
COMLEX: Visualizing Communication for Research and Saving Lives

William Billingsley

NICTA Queensland Research
Laboratory
Level 5, Axon Building (47)
Staff House Road
St Lucia, Qld 4072,
Australia
William.Billingsley@nicta.com.au

Andrew Smith

NICTA Queensland Research
Laboratory
Level 5, Axon Building (47)
Staff House Road
St Lucia, Qld 4072,
Australia
Andrew.Smith@nicta.com.au

Cindy Gallois

NICTA Queensland Research
Laboratory
Level 5, Axon Building (47)
Staff House Road
St Lucia, Qld 4072,
Australia
Cindy.Gallois@nicta.com.au

Marcus Watson

NICTA Queensland Research
Laboratory
Level 5, Axon Building (47)
Staff House Road
St Lucia, Qld 4072,
Australia
Marcus.Watson@nicta.com.au
and
Skills Development Centre,
Queensland Health

Abstract

One of the major causes of patient harm in hospital is poor communication. We are developing a video review and visualization platform to research and improve medics' communication skills. It intended for use by experimenters, as a deployable training tool for medics, and also for forensic review of communication. It supports pluggable analysis modules and visualizations for research teams, and configurable workflow for educators and hospital administrators.

Keywords

Communication skills, Professional development tools, Medical simulation, Video review, Visualization

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors, Experimentation, Measurement

Introduction

A typical hospital might harm as many as one in six of its patients [1]. Failures of communication are a major contributor to forty per cent of adverse patient events [2] and seventy per cent of those that cause patients serious harm [3]. (Harm, here, includes failures of inaction, as well as direct medical error.) In Australia

Copyright is held by the author/owner(s).

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

alone, communication has been attributed as the major cause for approximately 1,500 preventable deaths every year [4].

Although hospitals now have very sophisticated simulators that allow medics to practice operations, there is comparatively little technology to support and improve medics' communications skills. In the COMLEX project, we are looking at this problem from a human factors and computer science perspective. How can technology be used to identify communication or failures of communication, and how can that technology be used to improve clinicians' communications skills in practice?

As we describe in this poster abstract, one tool we are developing is a video review system that can replay a conversation together with combined analysis from various research-grade analytical engines (any that someone would care to write a plug-in for), but made more accessible to non-technical users. This is a tool for medics to review their own communication; it is a tool for communication researchers to analyze communication; and it is a tool for technical researchers to experiment and discover what kinds of analysis are useful and meaningful in practice.

A simple student scenario

One of the few communication exercises that medical trainees already undergo involves an actor. The actor is asked to play a patient with a particular medical condition, and is sent to have a consultation with the medical student. The actor might also have been asked to play the patient in a particular way – for example, to be open and forthcoming, or closed and distrustful. The consultation is video recorded and reviewed by a more

senior clinician or a communication expert. Reviewing the videos takes much longer than the consultation itself, and because the reviewers' time is limited and expensive, students get to perform very few of these exercises.

As these exercises are already video recorded, it is easy to imagine how some (hypothetically capable and meaningful) machine analysis could be used to allow students to perform more of them relatively cheaply. As we already know what is wrong with the patient, because we told the actor, we can give the machine analysis foreknowledge of what to look for, which makes the technical challenge somewhat easier.

Formative rather than summative

In anecdotal experience, medical workers tend to resist technology that might criticize or rate practitioners. For instance, some research colleagues developing an automated directory to help doctors find experts in particular fields were asked by a worried interviewee "But what if it doesn't rank me first?" Performance metrics are often resisted in many fields, but this seems to be particularly true of medicine.

Consequently, we would prefer to target the self-review tool towards formative advice rather than summative assessment. This means that it could not be used to fail poor communicators – the completion rate for Australian medical courses is approximately 97% [5] so it would seem that failing students is not part of the culture. Our theory, however, is that medics would be open to reviewing their own performance with suitable annotations, if they know they are not being overtly rated or criticized. In that review, they are likely to see for themselves how they performed.

So, for instance, we would like the machine analysis to be able to identify particularly important parts of the conversation, describe what happened, and offer some fairly potted suggestions for techniques to use in the future. In a sense, this use of the tool would be on the level of an automated self-help book.

Professional group communication

Hospital wards have shifts. Three times per day, one shift leaves the ward and another replaces them. The outgoing and incoming shifts have a hand-over meeting at which a large amount of information about the patients must be communicated. Furthermore, the medics on the incoming shift might not interact with any given patient for some hours after the hand-over meeting. It is an open question, then, how well information is communicated at the meeting, and also whether or not the transferred information is remembered or used when the medics from the new shift interact with the patients.

This suggests two more uses for machine analysis and review. First, if the hand-over meeting is recorded then, like the training consultations, it can be analyzed and reviewed. In this case, though, it would be for professional review rather than a training exercise for individual medics. Also, because there are more participants and no foreknowledge of patients' conditions, it is a much harder task for a machine to analyze. Second, if the interactions with the patients are also recorded, it may be possible to forensically analyze the communication over time – tracking information back from one medic's interaction with the patient, through the hand-over meeting, to the previous medic's interaction with the patient. For example, one medic might choose a course of action

partly based on something the patient told the previous shift (that may or may not be in the notes).

How to understand communication

There is also a fundamental scientific question. What is communication, and how could a machine tell whether it has or has not happened?

There are plenty of theories about communication in the psychology literature, and plenty of techniques that communication researchers use. There are also many different techniques and tools from the technology community, some of which are regularly used by communication researchers. Leximancer [6] performs unsupervised analysis of conversation transcripts to produce a map of the flow of topics. Latent Dirichlet Allocation [7] and Latent Semantic Analysis [8] are other popular word co-occurrence based analysis techniques. Machine analysis of the sound rather than the transcript can provide further information, for example a measure of the participants' cognitive load [9]. Individually, however, each of those tools usually only tells you about one aspect of the communication, be it the flow of topics, how hard a participant was thinking, or something else.

By bringing together many different analysis techniques, we hope to make it easier for communication researchers to use more of the technical techniques in their research. Ideally, we would even like to identify and validate some objective measures of communication (if such a thing can be identified). We are aware of at least one attempt to measure communication skills by in healthcare, by categorizing individual utterances in a transcript [10]. We hope that by using combinations of technologies

from Natural Language Processing (NLP) and Artificial Intelligence (AI) groups, we will be able to go further. For instance, we would like to capture some of the high level analysis that a communication researcher might perform to analyze a conversation with the tool, and be able to offer a potted automated form of that advice to clinicians. For instance, we may be able to identify common failure modes in a clinical consultation.

A platform for developing new techniques

Just as there is a fundamental scientific question about what communication is, there is also a practical question about how to make analysis and measures meaningful and useful to people. AI and NLP groups regularly develop and improve new techniques for understanding computable aspects of communication. The techniques are fairly complex, however, and the higher level the output, the more likely a user is to ask “how do you know that, and why should I believe you?” If the user is non-technical, that can be a very difficult question to answer in terms that the user understands and feels confident with. So, there is a challenge to finding user-meaningful depictions of scientist-meaningful analysis.

We do not expect to find a general solution for how to explain science to non-scientists. What we do expect is to provide a platform where it is (relatively) easy to plug in new analysis modules, and to create different visualizations based on the analysis. In this way, AI and NLP research groups, or other Human-Computer Interaction researchers, can experiment with different ways of depicting analysis to find out what is effective.

Technical description

To support these use cases, we are developing a video review platform that can support different analysis modules and different visualizations. There are three fundamental parts to this:

1. Repository – a bucket for different kinds of data
2. Workflow – a way of taking a video or audio recording and pass it through the different processing tools
3. Visualization client – which lets users review the result

We have also been building plug-ins that we need for our experiments and use cases (but we hope that other groups will also develop their own).

Repository

Repositories are now fairly common components, and we are currently using Apache Swing, which has a simple RESTful Web Service interface, and allows us to add modules that know how to handle particular kinds of data.

Server-side plug-ins need to take data from the repository, pass it to a module, and insert new data into the repository. For instance, a transcription plug-in takes the video or sound, and returns a transcript. A Leximancer plug-in takes the transcript, formats it for Leximancer, and returns a sequence of topic maps that represent parts of the conversation. Another plug-in might in turn take the Leximancer maps together with the sound, and produce something else. Plug-ins can

be written as scripts, in Ruby, Python, or other languages, or as Java modules.

Workflow

If our first use case, the student-training scenario, is to be deployed widely in teaching hospitals, then the system workflow must be configurable by hospital technical administrators. If the tool is to be useful to communication researchers, then the workflow should be configurable by the researchers (who are not programmers).

Accordingly, the workflow in our system can be configured using common Business Process Management tools. While these are by no means a perfect solution for end-user-programming, they are widely used, make it relatively straightforward to design the sequence of analysis, and it is unlikely we would be able to design a better solution in the scope of this project.

Also, Business Process Management tools have reasonably good support for *human steps* in the process. For example, although we can pass a video to a module for automatic transcription, there are likely to be errors in the transcript produced. *Correcting* the transcript is usually a human step in the workflow.

Visualization client

The visualization client allows visualizations of the different sets of data in the repository to be played together. The core client is fairly small, and most of the functionality comes through plug-ins.

One function of the core client is to reconcile between different timelines for different tracks. Some of the

timeline differences might be purely technical – an audio track from a separate microphone might be out of sync with the main video, or might use a different way of representing time. Some, however, might be more fundamental. For example, one study we plan to run involves reviewing think-aloud recordings of clinicians performing colonoscopies. As two clinicians might move the scope at different rates, it might be more helpful to align the recordings and data by centimeters of bowel rather than by seconds – in which case the audio, transcript, and other data would need to be synchronized to this scheme.

Plug-ins for the client are slightly more technical to write. There is a Java API, and plug-ins are loaded using a standard OSGi interface. Most plug-ins are between four and eight classes in size, depending on the number and complexity of visualizations they provide. Visualizations can be written for Java Swing or JavaFX. Depending on demand, we are considering developing a cut-down HTML5 version and also Silverlight.

Timeframes

During early 2010, we will be conducting experiments using the actor scenario described earlier in this paper, with as many different analysis technologies as we can implement or persuade others to implement. We will be reviewing videos and analysis with communications researchers to identify high-level heuristics with the students. We will also be analyzing colonoscopy videos with researchers to discover, in a user-centred way, how the tool can be made useful for comparative studies.

In 2011, we will be conducting experiments with medical hand-over meetings, and other larger group situations. Towards the end of 2011, we will move into a deployment phase, with a view to incorporating the system into hospital training processes.

Conclusion

The visualization and review tool we are developing is intended to be useful for a number of different groups: medical trainees, medical professionals, communication researchers, and technical researchers. We hope that as researchers you will be interested to use the tool and to develop your own visualizations and plug-ins for new kinds of analysis. Through our very close linkage with the Skills Development Centre of Queensland Health, a major state hospital system in Australia, we also have the opportunity to see the product of our research deployed. If we are successful, and if we can improve the communication training and review practices of medical professionals, we do hope to have a long-term impact on the rates of patient harm (“adverse patient events”) in hospitals. While “saving lives” is an attention-grabbing headline for a poster at a busy conference, it is also meant in earnest.

Acknowledgements

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.

The authors gratefully acknowledge the collaboration of Queensland Health and Leximancer, and the additional funding of NICTA Queensland Research Laboratory by the Queensland Government.

References

- [1] Wilson, R., Runciman, W., Gibberd, R., Harrison, B., Newby, L., and Hamilton, J. The quality in Australian healthcare survey. *The Medical Journal of Australia* (1995), 163:458-471.
- [2] Australian Institute of Health and Welfare. *Sentinel events in Australian public hospitals 2004-05*. Australian Institute of Health & Welfare (2005).
- [3] 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* (2004), 13:i85-i90.
- [4] Zinn, C. 14000 Preventable deaths in Australian hospitals. *British Medical Journal* (1995), 310:1487.
- [5] Australian Council for Education Research. University study in Australia: persistence, completion and beyond. *Longitudinal Surveys of Australian Youth Briefing Number 18*, (2008).
- [6] Booth, K., Maguire, P., and Hillier, V.F. Measurement of communication skills in cancer care: myth or reality? *Journal of Advanced Nursing* 30, 5 (1999), 1073-1079.
- [7] Smith, A.E. and Humphries, M.S. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behaviour Research Methods* 38, 2 (2006).
- [8] Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent Dirichlet allocation. *J of Machine Learning Research*, 3 (2003), 993 – 1022.
- [9] Landauer, T.K. and Dumais, S.T. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104, 2 (1997), 211 – 240.
- [10] Yin, B. and Chen, F. Towards automatic cognitive load measurement from speech analysis. In: *Human-Computer Interaction – HCI Intelligent Multimodal Interaction Environments*, Springer-Verlag (2007)