Queen Mary
University of London

Pump-Priming Research Funding for Cross-Faculty Initiatives 2011-12

Application Form

Name:	Matthew Purver / Rosemarie McCabe	***School / Institute:***	EECS/SMD

Project Summary
(maximum 500 words)

Project Title: Predicting Patient Adherence to Treatment from Consultation Transcripts

This interdisciplinary project will bring together recent research in the School of Medicine & Dentistry (SMD)'s Wolfson Institute of Preventive Medicine with research in the School of Electronic Engineering & Computer Science (EECS)'s Interaction Media & Communication group to develop methods and tools for improving the treatment of schizophrenia. Recent research in schizophrenia in SMD has found that patient communication in clinical consultations predicts adherence to treatment six months later, with the most important aspect being the pro-activity of the patient in checking and correcting their understanding of what the doctor has said (i.e. patient clarification). Research in human interaction in EECS has developed methods for modelling and detecting clarification in dialogue. This project will apply these methods to recorded doctor-patient interactions to produce methods for automatically detecting patient clarification and thus predicting adherence to treatment, producing tools which can aid doctors and help improve patient outcomes. We intend this as the start of a new ongoing interdisciplinary EECS/SMD research collaboration.

The project will run for 6 months; research will be carried out by one EECS-based PDRA, supervised jointly by Purver (EECS) and McCabe (SMD).

Objectives: The objective of this study is to develop automatic methods for predicting patient treatment adherence from doctor-patient transcripts. Past research shows that the presence of particular clarification phenomena in dialogue is a good predictor of treatment adherence, but the detection and categorisation of these phenomena has been done manually by human annotators. This project will apply supervised and/or semi-supervised machine learning techniques to develop a tool which can perform this detection and categorisation automatically, providing a useful resource for medical researchers and doctors. This will be of interest in medicine, where there is wide interest in improving adherence in chronic conditions, and in computational linguistics, where automatic processing of human-human dialogue is an emerging problem area.

Methods: The project will use existing transcript data available from previous SMD research, which has been manually annotated by humans with the clarification categories of interest. Statistical analysis of the transcripts will be used to automatically extract lexical and syntactic features associated with patient clarification; additional pragmatic features will be identified manually and methods for automatic extraction developed. A series of experiments using supervised classification methods (e.g. decision trees, support vector machines or maximum entropy modelling) will then be used to determine feature sets and classification methods for optimum accuracy. Depending on the outcome of these initial experiments, semi-supervised methods (e.g. active learning) may be investigated to improve accuracy and allow faster data annotation; additional features (e.g. prosodic information, lexical semantic classes) together with methods for automatically producing these features (syntactic or semantic parsing, phonetic processing of audio recordings) may be investigated.

Outputs: The final feature extraction and classification software will be released as open-source software, publicly available for use by researchers and clinicians. We also expect this project to seed further EECS/SMD research collaboration, with future research funded by EPSRC, NIHR and MRC.

Queen Mary
University of London

Fit with the aims of one of more of the above three themes
(maximum 300 words)

The proposal fits within the theme *"Delivering Impact - Sponsoring Pathways to Impact"*. The project will provide a way to develop recent interdisciplinary research in EECS and SMD and deliver its benefits to users in the health care sector, both health care professionals and ultimately patients. The project will apply recent research results in language processing to develop automatic methods for predicting the likelihood of patients' adherence to treatment by detecting the presence of patient clarification in doctor-patient transcripts.

In healthcare, there is wide interest in improving adherence in chronic conditions as about half of all medical patients are non-adherent to treatment, which is associated with a significant global burden of disease and high costs to society. The software we will develop to identify patient engagement in clinician-patient communication could be used to identify patients who are likely to be less adherent to treatment. Early interventions with these patients could be helpful in monitoring them more regularly and preventing future non-adherence. Researchers will also benefit from the availability of the software to help identify and investigate levels of patient engagement and their correlation with outcomes; this will seed new collaborative interdisciplinary research between EECS, SMD and/or others.

In computational linguistics, clarification has long been recognised as a key phenomenon in understanding human interaction. However, methods for automatic dialogue understanding and utterance classification have often concentrated on phenomena with higher information content (e.g. questions, statements) but which provide less insight into engagement and coordination, and have consequently shown limited accuracy on coordination and repair phenomena. An investigation into the pragmatic, syntactic and semantic features which characterise these phenomena will therefore allow more accurate dialogue understanding, benefiting researchers and system builders; this will seed new EPSRC-relevant research in internet search, human-human dialogue understanding systems and human-computer interaction

Alignment to which EPSRC Research and / or Mission Landscapes

(maximum 250 words)

The project is aligned with the EPSRC Research Landscape "Information and Communications Technology" and with the Mission Landscapes "Towards Next-Generation Healthcare" and "Digital Economy".

The research carried out will centre around applying and improving Information and Communications Technology techniques, from computational linguistics and machine learning, to medical communication data. The techniques have been developed through EECS research in ICT sponsored by EPSRC and ESRC (DiET EPSRC-EP/D057426/1; DynDial ESRC-RES-062-23-0962).

The project will contribute ot the Digital Economy by producing automatic software for dialogue analysis, a tool which will benefit researchers in computational linguistics and commercial and academic developers of computer dialogue systems. However, its main benefit will be in healthcare: by producing tools which are specifically targeted at detecting key features of clinician-patient consultations, it will provide a resource for predicting likelihood of patient adherence to treatment, allowing doctors to monitor cases likely to have poor outcomes and thus leading to improvements in healthcare.

Timeline and Deliverables

Month 1	processing of existing transcripts to extract baseline lexico-syntactic and pragmatic features
Month 2-3	initial experiments with baseline features, comparing feature sets and learning methods
Month 3-4	extraction of advanced pragmatic, syntactic and semantic features
Month 4-5	extraction of prosodic features
Month 4-5	experiments with advanced feature sets
Month 5-6	implementation and documentation of software

Deliverables:

- open-source software for transcript processing and feature extraction, with documentation
- open-source software for classification and adherence prediction, with documentation
- report detailing feature extraction and prediction methods

Outline Budget

Overall budget £26,571; of which directly incurred staff £21,571, travel & subsistence £1,500 and directly incurred costs £3,500. Project duration 6 months. More details in attached spreadsheet, but a summary follows:

Staff	1 PDRA for 6 months, grade RESEARCH 5, spinal 31	£21,571.00
Travel	1 conference visit for dissemination, inside UK	£500.00
	1 conference visit for dissemination, outside UK	£1,000.00
Other costs	Data recording and transcription	£2,000.00
	Recruitment of PDRA	£1,000.00
Total		*£26,571.00*