and personality characteristics. These portrayals are often standardised ("Standardised Patients"). The trainee doctor is asked to examine this simulated patient. At some institutions, these simulated patient sessions are video recorded for later review, rather than being reviewed live. Reviewing a video session thoroughly is a time-intensive task, so there are practical limitations on how many scenarios a trainee doctor can undertake.

As the sessions are already being recorded, they make an ideal first candidate for automatic analysis. The video recording can be processed by a transcriber, sound extractor, and multiple different analysis engines. This data can then be compiled, and the video played back together with the analysis results. Because the scenarios are standardised and well-understood, they make a very good research test-bed for analysis technologies. The fact that the educators already know what is wrong with the patient and what kinds of interactions to look for means that the analysis can be pre-seeded with information. This can help to mitigate any limitations of the analysis algorithms. If the algorithmic analysis proves successful, then we can begin to apply it educationally. Because algorithmic analysis would not involve the reviewer (though it would require time from the actors and someone to correct transcription errors) it would be possible for medical students to perform additional practice sessions using the system.

There have been previous attempts at automated teaching of patient history taking. Particularly, we are aware of interactive computer simulations of patients [17, 18]. The virtual patient listens to the doctor through voice recognition or free text entry, and responds through pre-recorded video. The computer simulations are necessarily lower fidelity than using a live actor, and they are also lower fidelity in their analysis than the multi-modal detailed algorithmic analysis we have proposed. We would be interested in augmenting a computerised simulation with fine-grained AI and NLP based communication analysis, to see whether it can be effective in that situation. As the computerised patient cannot vary its intonation, body language, and emotion as effectively as the human actor, however, it is possible that doctor-computer conversations would always appear somewhat stilted.

4 More Complex Use Cases

A second scenario we are interested in investigating is the shift hand-over meeting. Shift-handover is recognised as being a critical communication system, which has informational, social, educational functions [19]. Despite the hand-over's importance,

though, there are comparatively few studies into factors that make a hand-over successful or unsuccessful. Likewise, in medical practice, it is difficult to monitor how well hand-overs are being performed.

We would like to see what automated AI and NLP analysis can reveal about these meetings. By recording hand-over meetings and processing them through the analysis tools, we hope to be able to inspect them in a way that is both more detailed and more efficient than manually coding utterances.

Technically, shift hand-overs are a very much more challenging problem than simulated patient scenarios. There are many more people involved in the meeting, and the number of communication channels (the number of possible two-way conversations) grows with the square of the number of participants. Similarly, whereas the simulated patient scenarios discuss a single medical case, the shift hand-overs discuss an entire ward. Furthermore, the shift cases are not standardised, so the software would not know ahead of time what the patients' conditions and issues are. Nonetheless, it may be possible to identify common failure modes in the communication. At the very least, we hope to uncover more detail about how the meetings function.

If a hand-over meeting is transcribed and analysed by a NLP and AI components, then this also provides a way of indexing the meeting, not just against words but against topics and other measures. If other clinical conversations are also recorded and indexed, it may be possible to forensically analyze how information is conveyed from one conversation to another. For example, it might eventually be possible to follow a piece of information from one medic's interaction with a patient, through a shift hand-over, to another medic's interaction with the same patient.

5 Status and Plans

We have constructed prototype software for our diagnostic tool, which is being continuously refined. We have begun to build our first processing and visualisation plug-ins, and have established collaborations with AI and NLP teams in order to add more. There is still, however, much work to be done before we can place an educational tool in front of a doctor.

Firstly, we need to experiment with the tool. We need to process different kinds of conversation through various plug-ins, to find out what kinds of analyses combine well and add meaning – which combinations are more than the sum of their parts, and which are not. This is an experimental discovery process. Then, we need to work out ways of showing this data meaningfully to doctors. The data has been derived through complex algorithms, but must be understood by someone who has no knowledge of those algorithms, but who nonetheless works in an evidence-driven field and who will probably want to know the reasons not just the result.

The visualisations that the team needs during the discovery process might be different to those the doctors need in their professional development. So, we might find we need to develop visualisations and measures twice. The system is designed to be very adaptable, however, and can be configured differently for researchers than for doctors. We also hope to make it available to other researchers to use in their work.

Our first experiments are scheduled to take place in 2010, with the simulated patient scenarios. We also intend to apply the tool to a separate study comparing think-alouds by different practitioners. We hope that applying it in this study will help us to see how we can make the tool more useful for researchers. In 2011, we are scheduled to begin our investigations into hand-over meetings, and also hope to have developed a useful educational tool with the system.

4 Conclusion

The diagnostic toolbox we are creating is intended to be useful to both researchers and practitioners. We have a collaborative relationship with the Skills Development Centre of Queensland Health (the state hospital system), which gives us both the access to experimental subjects and scenarios that we need, and also a route to real impact if our research is successful. Our first goal is to uncover combinations of analysis that can provide meaningful insights into medical communication. Our second goal is to provide a useful educational tool that can present that meaningful analysis of communication to doctors, so they can understand more about their own communication. Through those two goals, we hope to provide a way of improving medical communication, at least in the area of patient history taking, and reduce the rates of patient harm.

Acknowledgments. The authors gratefully acknowledge their colleagues on the project: Harold Cruz and Michael Pickard. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

5 References

1. Brennan, T. A., Leape, L.L., Laird, N.M., Hebert, L., Localio, A.R., Lawthers, A.G., Newhouse, J.P., Weiler, P.C., Hiatt, H.H.: Incidence of adverse events and negligence in hospitalized patients: results of the Harvard medical practice study. Quality and Safety in Health Care, vol. 13, no. 2, pp. 145 (2004)
2. Wilson, R., Runciman, W., Gibberd, R., Harrison, B., Newby, L., and Hamilton, J. The quality in Australian healthcare survey. The Medical Journal of Australia, vol. 163, pp. 458-471 (1995)
3. Australian Institute of Health and Welfare. Sentinel events in Australian public hospitals 2004-05. Australian Institute of Health & Welfare (2005)
4. Leonard, M., Graham, S., and Bonacum, D. The human factor: the critical importance of effective teamwork and communication in providing safe care. Quality and Safety in Healthcare, vol. 13, pp. i85—i90, (1995)
5. Zinn, C. 14000 Preventable deaths in Australian hospitals. British Medical Journal, vol. 310, pp.1487, (1995)

6. Roter, D., Larson, S.: The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. Patient Education and Counseling, vol. 46, issue 4, pp.243-51 (2002)
7. Apker, J., Mallak, L.A., Applegate, E.B., Gibson, S.C., Ham, J.J., Johnson, N.A., Street, R.L.: Exploring emergency physician-hospitalist handoff interactions: development of the Handoff Communication Assessment., Annals Of Emergency Medicine, vol. 55, pp.161-70. (2010)
8. Siminoff, L.A., Graham, G.C., Gordon, N.H.: Cancer communication patterns and the influence of patient characteristics: Disparities in information-giving and affective behaviors. Patient Education and Counseling, vol. 62, issue 3, pp.355-360 (2006)
9. Street, R.L., Gordon, H.S., The clinical context and patient participation in post-diagnostic consultations. Patient Education and Counseling, vol. 64, issue 1, pp.217-224 (2006)
10. Connor, M., Fletcher, I., Salmon, P.: The analysis of verbal interaction sequences in dyadic clinical communication: A review of methods. Patient Education and Counseling, vol. 75, issue 2, pp.169-177 (2009)
11. Smith, A.E.: Automatic extraction of semantic networks from text using Leximancer. In: HLT-NAACL 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: Companion volume, pp. Demo23—Demo24, ACL, Edmonton (2003)
12. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science, vol. 41, no. 6, pp.391—407 (1990)
13. Blei, D. M., Ng, A.Y., Jordan, M. I.: Latent Dirichlet allocation. Journal of Machine Learning Research, vol. 3, pp.993-1022 (1993)
14. Smith, A.E., Humphries, M.S.: Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. Behaviour Research Methods, vol. 38, issue 2, pp.262—279 (2006)
15. Landauer, T.K., Laham, D., Rehder, B., Schreiner, M.E.: How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. Proc. of the Nineteenth Annual Conference of the Cognitive Science Society, pp.412 (1997)
16. Yin, B., Chen, F.: Towards automatic cognitive load measurement from speech analysis. In: Human-Computer Interaction: HCI Intelligent Multimodal Interaction Environments, Springer-Verlag, Berlin, pp.1011-1020 (2007)
17. Bergin, R.A., Fors, U.G.H.: Interactive simulated patient—an advanced tool for student-activated learning in medicine and healthcare. Computers & Education, vol. 40, issue 4, pp.361—376 (2003)
18. Stevens, A.: The use of virtual patients to teach medical students history taking and communication skills. American Journal of Surgery, vol. 191, no. 6, pp.806—811 (2006)
19. Kerr, M.P.: A qualitative study of shift handover practice and function from a socio-technical perspective. Journal of Advanced Nursing, vol. 37, no. 2, pp.125—134 (2002)