The Third Australasian Workshop on
Artificial Intelligence in Health
AIH 2013

Fourth International Workshop on Collaborative Agents --
REsearch and Development (CARE) 2013
"CARE for a Smarter Society"

held in conjunction with the

26th Australasian Joint Conference on Artificial Intelligence (AI 2013)

Tuesday, 3rd December 2013
University of Otago, Dunedin, New Zealand

JOINT WORKSHOP PROCEEDINGS

Editors :
Sankalp Khanna[1], Christian Guttmann[2],
Abdul Sattar[3], David Hansen[1], Fernando Koch[4]

1 The Australian e-Health Research Centre, CSIRO Computational Informatics, Australia
2 IBM Research, Australia
3 Institute for Integrated and Intelligent Systems, Griffith University, Australia
4 Samsung Research Institute, Brazil

AIH 2013 - ACKNOWLEDGEMENTS

Program Chairs

- Abdul Sattar (Griffith University, Australia)
- David Hansen (CSIRO Australian e-Health Research Centre, Australia)

Workshop Chair

- Sankalp Khanna (CSIRO Australian e-Health Research Centre, Australia)

Senior Program Committee

- Aditya Ghose (University of Newcastle, Australia)
- Jim Warren (University of Auckland, New Zealand)
- Wayne Wobcke (University of New South Wales, Australia)
- Mehmet Orgun (Macquarie University, Australia)
- Yogesan (Yogi) Kanagasingam (CSIRO Australian e-Health Research Centre, Australia)

Program Committee

- Simon McBride (CSIRO AEHRC)
- Adam Dunn (University of New South Wales)
- Stephen Anthony (University of New South Wales)
- Lawrence Cavedon (Royal Melbourne Institute of Technology / NICTA)
- Diego Mollá Aliod (Macquarie University)
- Michael Lawley (CSIRO AEHRC)
- Anthony Nguyen (CSIRO AEHRC)
- Amol Wagholikar (CSIRO AEHRC)
- Bevan Koopman (CSIRO AEHRC)
- Kewen Wang (Griffith University)
- Vladimir Estivill-Castro (Griffith University)
- John Thornton (Griffith University)
- Bela Stantic (Griffith University)
- Byeong-Ho Kang (University of Tasmania)
- Justin Boyle (CSIRO AEHRC)
- Guido Zuccon (CSIRO AEHRC)
- Hugo Leroux(CSIRO AEHRC)
- Alejandro Metke (CSIRO AEHRC)

Key Sponsors

- CSIRO Australian e-Health Research Centre
- Institute for Integrated and Intelligent Systems, Griffith University

Supporting Organisations

- The Australasian College of Health Informatics
- The Australasian Medical Journal
- The Australasian Telehealth Society

CARE 2013 - ACKNOWLEDGEMENTS

General Chairs

- Christian Guttmann (IBM Research -- Australia)
- Fernando Koch (Samsung Research Institute -- Brazil)

ABSEES Track Chairs:

- Maryam Purvis, maryam.purvis@otago.ac.nz
- Takayuki Ito, ito.takayuki@nitech.ac.jp

Program Committee

- Andrew Koster
- Anthony Patricia
- Bastin Tony Roy Savarimuthu
- Benjamin Hirsch
- Carlos Cardonha
- Cristiano Castelfranchi
- David Morley
- Diego Gallo
- Frank Dignum
- Franziska Klügl
- Gordon McCalla
- Ingo J. Timm
- Inon Zuckerman
- Jaime Sichman
- Kobi Gal
- Lars Braubach
- Lawrence Cavedon
- Leonardo Garrido
- Liz Sonenberg
- Magnus Boman
- Marcelo Ribeiro
- Martin Purvis
- Meritxell Vinyals
- Michael Thielscher
- Neil Yorke-Smith
- Priscilla Avegliano
- Rainer Unland
- Ryo Kanamori
- Sankalp Khanna
- Sarvapali Ramchurn
- Sascha Ossowski
- Shantanu Chakraborty
- Sherief Abdallah
- Simon Thompson
- Simon Goss
- Toby Walsh
- Wayne Wobcke
- Wei Chen
- Zakaria Maamar

TABLE OF CONTENTS

Health Informatics and Artificial Intelligence solutions: Addressing the Challenges at the Frontiers of Modern Healthcare

Keynote Address

Professor Michael Blumenstein

Griffith University, Australia
m.blumenstein@griffith.edu.au

Speaker Profile

Michael Blumenstein is a Professor and Head of the School of Information and Communication Technology at Griffith University, where he previously served as the Dean (Research) in the Science, Environment, Engineering and Technology Group. In addition, Michael currently serves as the Leader for the Health Informatics Flagship Program at the Institute for Integrated and Intelligent Systems.

Michael is a nationally and internationally recognised expert in the areas of automated Pattern Recognition and Artificial Intelligence, and his current research interests include Document Analysis, Multi-Script Handwriting Recognition and Signature Verification. He has published over 132 papers in refereed books, conferences and journals. His research also spans various projects applying Artificial Intelligence to the fields of Engineering, Environmental Science, Neurobiology, Coastal Management and Health. Michael has secured internal/nationally competitive research grants to undertake these projects with funds exceeding AUD$4.3 Million. Components of his research into the predictive assessment of beach conditions have been commercialised for use by local government agencies, coastal management authorities and in commercial applications.

Following his achievements in applying Artificial Intelligence to the area of bridge engineering (where he has published widely and has been awarded federal funding), he was invited to serve on the International Association for Bridge and Structural Engineering's Working Commission 6 to advise on matters pertaining to Information Technology. Michael is the first Australian to be elected onto this committee. In addition, he was previously the Chair of the Queensland Branch of the Institute for Electrical and Electronic Engineers (IEEE) Computational Intelligence Society. He is also the Gold Coast Chapter Convener and a Board Member of the Australian

Computer Society's Queensland Branch Executive Committee as well as the Chairman of the IT Forum Gold Coast and a Board Member of IT Queensland. Michael currently serves on the Australian Research Council's (ARC) College of Experts on the Engineering, Mathematics and Informatics (EMI) panel. In addition, he has recently been elected onto the Executive of the Australian Council of Deans of Information and Communication Technology (ACDICT). Michael also serves on a number of Journal Editorial Boards and has been invited to act as General Chair, Organising Chair, Program Chair and/or Committee member for numerous national/international conferences in his areas of expertise.

In 2009 Michael was named as one of Australia’s Top 10 Emerging Leaders in Innovation in the Australian’s Top 100 Emerging Leaders Series supported by Microsoft. Michael is a Fellow of the Australian Computer Society and a Senior Member of the IEEE.

Abstract

Numerous challenges currently exist in the Health Sector such as effective treatment of patients with chronic diseases, early diagnosis and prediction of health conditions, patient data administration and adoption of electronic health records, strategic planning for hospitals and engagement of health professionals in training. This presentation focuses on these challenges and examines some innovative Health Informatics solutions with prospective deployment of automated artificial intelligence tools to augment current practices.

Some challenges are examined at a brand new University Hospital in Queensland, whereby a number of automated solutions are investigated using technology and intelligent approaches such as mobile devices for understanding patient chronic health conditions over time, image analysis and pattern recognition for the early diagnosis and treatment of such brain disorders as Parkinson's disease, social media analytics for patient engagement in the adoption of electronic health records, on-line collaborative tools for strategic planning in the hospital and the use of 3D virtual worlds for realistic training and professional development for medical staff.

Finally, the presentation will conclude with a discussion about the emerging "Research Triangle" present at the Gold Coast, in Queensland, which includes the new Gold Coast University Hospital and is directly adjacent to Griffith University's Gold Coast campus with proximity to the emerging Health and Knowledge Precinct. This special zone presents a unique opportunity to nurture cutting edge health-related research intersecting information technology in collaboration with industry and government, which may have a profound impact on the future landscape of Health Informatics innovation in the region.

Classification Models in Intensive Care Outcome Prediction-can we improve on current models?

Nicholas A. Barnes,

Intensive Care Unit,
Waikato Hospital, Hamilton, New Zealand.

Lynnette A. Hunt,

Department of Statistics,
University of Waikato, Hamilton, New Zealand.

Michael M. Mayo,

Department of Computer Science,
University of Waikato, Hamilton, New Zealand.

Corresponding Author: Nicholas A. Barnes.

Abstract
Classification models (“machine learners” or “learners”) were developed using machine learning techniques to predict mortality at discharge from an intensive care unit (ICU) and evaluated based on a large training data set from a single ICU. The best models were tested on data on subsequent patient admissions. Excellent model performance (AUCROC (area under the receiver operating curve) =0.896 on a test set), possibly superior to a widely used existing model based on conventional logistic regression models was obtained, with fewer per-patient data than that model.

1 Introduction

Intensive care clinicians use explicit judgement and heuristics to formulate prognoses as soon as reasonable after patient referral and admission to an intensive care unit [1].

Models to predict outcome in such patients have been in use for over 30 years [2] but are considered to have insufficient discriminatory power for individual decision making in a situation where patient variables that are difficult or impossible to measure may be relevant. Indeed even variables that have little or nothing to do with the patient directly (such as bed availability or staffing levels [3]) may be important in determining outcome.

There are further challenges for model development. Any model used should be able to deal with the problem of class imbalance, which refers in this case to the fact that mortality should be much less common than survival. Many patient data are probably only loosely or indeed not related to outcome and many are highly correlated. For example, elevated measurements of serum urea, creatinine, urine output, diagnosis of renal failure and use of dialysis will all be closely correlated.

Nevertheless, models are used to risk adjust for comparison within an institution over time or between institutions, and model performance is obviously important if this is to be meaningful. It is also likely that a model with excellent performance could augment clinical assessment of prognosis. Furthermore, a model that performs well while requiring fewer data would be helpful as accurate data acquisition is an expensive task.

The APACHE III-J (Acute Physiology and Chronic Health Evaluation revision III-J [4]) model is used extensively within Australasia by the Centre for Outcomes Research of the Australian and New Zealand Intensive Care Society (ANZICS) and a good understanding of its local performance is available in the published literature [4]. It should be noted that death at hospital discharge is the outcome variable usually considered by these models. Unfortunately the coefficients for all variables for this model are no longer in the public domain so direct comparison with new models is difficult. The APACHE (Acute Physiology and Chronic Health Evaluation) models are based largely on baseline demographic and illness data and physiological measurements taken within the first day after ICU admission.

This study aims to explore machine learning methods that may outperform the logistic regression models that have previously been used.

The reader may like to consult a useful introduction to the concepts and practice of machine learning [5] if terms or concepts are unfamiliar.

2 Methods

The study is comprised of three parts:

1. An empirical exploration of raw and processed admission data with a variety of attribute selection methods, filters, base classifiers and metalearning techniques (which are overarching models that have other methods nested within them) that were felt to be suitable to develop the best classification models. Metamodels and base classifiers may be nested within other metamodels and learning schemes can be varied in very many ways .These experiments are represented below in Figure 1 where we used up to two metaclassifiers with up to two base classifiers nested within a metaclassifier.

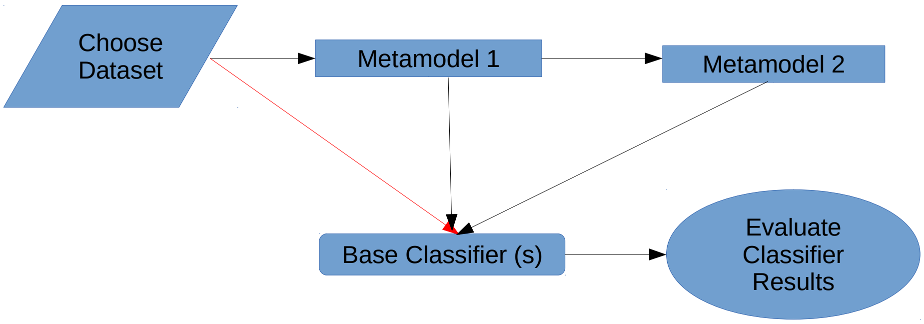


Fig. 1. Schematic of phase 1 experiments. Different color arrows indicate that one or more metamodels and base classifiers may optionally be combined in multiple different ways. One or more base classifiers are always required.

2. Further testing with the best performing data set (full unimputed training set) and learners with manual hyperparameter setting. A hyperparameter is a particular model configuration that is selected by the user, either manually or following an automatic tuning process. This is represented in a schematic below:

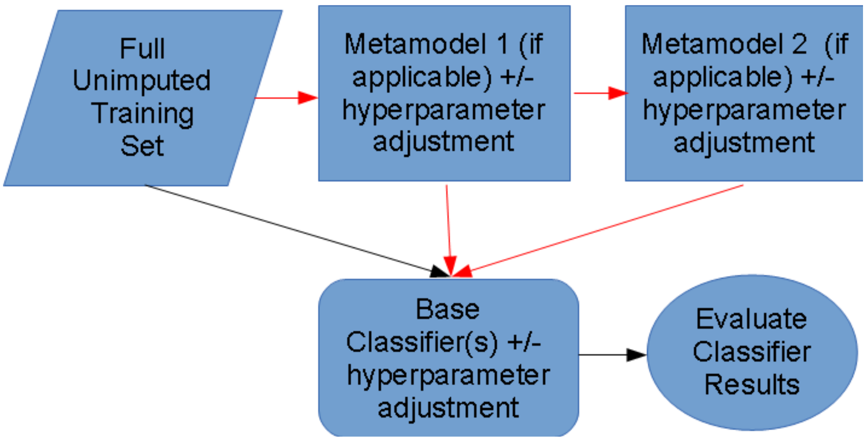


Fig. 2. Schematic of phase 2 experiments. As in phase 1, one or more metamodels may be optionally combined with one or more base classifiers.

3. Testing of the best models from phase 2 above on a new set of test data to better understand generalizability of the models. This is depicted in Figure 3 below.

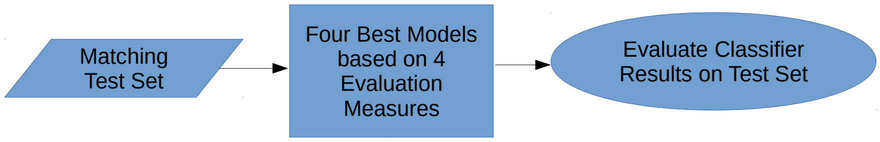


Fig. 3. Schematic of phase 3

The training data for adult patients (8122 patients over 16 years of age) were obtained from the database of a multidisciplinary ICU in a tertiary referral centre from a period between July 2004 and July 2012.Data extracted were comprised of a demographic variable (age), diagnostic category (with diagnostic coefficient from the APACHE III-J scoring system, including ANZICS modifications), and an extensive list of numeric variables relating to patient physiology and composite scores based on these, along with the classification variable: either survival, or alternatively, death at ICU discharge (as opposed to death at hospital discharge as in the APACHE models). Much of the data collected is used in APACHE III-J model mentioned above, and represents a subset of the data used in that model. Training data, prior to the imputation process, but following discretization of selected variables are represented in Table 1. Test data for the identical variable set were obtained from the same database for the period July 2012 to March 2013.

Of particular interest is that the data is clearly class imbalanced with mortality during ICU stay of approximately 12%. This has important implications for modelling the data.

There were many strongly correlated attributes within the data sets. Many of the model variables are collected as highest and lowest measures within twenty four hours of admission to the ICU. Correlated variables may bring special problems with conventional modelling including logistic regression. The extent of correlation is demonstrated in Figure 4.

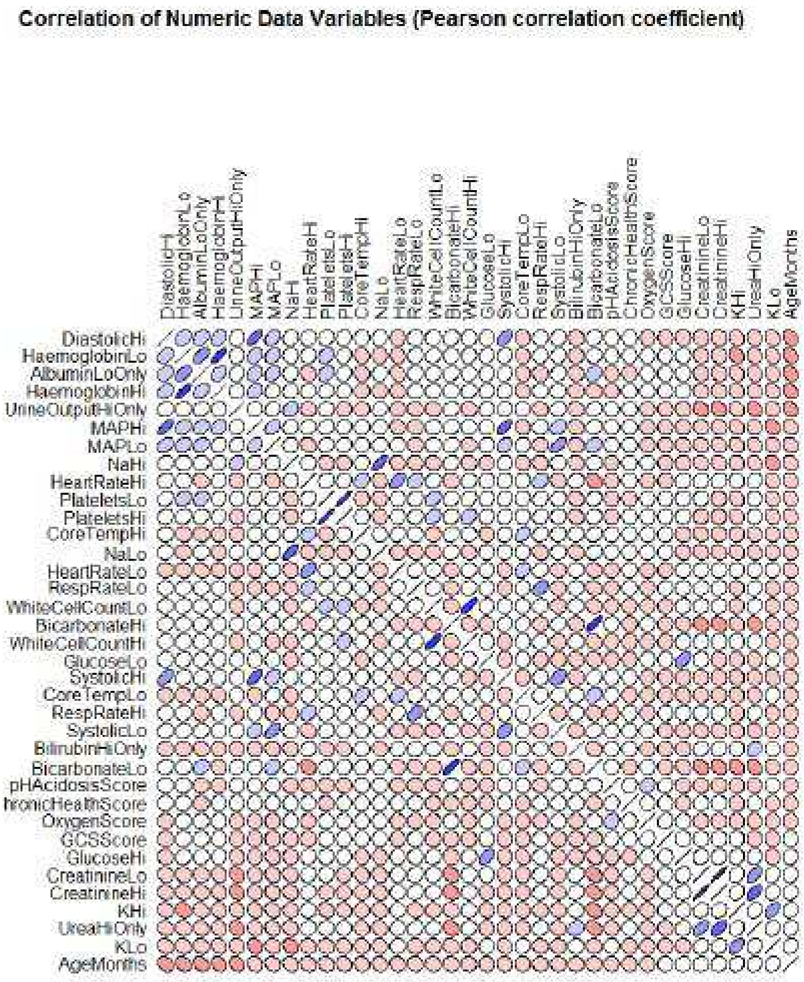


Fig. 4. Pearson correlations between variables are shown using colour. Blue colouration indicates positive correlation. Red colouration indicates negative correlation. The flatter the ellipse, the higher the correlation. White circles indicate no significant correlation between variables.

Patterns of missing data are indicated in Table 1 and represented graphically in Figure 5.

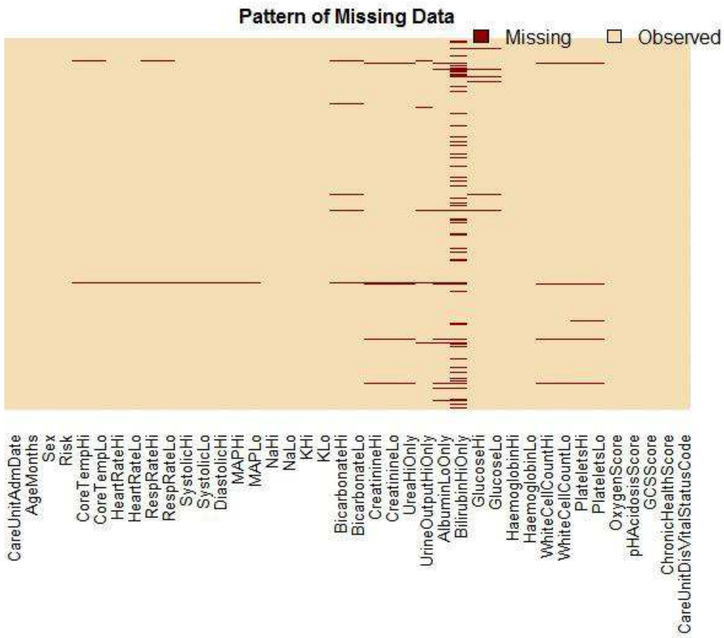


Fig. 5. Patterns of missing data in the raw training set. Missing data is represented by red colouration.

Missing numeric data in the training set was imputed using multiple imputation with the R program [6] and the R package Amelia [7], which utilises bootstrapping of non-missing data followed by imputation by expectation maximisation. We initially used the average of five multiple imputation runs.

Using the last imputed set was also trialled, as it may be expected to be the most accurate based on the iterative nature of the Amelia algorithm. No categorical data were missing. Date of admission was discretized to the year of admission, age was converted to months of age, and the diagnostic categories were converted to five to eight (depending on study phase) ordinal risk categories by using coefficients from the existing APACHE III-J risk model.

A summary of data is presented below in Table 1.

Table 1. Data Structure

Variable	Type	Missing	Distinct values	Min.	Max.
CareUnitAdmDate	numeric	0	9	2004	2012
AgeMonths	numeric	0	880	192	1125
Sex	pure factor	0	2	F	M
Risk	pure factor	0	8	Vlow	High

CoreTempHi	numeric	50	89	29	42.3
CoreTempLo	numeric	53	102	25.2	40.7
HeartRateHi	numeric	25	141	38.5	210
HeartRateLo	numeric	26	121	0	152
RespRateHi	numeric	38	60	8	80
RespRateLo	numeric	40	42	2	37
SystolicHi	numeric	27	161	24	288
SystolicLo	numeric	55	151	11	260
DiastolicHi	numeric	27	105	19	159
MAPHi	numeric	28	124	20	200
MAPLo	numeric	43	103	3	176
NaHi	numeric	46	240	112	193
NaLo	numeric	51	245	101	162
KHi	numeric	46	348	2.7	11.7
KLo	numeric	51	275	1.4	9.9
BicarbonateHi	numeric	218	322	3.57	48
BicarbonateLo	numeric	221	319	2	44.2
CreatinineHi	numeric	130	606	10.2	2025
CreatinineLo	numeric	134	552	10	2025
UreaHiOnly	numeric	232	433	1	99
UrineOutputHiOnly	numeric	184	3501	0	15720
AlbuminLoOnly	numeric	281	66	5	65
BilirubinHiOnly	numeric	1579	183	0.4	618
GlucoseHi	numeric	172	255	1.95	87.7
GlucoseLo	numeric	177	198	0.1	60
HaemoglobinHi	numeric	54	153	1.8	25
HaemoglobinLo	numeric	59	151	1.1	25
WhiteCellCountHi	numeric	131	470	0.1	293
WhiteCellCountLo	numeric	135	393	0.08	293
PlateletsHi	numeric	149	653	7	1448
PlateletsLo	numeric	153	621	0.27	1405
OxygenScore	numeric	0	8	0	15
pHAcidosisScore	numeric	0	9	0	12
GCSScore	numeric	0	11	0	48
ChronicHealthScore	numeric	0	6	0	16
Status at ICU Discharge	pure factor	0	2	A	D

Phase 1 consisted of an exploration of machine learning techniques thought suitable to this classification problem, and in particular those thought to be appropriate to a class imbalanced data set. Attribute selection, examining the effect of using imputed and unimputed data sets and application of a variety of base learners and metaclassifiers without major hyperparameter variation occurred in this phase. The importance of attributes was examined in multiple ways including using random forest methodology for variable selection, using improvement in Gini index using particular attributes. This information is displayed in figure 6.

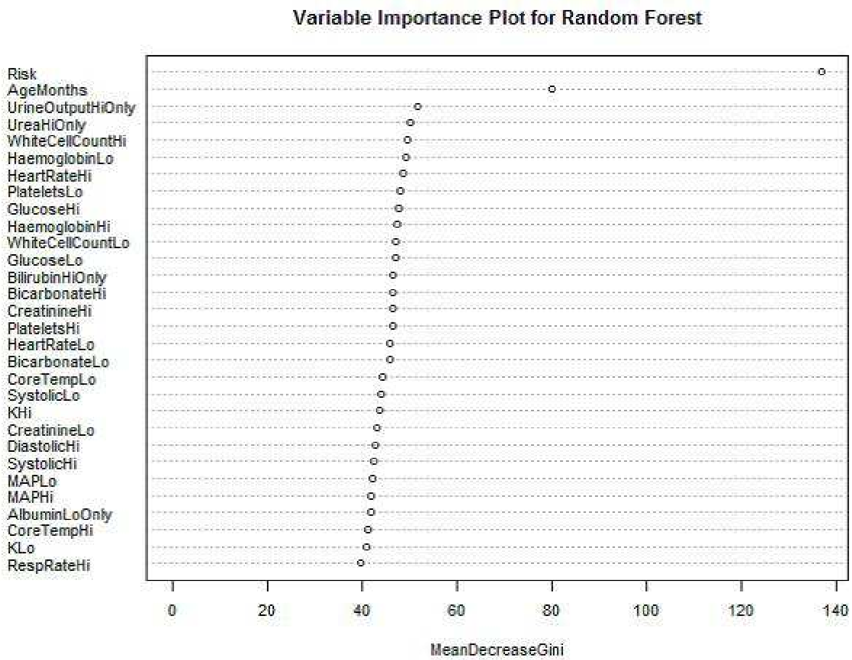


Fig. 6. Variable importance as measured by Gini index using random forest methodology. A substantial decrease in Gini index indicates better classification with variable inclusion. Variables used in the study are ranked by their contribution to Gini index.

A comprehensive evaluation of all techniques is nearly impossible given the enormous variety of techniques and the ability to combine up to several of these at a time in any particular model. Techniques were chosen based on the likely success of their application. WEKA [8] was used to apply learners and all models were evaluated with tenfold cross validation. WEKA default settings were commonly used in phase 1 and the details of these defaults are widely available [9]. Unless otherwise stated all settings in all study phases were the default settings of WEKA for each classifier or filter. Two results were used to judge overall model performance during phase 1. These were:

1. Area under the receiver operating curve (AUC ROC)
2. Area under the precision recall curve (AUC PRC)

The results are presented in Table 3 in the results section.

Phase 2 of our study involved training and evaluation on the same data sets with learners that had performed well in phase 1. Hyperparameters were mostly selected manually, as automatic hyperparameter selection in any software is limited and hampered by a lack of explicitness. Class imbalance issues were addressed with appropriate WEKA filters (spread subsample and SMOTE, a filter which generates a synthetic data set to balance the classes [10]), or the use of cost sensitive learners [11]. Unless otherwise stated in Table 3, WEKA default settings were used for each filter or classifier. Evaluation of these models proceeded with tenfold cross-validation and the results were examined in light of four measures:

1. Area under the receiver operating curve with 95% confidence intervals by the method of Hanley and McNeill [12]
2. Area under the precision recall curve
3. Matthews correlation coefficient and,
4. F-measure

Additionally, scaling the quantitative variables by standardizing or normalizing the data was explored as this is known to sometimes improve model performance [13].
The results of phase 2 are presented in Table 2 in the results section.
Phase 3 involved evaluating the accuracy of the best classification models from phase 2 on a new test set of 813 patient admissions. Missing data in the test set were not imputed. Results are shown in Table 3.

3 Results

Table 2 presents the results following tenfold cross validation on a variety of techniques thought suitable for trial in the modelling problem. These are listed in order of descending area under the curve of the receiver operating curve and the area under the precision recall curve is also presented.

Table 2. Phase 2 of study.

Data	Preprocess	Meta Model 1	Meta model 2	Meta model 3	Base classifier 1	Base classifier 2	ROC	PRC
Unimputed all variables	NA	Cost Sensitive Classifier matrix 0,5;1,0	NA	NA	Random Forest 500 trees	NA	0.895	0.629
Unimputed all variables	NA	Cost Sensitive Classifier matrix 0,5;1,0	NA	NA	Random Forest 200 trees	NA	0.894	0.416
Unimputed all variables	NA	Cost Sensitive Classifier matrix 0,5;1,0	NA	NA	Naïve Bayes	NA	0.864	0.418

Unimputed all variables	Spread-subsample uniform	Filtered Classifier	Attribute selected classifier 20 variables selected on info. Gain and ranked	Vote	J4.8 tree	Naïve Bayes	0.854	0.439
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Logistic regression	NA	Logistic Regression	NA	0.766	0.283
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	NA	NA	SimpleLogis-tic	NA	0.766	0.28
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Random Comm	NA	REP tree	NA	0.753	0.259
Imputed ten variables	NA	Filtered Classifier	NA	NA	Naïve Bayes	NA	0.742	0.248
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Adaboost M1	NA	J48	NA	0.741	0.254
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Vote	NA	Random Forest 10 trees	Naïve Bayes	0.741	0.252
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Bagging	NA	J48	NA	0.736	0.258
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	Decorate	NA	Naïve Bayes	NA	0.735	0.238
Imputed all variables	Spread-subsample uniform	Filtered Classifier	Attribute selected classifier 20 variables selected on info. Gain and ranked	Vote	J4.8 tree	Naïve Bayes	0.735	0.238
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	NA	NA	J4.8 tree	NA	0.734	0.234
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	NA	NA	Random Forest 10 trees	NA	0.713	0.221
Imputed ten variables	Spread-subsample uniform	Filtered Classifier	SMO	NA	SMO	NA	0.5	0.117

ROC-area under receiver operating characteristic curve
CI-confidence interval
PRC-area under precision-recall curve
NA-not applicable

Table 3 presents the results of tenfold cross validation on the best models from phase 1 trained on the training set in phase 2 of our study. Models are listed in descending order of AUC ROC. The data set used in the modelling is indicated, along with any pre-processing of data, base learners, metalearners if applicable, and other evaluation tools as listed in the methods section above. The model which performs best of all models on any of the four classification methods is shaded in red to emphasise that no one performance measure dominates a classifier's overall utility.

Table 3. Phase 2 results

Preprocess	Metamodel1	Metamodel2	Base Model 1	Base model 2	ROC	ROC 95% CI's	PRC	MCC	F-measure
Spread subsample uniform	Filtered classifier	Rotation forest 100 iterations	Alternating decision tree 100 iterations	NA	0.903	(0.892,0.912)	0.622	0.47	0.51
NA	Cost sensitive classifier 0,5;1,0	NA	Rotationforest 500 iterations	J 48	0.901	(0.881,0.921)	0.625	0.482	0.481
Spread subsample uniform	Filtered classifier	Rotationforest 200 iterations	NA	J 48	0.897	(0.888,0.906)	0.606	0.452	0.494
Spread subsample uniform	Filtered classifier	NA	Rotationforest 500 iterations	J 48	0.897	(0.888,0.906)	0.608	0.45	0.493
Spread subsample uniform	Filtered classifier	NA	Rotation forest 500 iterations	J48 graft	0.897	(0.888,0.906)	0.611	0.456	0.5
Spread subsample uniform	Filtered classifier	Rotation forest 50 iterations	Alternating decision tree 50 iterations	NA	0.896	(0.887,0.905)	0.608	0.452	0.495
Spread subsample uniform	Filtered classifier	NA	Rotation forest 100 iterations	J 48	0.895	(0.886,0.904)	0.602	0.443	0.488
NA	Cost sensitive classifier 0,5;1,0	NA	Random forests (RF) 1000 trees 2 features each tree	NA	0.893	(0.879,0.907)	0.599	0.506	0.561
NA	Cost sensitive classifier 0,5;1,0	NA	RF 500 trees 2 features each tree	NA	0.892	(0.878.0.906)	0.598	0.511	0.567
NA	Cost sensitive classifier 0,1;1,0	NA	RF 500 trees 2 features each tree	NA	0.891	(0.867,0.915)	0.602	0.416	0.398
NA	Cost sensitive classifier 0,1;1,0	NA	RF 1000 trees 2 features each tree	NA	0.891	(0.867,0.915)	0.603	0.422	0.391
NA	Cost sensitive classifier 0,10;1,0	NA	RF 500 trees 2 features each tree	NA	0.891	(0.878,0.904)	0.594	0.497	0.558
NA	Cost sensitive classifier 0,5;1,0	NA	Rotation Forest 50 iterations	J48	0.891	(0.871,0.911)	0.606	0.479	0.485
Spread subsample	Filtered classifier	Bagging 150 iterations	J 48 C 0.25 M 2	NA	0.89	(0.869,0.911)	0.609	0.474	0.471
Spread subsample	Filtered classifier	Bagging 200 iterations	J 48 C 0.25 M 3	NA	0.889	(0.868,0.910)	0.61	0.474	0.473
NA	Cost sensitive classifier 0,1;1,1	NA	RF 200 trees 2 features each tree	NA	0.889	(0.865,0.913)	0.598	0.425	0.395
Spread subsample	Filtered classifier	Bagging 100 iterations	J 48 C 0.25 M 2	NA	0.888	(0.867,0.909)	0.605	0.47	0.467

NA	Cost sensitive classifier 0,5;1,0	NA	RF 100 trees 2 features each tree	NA	0.888	(0.864,0.912)	0.594	0.42	0.396
Spread subsample uniform	Filtered classifier	NA	Random committee 500 iterations	Random tree	0.887	(0.879,0.895)	0.578	0.373	0.409
Spread subsample	Filtered classifier	Adaboost M1 150 iterations	J 48 C 0.25 M 2	NA	0.886	(0.865,0.907)	0.584	0.48	0.476
Spread subsample	Filtered classifier	Adaboost M1 100 iterations	J 48 C 0.25 M 2	NA	0.884	(0.863,0.905)	0.577	0.469	0.467
Spread subsample	Filtered classifier	Bagging 50 iterations	J 48 C 0.25 M 2	NA	0.883	(0.862,0.904)	0.597	0.465	0.465
Spread subsample uniform	Filtered classifier	NA	Random subspace 100 iterations	REP tree	0.877	(0.868,0.886)	0.563	0.423	0.473
Spread subsample uniform	Filtered classifier	NA	Multiboost AB 50 iterations	J 48	0.874	(0.864,0.884)	0.428	0.435	0.482

RF-random forest
REP-representative
NA-not applicable
MCC-Matthews correlation coefficient

Normalizing or standardizing the data did not improve model performance and indeed tended to moderately worsen it.

Table 4 presents the results of applying four of the best models from phase 2 on a test data set of 813 patient admissions which should be from the same population distribution (if date of admission is not a relevant attribute). Evaluation is based on AUC ROC, AUC PRC, Matthews's correlation coefficient and F-measure. These evaluations were obtained by WEKA's knowledge flow interface.

Table 4. Model results with new test set in Phase 3

Data preprocessing	Metamodel 1	Metamodel 2	Base Classifer 1	Base Classifier 2	ROC	95% CI ROC	PRC	MCC	F-meas
Spread subsample uniform	Filtered classifier	Rotation forest 100 iterations	Alternating decision tree 100 iterations	NA	0.896	(0.854,0 .938)	0.592	0.401	0.426
Spread subsample uniform	Filtered classifier	Rotation forest 200 iterations	NA	J 48	0.893	(0.863,0 .923)	0.571	0.525	0.534
NA	Cost sensitive classifier 0,5;1,0	NA	Rotation forest 500 iterations	J 48	0.887	(0.821,0 .953)	0.561	0.386	0.411
NA	Cost sensitive classifier 0,5;1,0	NA	Random forest 500 trees, 2 features each tree	NA	0.885	(0.855,0 .915)	0.551	0.51	0.555

ROC-area under receiver operating characteristic curve
CI-confidence interval
PRC-area under precision-recall curve
MCC-Matthews correlation coefficient
F-meas-F-measure

4 Discussion

It is unrealistic to expect models to perfectly represent such a complex reality as that of survival from critical illness. Perfect classification is impossible because of the limitations of any combination of currently available measurements made on such patients to accurately reflect survival potential. Patient factors such as attitudes towards artificial support and presumably health practitioner and institution related factors are important. Additionally non-patient related factors which may be purely logistical will continue to thwart perfect prediction by any future model. For instance, a patient may die soon after discharge from the ICU if a ward bed is available and conversely will die within the ICU if a ward bed is not available and transfer cannot proceed. Models currently employed generally consider death at hospital discharge, but new factors that increase randomness can enter in the hospital stay following ICU discharge, so problems are not necessarily decreased with this approach.

The best models we have studied have excellent performance when evaluated following tenfold cross validation in the single ICU setting with use of fewer data points than the current gold standard model. Machine learning techniques usually make few distributional assumptions about the data when compared with the traditional logistic regression model. Missing data are often dealt with effectively with machine learning techniques while complete cases are generally used in traditional general linear modelling such as logistic regression. Clinical data will never be complete, as some data will not be required for a given patient, while some patients may die prior to collection of data which cannot subsequently be obtained. Imputation may be performed on data prior to modelling but has limitations. It is interesting that models trained on unimputed data tend to perform better than imputed data, both in phase 2 and with the test set in phase 3.

The best comparison we can make in the published literature is the work of Paul et al [4] which demonstrates that the AUC ROC of the APACHE-III-J model has varied between 0.879 and 0.890 when applied to over half a million adult admissions to Australasian ICUs between 2000 and 2009. Routine exclusions in this study included readmissions, transfers to other ICUs, and missing outcome and other data, and admission post coronary artery bypass grafting prior to introduction of the ANZICS modification to APACHE-III-J for this category. None of these were exclusions in our study. The Paul et al paper looks at outcome at hospital discharge, while ours examines outcome at ICU discharge. For these reasons the results are not directly comparable but our results for AUC ROC of up to 0.896 on a separate validation set clearly demonstrate excellent model performance.

The techniques associated with the best performance involve addressing class imbalance (i.e. pre-processing data to create a dataset with similar numbers of those who survive and those that die). This class imbalance is a well-known problem in classification. Mortality data from any healthcare setting tend to be class imbalanced. Our study shows that any approach to class imbalance in the data greatly enhance model performance. Cost sensitive metalearners [11], synthetic minority generation techniques (SMOTE [10]) and creating a uniform class distribution by subsampling across the data all improve model performance.

A cost sensitive learner indicates a technique that reweights cases according to a cost matrix that the user sets to reflect differing "cost" of misclassification of positive and negative cases. This intuitively lends itself to the intensive care treatment process where such a framework is likely implemented at least subconsciously by the intensive care clinician. For instance the cost of clinically "misclassifying" a patient may be substantial and clinicians would likely try hard to avoid this situation.

In our study, the ensemble learner random forests [14] with or without a technique to address class imbalance tends to outperform many more complex metalearners, or enhancements of single base classifiers such as bagging [15] and boosting [16]. Random forests involve generation of many different tree models, each of which splits the cases based on different variables and a criterion to increase information gain. Voting then occurs across the "forest" to decide on the best way to split the cases and this produces the model. The term ensemble simply represents the fact that multiple learners are involved, rather than a single tree. As many as 500 or 1000 trees are commonly required before the error of the forest is at a minimum. The number of variables to be considered by each tree may also be set to try and improve performance. The other techniques that produced excellent results were rotation forests either alone, with a cost sensitive classifier, or in combination with a technique known as alternating decision tree. Alternating decision tree takes a "weak" classifier (such as a tree classifier) and uses a technique similar to boosting to improve performance.

The reason extensive experimentation may be required to produce the best model is attributed to Wolpert [17] and described as the "no free lunch theorem", meaning that there is no one single technique that will model the best in every given scenario. Of course the same is true of any conventional statistical technique applied to multidimensional problems. Data processing and model selection are crucial to performance although if prediction alone is important, a pragmatic approach can be taken to the usual statistical assumptions. Machine learning techniques are generally not a "black box" approach however and deserve the same credibility as any older method, if application is appropriate.

Similarly, no single evaluation measure can summarize a classifier's performance and different model strengths and weaknesses may be more or less tolerable depending on the circumstances of model use and hence a range of measures are usually presented as we have done.

There are several weaknesses to our study. It is clearly from a single centre and may not generalize to other ICUs in other healthcare systems. Mortality remains a crude measure of ICU performance but remains simple to measure and of great relevance nevertheless. The existing gold standard models usually measure classification of survival or death at hospital discharge, so are not necessarily directly comparable to our models which measures survival or death at ICU discharge.

We are unable to directly compare our models with what may be considered gold standards as some of these (e.g. APACHE IV) are only commercially available, and as mentioned before, even the details of APACHE-III-J are not in the public domain. The best comparison involving Australasian data using APACHE-III-J comes from the paper of Paul et al. [4] but as with all APACHE models, this predicts death at hospital discharge. Additionally, re-admissions were excluded which may be a significant factor beyond what are often relatively small numbers of re-admissions in any given ICU, as re-admissions suffer a disproportionately high mortality.

Exploration of the available hyperparameters of the many models examined has been relatively limited. The ability to do this automatically, and explicitly or in a reproducible way in WEKA and indeed any available software is limited although this may be changing [18]. Yet minor changes to these hyperparameters may produce meaningful enhancements in model performance. Tuning hyperparameters runs the risk of overfitting a model, but we have tried to guard against this by testing the data on a separate validation set.

Likewise, the ability to combine models with the best characteristics [19], which is becoming more common in prediction of continuous variables [20] is not yet easily performed with the available software.

We have not examined the calibration of our models. Good calibration is not required for accurate classification. Accurate performance across all risk categories is highly desirable in a model. Similarly, performance including calibration for different diagnostic categories that may become more significant in an ICU's case mix is not accounted for.

Modelling using imputed data in every phase of our study tends to show inconsistent or suboptimal performance. It may be that imputation could be applied more accurately by another approach that would improve model performance.

The major current use of these scores is in quality improvement activities. Once a score is developed which accurately quantitates risk, the expected number of deaths may be compared to those observed [21]. The exact risk for a given integer valued number of deaths may be derived from the Poisson binomial distribution and compared to the number observed [22]. A variety of risk adjusted control charts can be constructed with confidence intervals [23].

5 Conclusions

We have presented alternative approaches to the classification problem involving prediction of mortality at ICU discharge using machine learning techniques. Such techniques may hold substantial advantage over traditional logistic regression approaches and should be considered to replace these. Complete clinical data may be unnecessary when using machine learning techniques, and in any case are frequently not available. Out of the techniques studied, random forests seems to be the modelling approach with the best performance and has an advantage that it is relatively easy to conceptualise and implement with open source software. During model training a method to address class imbalance should be used.

6 Bibliography

[1]. Downar, J. (2013, April 18). Even without our biases, the outlook for prognostication is grim. Available from ccforum: http://ccforum.com/content/13/4/168

[2]. Knaus WA, W. D. (1981). APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med, 591-597.

[3]. Tucker, J. (2002). Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. Lancet, 99-107.

[4]. Paul, E., Bailey, M., Van Lint, A., & Pilcher, D. (2012). Performance of APACHE III over time in Australia and New Zealand: a retrospective cohort study. Anaesthesia and Intensive Care, 980-994.

[5]. Domingos, P. (2013, May 6). A few useful things to know about machine learning. Available from Washington University: http://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

[6]. R Core Team. (2013, April 25). Available from CRAN: http://www.R-project.org/.

[7]. Honaker, J., King, G., & Blackwell, M. (2013, April 25). Amelia II: a program for missing data. Available from Journal of Statistical Software: http://www.jstatsoft.org/v45/i07/.

[8]. Hall, M., Eibe, F., Holmes, G., Pfahringer, B., & Reutemann, P. (2009, 1). The WEKA Data Mining Software: An Update. SIGKDD Explorations.

[9]. Weka overview. (2013, April 25). Available from Sourceforge: http://weka.sourceforge.net/doc/

[10]. Chawla, N. O., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Ariticial Intelligence Research, 321-357.

[11]. Ling, C. X., & Sheng, V. S. (2008). Cost-sensitive learning and the class imbalance problem. In C. Sammat; G. Webb, editors. Encyclopaedia of Machine Learning. Springer.p.231-235.

[12]. Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 29-36.

[13]. Aksoy, S., & Haralick, R. M. (2013, May 20). Feature Normalization and Likelihood-based Similarity Measures for Image Retrieval. Available from cs.bilkent.edu: http://www.cs.bilkent.edu.tr/~saksoy/papers/prletters01_likelihood.pdf

[14]. Breiman, L. (2001). Random Forests. Machine Learning, 5-32.

[15]. Breiman, L. (1996). Bagging predictors. Machine Learning, 123-140.

[16]. Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm.In Machine Learning:Proceedings of the Thirteenth International Conference on Machine Learning, (pp. 148-156). San Francisco.

[17]. Wolpert, D. (1996). The lack of a priori distinctions between learning algorithms. Neural computation, 1341-1390.

[18]. Thornton, C., Hutter, F., Hoos, H., & Leyton-Brown, K. (2013, April 21). Auto-WEKA: Combined selection and hyperparameter optimisation of classification algorithms. Available from arxiv.org: `http://arxiv.org/pdf/1208.3719.pdf`

[19]. Carauna, R., Nikilescu-Mizil, A., Crew, G., & Ksikes, A. (2013, May 20).[Internet]Ensemble selection from libraries of models. Available from cs.cornell.edu: `http://www.cs.cornell.edu/~caruana/ctp/ct.papers/caruana.icml04.icdm06long.pdf`

[20]. Meyer, Z. (2013, April 21). New package for ensembling R models [Internet]. Available from Modern Toolmaking: `http://moderntoolmaking.blogspot.co.nz/2013/03/new-package-for-ensembling-r-models.html`

[21]. Gallivan, S; (2003) How likely is it that a run of poor outcomes is unlikely? European Journal of Operational Research , 150 46 - 52.

[22]. Hong, Y. (2013) On computing the distribution function for the Poisson binomial distribution. Computational Statistics and Data Analysis 59 41–51

[23]. Sherlaw-Johnson C. 2005 A method for detecting runs of good and bad clinical outcomes on Variable Life-Adjusted Display (VLAD) charts. Health Care Manag Sci. Feb;8(1):61-5.

Towards a Visually Enhanced Medical Search Engine

Lavish Lalwani[1,2], Guido Zuccon[1], Mohamed Sharaf[2], Anthony Nguyen[1]

[1] The Australian e-Health Research Centre, Brisbane, Queensland, Australia;
[2] The University of Queensland, Brisbane, Queensland, Australia.
lavish.lalwani@uqconnect.edu.au, m.sharaf@uq.edu.au,
{guido.zuccon, anthony.nguyen}@csiro.au

Abstract. This paper presents the prototype of an information retrieval system for medical records that utilises visualisation techniques, namely word clouds and timelines. The system simplifies and assists information seeking tasks within the medical domain. Access to patient medical information can be time consuming as it requires practitioners to review a large number of electronic medical records to find relevant information. Presenting a summary of the content of a medical document by means of a word cloud may permit information seekers to decide upon the relevance of a document to their information need in a simple and time-effective manner. We extend this intuition, by mapping word clouds of electronic medical records onto a timeline, to provide temporal information to the user. This allows exploring word clouds in the context of a patient's medical history. To enhance the presentation of word clouds, we also provide the means for calculating aggregations and differences between patient's word clouds.

Keywords. Visualisation, Timeline, Word Cloud, Medical Search.

Introduction

Current information systems deployed in clinical settings require practitioners and information seekers to review all medical records for a patient or enter database-like queries in order to retrieve patient information. Clinical data is often organised primarily by data source, without supporting the cognitive information seeking processes of clinicians and other possible users. For example, "The Viewer" application deployed by Queensland Health allows clinicians to access all patient electronic medical records collected by Queensland Health hospitals and facilities[1]. To access this information, clinicians need to enter data that allows them to select a patient (e.g., name, date of birth, Medicare number, etc.); afterwards they are given access to all information collected for that patient. However, they are unable to search through the medical records of the selected patient: if clinicians require a patient's past medical history, they have to read all medical records for that patient (organised by type of data, e.g. discharge notes, laboratory reports, etc., and clinical facility). This can be a very time consuming and tedious way of accessing information, particularly when clinicians

[1] Electronic medical record viewer solution, http://www.health.qld.gov.au/ehealth/theviewer.asp

want to review a large number of cases for research purposes, e.g. observe the effect a treatment had on their patient population.

An alternative solution is to deploy an information retrieval system where searches over patient records can be conducted with keywords, and medical records are ranked against the user query. We argue that this is a more efficient way for accessing patient information; previous research has developed systems that are able to search for relevant information in medical records [1, 2]. This paper considers how these systems could be improved by enhancing the presentation of results retrieved in answer to information seekers' queries. Search results are commonly shown to users as textual snippets that attempt to capture relevant portions of the medical record. Since these snippets are small chunks of text extracted from the original document (extractive summarisation), they often lack important information or can be misleading, especially if the original document is a medical record [3]. In addition, textual snippets do not convey an overview of the general clinical picture of a patient. For this reason, it is difficult to determine whether a medical case matches a search and whether it should be explored further; this thus requires the information seeker to access and read much of the document to determine its relevance to the query.

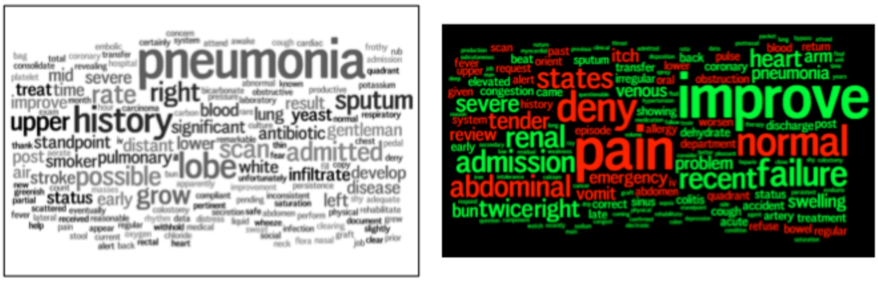
This paper investigates the use of data visualisation as a means for solving this problem. Data visualisation has the potential to provide a meaningful overview of medical reports, visits or even a patient's life and therefore may assist searchers to determine whether a medical document is relevant and worth further examination. Data visualisation may provide a simpler approach to augment standard searching methods for medical data. The remainder of the paper describes a system prototype that implements two data visualisation techniques: word clouds and timelines.

1. Related Work

Word clouds provide a visual representation of the content of a document by displaying words considered important in a document. Words are arranged to form a cloud of words of different sizes. The size of a word within a cloud is used to represent the importance of that word in the document; often, the importance of a word is computed as a function of the frequency of that word within a document. Figure 1 shows examples of word clouds.

In this paper we posit that word clouds have the ability to provide a better summary of the information contained in a medical record than textual snippets. This is supported by existing research on employing word clouds within information retrieval systems. For example, Gottron used a technique akin to word clouds to present news web pages [4]. In that study it was found that word clouds helped users to decide upon the relevance of news articles to their search query. Kaptein and Marx used word clouds to enhance information access to debate transcripts from the Dutch parliament [5]; they found that word clouds provided an effective first impression of the content of a debate.

Timelines are an additional data visualisation technique providing a map of events over time. The visualisation of events on a timeline provides the user with information related to which events occurred prior (and after) to an event of interest;. In our scenario, medical records belonging to a patient represents an event. Visualising medical records over a timeline allows for the possibility of mapping an entire patient's medical history within a unique visual representation. Previous research found that employing timelines for displaying patient medical records has the benefit of enabling clinical audit, reduced clinical errors, and improved patient safety [6]. Bui et al. have explored the use of timelines to give a problem-centric visualisation of medical reports, where patient reports are organised around diseases and conditions and mapped to a timeline [7].



(a) An example of word cloud computed from a medical record.

(b) An example of differential word cloud.

Figure 1. Word clouds computed from a medical record.

2. Word Clouds and Timelines

As supported by the previous research already outlined, this paper posits that word clouds and timelines can be effective visualisation techniques to provide quick information access to clinical records. The clinical records used to develop the prototype system were obtained from the TREC Medical Records Track corpus, a collection of 100,866 medical record documents taken from U.S. hospitals. Note that documents belonging to a single patient's admission were grouped together, obtaining a total of 17,198 groups of records. Next, we present the algorithms used within the system to generate word clouds and timelines.

2.1. Word Cloud Generation

The generation of a word cloud within our prototype system is a multi-step process.

The first step consists of removing tokens and words from the documents that convey limited or no information (stop word removal). These may include symbols, special characters, and words contained in a ‘stoplist’ (e.g. “the”, “a”, “when”, etc.). This step is used to avoid displaying irrelevant or non-informational words within the word clouds.

The second step involves stemming the text of the medical reports. Stemming consists of reducing a word to its base form (stem). Stemming is applied to conflate syntactical variations of the same word (e.g. plurals, gerund forms, past tense, etc.) into a single token to represent the fact that they may have the same or similar meaning.

The third step consists of generating a probability distribution over the vocabulary words *w,* in a document *d*, $P(w|d)$. Since a word cloud cannot display all the words in a document, this distribution is used to derive the list of words that will form the word cloud and their final font size (step four). Language models are used to compute such probability distributions. The probability of a word *w* in a document *d* is computed as a function of the occurrence of w in the medical records as the following equation mathematically explains.

$$P_\lambda(w \mid d) = (1-\lambda)P(w \mid d) + \lambda P(w \mid C) \qquad (1)$$

In Equation 1, $P(w|d)$ is calculated as the ratio between the number of occurrences of w in d and the total number of words in d (maximum likelihood estimate). Similarly, $P(w|C)$ is calculated as the ratio between the number of occurrences of w in the whole corpus of medical reports C and the total number of words in C. These probabilities are interpolated according to the parameter λ, which controls the importance of background information (i.e., $P(w|C)$) when determining the importance of word w in the context of document d. The use of both the maximum likelihood estimate and the background language modelling are referred to as Jelinek-Mercer smoothing; more details on language modelling can be found in [8].

The last step (fourth step) is the generation of the actual word cloud. Words in a document are ranked in decreasing order of their probability $P(w|d)$, and only the top ranked words are selected to be included in the word cloud. The probabilities of the selected words are mapped into font sizes, and the appropriately sized words are placed in the word cloud for document d. Figure 1a shows an example of a word cloud generated from a patient medical report.

2.2. Word Cloud Aggregation

Individual word clouds could be merged to visualise an entire patient hospital visit or medical history as a unique word cloud. Two word clouds wc_1 and wc_2 are merged according to the following equation:

$$P(w) = P(w \mid wc_1)P(wc_1) + P(w \mid wc_2)P(wc_2) \qquad (2)$$

where $P(w|wc_i)$ represents the probability[2] of word w in word cloud wc_i, and $P(wc_i)$ is the probability associated to wc_i. Currently, we consider word clouds to be uniformly distributed (thus $P(wc_1) = P(wc_2)$); however future developments may consider biasing word clouds according to temporal relations or document types when merging. As previously stated, Equation 2 can also be used to create a word cloud representing a complete patient medical history by merging all the word clouds associated to their medical records. Similarly, Equation 2 can be applied for merging word clouds associated with reports belonging to different patients.

2.3. Word Cloud Differential

A differential word cloud is designed to highlight the differences between two word clouds (i.e. between two documents). Since two word clouds are effectively two probability distributions, their difference can be computed using the Kullback-Leibler (KL) divergence. Equation 3 provides the means for computing the difference between word clouds, given the source word clouds wc_1 and wc_2.

[2] $P(w|wc_i)$ is equivalent to $P(w|d)$ if wc_i represents the word cloud for document d; however, note that wc_i may have also be computed from the merging of other previously computed word clouds.

$$D_{KL}(wc_1 \| wc_2) = \sum_i P(w_i \mid wc_1) \log \frac{P(w_i \mid wc_1)}{P(w_i \mid wc_2)} \tag{3}$$

The magnitude of the KL divergence can be thought of as the degree of difference between the two word clouds. The value of KL divergence for each word can be used to generate a word cloud that provides visual information about how the two original word clouds differ. We refer to this type of word cloud as a differential word cloud (between wc_1 and wc_2). In a differential word cloud, the sign of D_{KL} for each word (i.e. $D_{KL}(w,wc_1||w,wc_2) = P(w|wc_1)log[P(w|wc_1)/P(w|wc_2)]$) determines the colour the word would be painted with. Words with positive D_{KL} values are painted green and words with a negative D_{KL} values are painted red. In this case, if a word is painted green it means it has a stronger presence (i.e. higher probability) in wc_1. The degree to which this presence is stronger is signified by the size of the word in the cloud (the bigger the word, the stronger the difference in presence). The opposite applies for a red colour word in the differential word cloud. Note that if the calculation was conducted with the probabilities in reverse order, the colours on the differential word cloud will reverse. An example of a differential word cloud is shown in Figure 1b.

2.4. Timeline Generation

The generation of timelines involves, for each medical report, extracting the date and time it was created. This was achieved using metadata information present in the reports from the TREC Medical Records Track corpus; however, it is acceptable to assume that similar metadata is present in records from other hospital providers. Since entire patient admissions were mapped to timelines, after dates and times are extracted for all records in a patient admission, this metadata, along with the medical record data are rendered within a timeline created using the Java Script library, Timeline JS[3]. This means that when retrieving a particular medical record, it can be displayed within context of the other reports produced for that patient admission.

3. Integration of Word Clouds and Timelines

The prototype described here is a modular information retrieval system, developed based on the Apache Lucene 4 framework, specifically for searching archives of medical records. Its architecture consists of three main modules: the indexer, the visualiser, and the searcher.

Within the indexer module, medical records are parsed and stored within a representation appropriate for supporting the retrieval stage (inverted file). The indexer is built using the Apache Lucene 4.0 incremental indexing capabilities, thus allowing new documents to be included in the index without re-indexing the previous documents. The indexer also maintains the relation between medical records and patients.

The searcher module is responsible for retrieving documents from the index that match a user query. A ranked list of medical admissions is produced as the result of querying the system.

[3] http://timeline.verite.co/

The visualiser module has the responsibility of rendering the results of a search and supporting navigation across search results. The modular architecture of the system integrates the visualisation methods described in Section 2 within the visualiser module without modifying the approaches used to index and retrieve documents. Indeed, the visualiser module is independent of the processes used in the other modules, allowing for flexibility when devising and testing new visualisation algorithms, as well as deploying versions of the system tailored to specific scenarios. Figure 2 shows a screenshot of an implementation of the methods described in Section 2 within the prototype system visualiser module. The figure illustrated a situation where a user has submitted a query and is in the process of examining a specific medical record. The content of the record is rendered as a word cloud allowing the user to quickly understand the content of the record itself. The text of the record can be accessed through the “Reports view” button above the word cloud. The record is also placed within the timeline of the patient admission to the hospital (bottom of Figure 2).

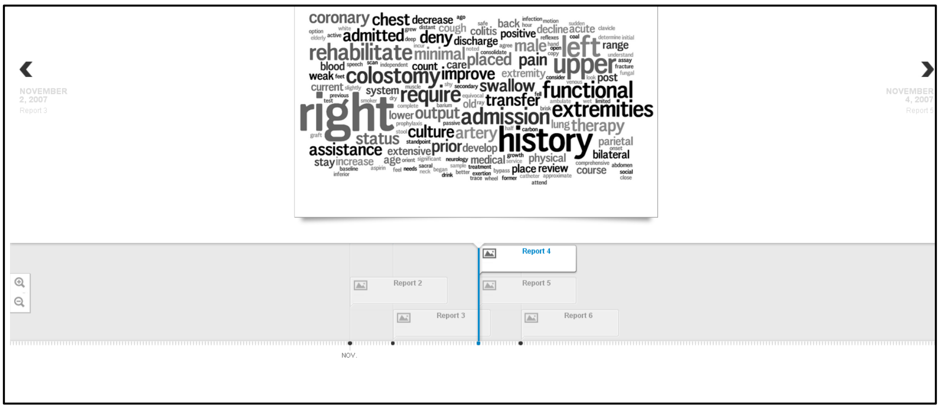


Figure 2. A screenshot of the visual interface of the system showing the use of word clouds and timelines.

4. Conclusion

In this paper we have presented two techniques, word clouds and timelines, to enhance search results presentation within medical records search. Word clouds have the potential to provide a rapid overview of an entire medical report, admission and patient history. Timelines provide a visual means to represent patient journeys as well as to place a medical record within the temporal context of other existing records. These techniques were integrated within the visualiser module of our prototype, a state-of-the-art medical information retrieval system. Future work will be directed towards a formal evaluation of the proposed techniques in a real scenario. Possible improvements will consider n-grams (sequences of *n* words, e.g. ‘heart attack’) and medical concept detection and reasoning (e.g. “heart attack” and “myocardial infarction” within a record should contribute towards the same medical concept) when building and rendering word clouds.

References

[1] Voorhees, E., & Tong, R. Overview of the TREC 2011 Medical Records Track. *In Proceedings of TREC* (2011).
[2] Zuccon, G., Koopman, B., Nguyen, A., Vickers, D., & Butt, L. Exploiting Medical Hierarchies for Concept-Based Information Retrieval. *In Proceedings of ADCS* (2012), 111-114.
[3] S. Afantenos, V. Karkaletsis & P. Stamatopoulos, Summarization from Medical Documents: a survey, *Artificial Intelligence in Medicine* **33** (2005), 157-177.
[4] T. Gottron, Document Word Clouds: Visualising Web Documents as Tag Clouds to Aid Users in Relevance Decisions, *Lecture Notes in Computer Science* **5714** (2009), 94-105.
[5] Kaptein, Rianne, and Maarten Marx. Focused Retrieval and Result Aggregation with Political Data. *Information retrieval* 13.5 (2010): 412-433.
[6] Gill, J., Chearman, T., Carey, M., Nijjer, S., & Cross, F. Presenting Patient Data in the Electronic Care Record: the role of timelines. JRSM short reports, 1(4), (2010).
[7] Bui, A. A., Aberle, D. R., & Kangarloo, H. TimeLine: visualizing integrated patient records. Information Technology in Biomedicine, IEEE Transactions on, 11(4), (2007), 462-473.
[8] Zhai, C. Statistical Language Models for Information Retrieval. Synthesis Lectures on Human Language Technologies, 1(1), (2008), 1-141.

Using Fuzzy Logic for Decision Support in Vital Signs Monitoring

Shohas Dutta, Anthony Maeder, Jim Basilakis
School of Computing, Engineering & Mathematics, Telehealth Research & Innovation Laboratory
University of Western Sydney
Private Bag 1797, Penrith 2751, NSW

shohas6@gmail.com, a.maeder@uws.edu.au, j.basilakis@uws.edu.au

Abstract

This research investigated whether a fuzzy logic rule-based decision support system could be used to detect potentially abnormal health conditions, by processing physiological data collected from vital signs monitoring devices. An application of the system to predict postural status of a person was demonstrated using real data, to mimic the effects of body position changes while doing certain normal daily activities. The results gathered in this experiment achieved accuracies of >85%. Applying this type of fuzzy logic approach, a decision system could be constructed to inform necessary actions by caregivers or for a person themself to make simple care decisions to manage their health situation.

Keywords: fuzzy logic, patient monitoring, decision support, assistive technologies, care management.

1 Introduction

Current trends in health within our society include the move towards an ageing population profile, and increased needs for complex care management for people with chronic diseases and multiple co-morbidities. These are fast growing segments of the population; and so is the need for covering their broad ranging and diverse care requirements. External support to manage high-risk (or unsafe) health situations is often needed for them to continue their everyday living routines. This support is typically given by both professional and informal caregivers.

Due to technological advances in wireless data communication systems in the last decade, the application of wireless-based vital sign monitoring devices for patient monitoring has gained increasing attention in the clinical arena. Patient health status can be determined based on the acquisition of basic physiological vital signs, suggesting that a system providing wireless monitoring of vital signs has potential benefits for clinical care management of independently living patients as well as their carers. A patient's physiological state, which includes heart rate, blood pressure, body temperature etc., can be monitored continuously using wearable medical body sensor devices. The remaining challenge is to gain sufficient understanding of this data to assist in health care needs.

The overall aim of this research was to utilise information gathered from personal vital signs monitoring in a laboratory-based smart home environment, and to assist with clinical care decisions using a fuzzy logic rule-based clinical decision support system. Fuzzy logic has benefits over other algorithmic approaches, as it has the potential to incorporate values from ordinal, nominal and continuous datasets within its rules, and can capture the knowledge associated with these rules in ways that are more intuitive to humans.

2 Vital Signs Monitoring Concepts

There are numerous examples in literature describing how monitoring of basic vital signs (i.e. heart rate, blood pressure, temperature and respiration rate) can play a key role in health care, e.g. Norris (2006) [39]. This approach requires software to discover patterns and irregularities as well as to make predictions. By collecting and analysing vital signs continuously it can be shown how well the vital organs of the body are working, e.g. heart and lungs (Harries et al. 2009) [40].

Lockwood et al. (2004) [30] provided a review of the clinical usage of vital signs, including monitoring purpose, limitations, frequency and importance of vital signs measurements. They suggested that vital signs monitoring should become a routine procedure in chronic disease patients' care. Bentzen (2009) [43] defined chronic diseases as:

"diseases which are long in duration, having long term clinical course with no definite cure, gradually change over time, and having asynchronous evolution and heterogeneity in population susceptibility."

Living with a chronic disease, which increases in severity with age, has a significant impact on a person's quality of life and on their family. Chronic disease patients would be able to play a more active role in managing their own health by taking vital signs measurements daily and participating in meaningful electronic information exchanges with clinicians.

A number of authors have suggested that using smart homes for health monitoring is a promising area for health care. Chan et al. (2009) [2] in their review paper described the smart home as a promising and cost-effective way to improve home care for elderly people and people suffering with different chronic diseases.

Vincent et al. (2002) [19] identified three research areas, which combined to produce the concept of "health smart home". These three areas are *medicine*, *information systems*, and home based automatic and remote *control* *devices*. A smart home contributes to monitoring of the patient's health status continuously, taking into consideration the patient's personal needs and wishes in addition to their specific medical requirements. The information gathered through health status monitoring systems can feed into an access controlled electronic patient records system for further medical interpretation.

LoPresti et al. (2008) [21] identified different assistive technologies which can be used in smart homes to reduce the effect of disabilities and improve quality of life. Wearable and portable devices are used which help to monitor the vital signs or physiological behaviour of a person living in a smart home. Those devices are worn by the user or embedded in the smart home. They are wired or wirelessly connected to a monitoring centre. Recently, robotic technology has been developed to support basic activities and mobility for elderly people too.

3 Fuzzy Logic Concepts

Fuzzy logic (Zadeh 1990) [68] is a well established computational method for implementing rules in imprecise settings, where some adaptability for prescribing the rules is necessary. A fuzzy system can be used to match any set of input-output combinations. Fuzzy logic can provide us with a simple way to draw definite results from vague, ambiguous or imprecise information. The rule inference system of the fuzzy model (Jang 1993) [67] consists of a number of conditional IF-THEN rules. For the designer who understands the system, these rules are easy to write, and as many rules as are necessary can be supplied to describe the system adequately.

To improve clinician performance, fuzzy logic-based expert systems have shown potential for imitating human thought processes in the complex circumstances of clinical decision support (Pandey 2009) [75]. A key advantage of using fuzzy logic in such situations is that the fuzzy rules can be programmed easily, and as a result they are easily understood by clinicians. It is different from neural networks and other regression approaches, where the system behaves more like a black box to clinicians. Schuh (2008) [73] found that fuzzy logic holds great promise for increasing efficiency and reliability in health care delivery situations requiring decisions based on vital signs information. This has also been observed in specialised situations such as intensive care (Cicilia et al 2011) [81].

Fuzzy control is the core computational component of a fuzzy logic system. It includes the processing of the measured input values based on the fuzzy rules, and their conversion into decisions with the help of fuzzy combination logic. A full description of fuzzy control principles is beyond the scope of this paper and can be found in numerous fuzzy logic texts. The functional elements of fuzzy control can be represented in a block diagram in Figure 1, based on fuzzy membership functions of variables of interest, as shown in Figure 2 for the example of body temperature represented by the variable T.

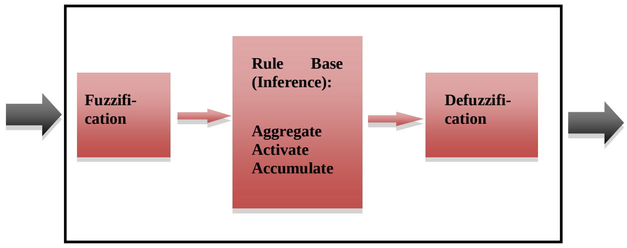


Figure 1. Elements and structure of fuzzy control.

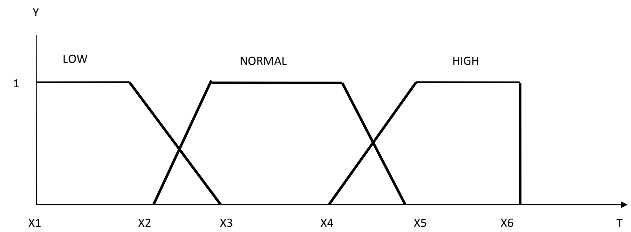


Figure 2. Fuzzy membership functions of variable T.

4 Experimental Methodology

This section will discuss the design of a laboratory experiment to undertake validation of the approach, using a longitudinal data set of physiological signals which have been gathered from an experiment involving monitoring of blood pressure and heart rate signals. It is well known that changes to these vital signs will occur if the body position is changed from vertical to horizontal. The nature and rapidity of these changes mimics the changes in vital signs that may occur with onset of some exacerbated or acute health status in patients.

The laboratory setup used a tilt table to generate changes in heart rate and blood pressure measurements that were correlated with the angle of the tilt table (Figure 3). These physiological changes would be similar to changes one would expect in circumstances such as changing health status or other physiological stressors such as an infection or blood loss. The result of the fuzzy logic analysis of such data can be used to detect a change in physiological state occurring when the vital signs measures are either increasing or decreasing, compared to a steady state where there are no longitudinal changes in the vital sign measures. This output can be compared against the angle of the tilt table, that will serve as a gold standard for determining whether the system is in a steady state or not.

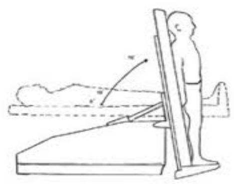


Figure 3. Movement range of tilt table.

The tilt table used was a motorized table with a metal footboard. The subject's feet were rested on the footboard. Soft Velcro straps were placed across the body for safety reasons, to secure the person when the table was tilted during the test. When using the tilt table, it was always tilted upright so that the head of the subject was above his feet. Small, sticky patches containing electrodes were placed on the subject's chest. These electrodes were connected to an electrocardiograph monitor (ECG) to record the electrical activity of the person heart to be shown as an ECG graph. The ECG showed the heart rate and rhythm during the test, at a raw sampling rate of 100Hz and an accuracy of 3%. A blood pressure measuring device was also attached on the subject's finger. This was connected to monitors so that the blood pressure could be observed during the test as well as being recorded.

At the very beginning of the test, the subject was laid flat on his back on the tilt table. At that time his initial blood pressure, ECG, and his position angle data were recorded. After resting for few minutes, the test was started. The blood pressure and ECG was constantly monitored throughout the test and instantaneous readings of the data stream were recorded every second for subsequent analysis. The following protocol was applied for changing the positioning of the tilt table:

1. Lying flat at rest for ~60 sec (to gain statistics of resting state)
2. Fast tilt upwards over ~10 sec
3. Very slow tilt downwards over ~30 sec
4. Lying flat resting state ~30 sec
5. Medium tilt upwards over ~20 sec
6. Upright resting state ~30 sec
7. Fast tilt downwards over ~10 sec
8. Lying flat resting state ~30 sec
9. Fast tilt upwards over ~10 sec
10. Upright resting state ~30 sec
11. Medium tilt downwards over ~20 sec
12. Lying flat resting state ~30 sec

A sample data set collected recorded using the above protocol is shown in the graphs in Figure 4. Data sets from three repetitions of the protocol were captured using one of the investigators as the subject, as a pre-ethics proof-of-concept exercise needed to justify a full human research ethics application for extending the work for recruited subjects in the future. Little variability was observed in the three data sets, so it was considered unnecessary to collect further test data.

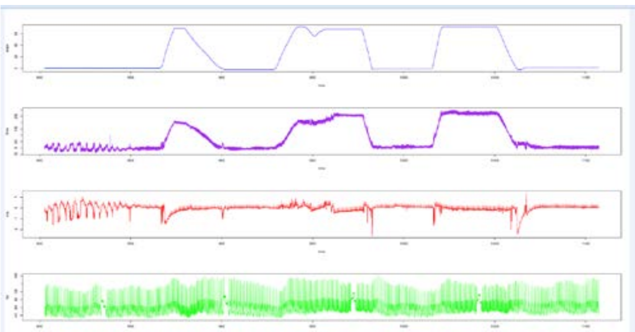


Figure 4. Data captured from the experiment: (top to bottom): angle, footplate force, ECG, blood pressure.

5 Experimental Results

The fuzzy logic rules were derived using the blood pressure and heart rate signals from the first of the three cycles. These signals were pre-processed to find a smoothed curve of the recorded raw signals. In this smoothing process, the averages of the values of heart rate and blood pressure were calculated for every five timestamps using non-overlapping windows. Then these average values were used to plot a smooth curve of the systolic blood pressure and peak-to-peak heart rate to establish the trends. Figure 5 shows the training dataset.

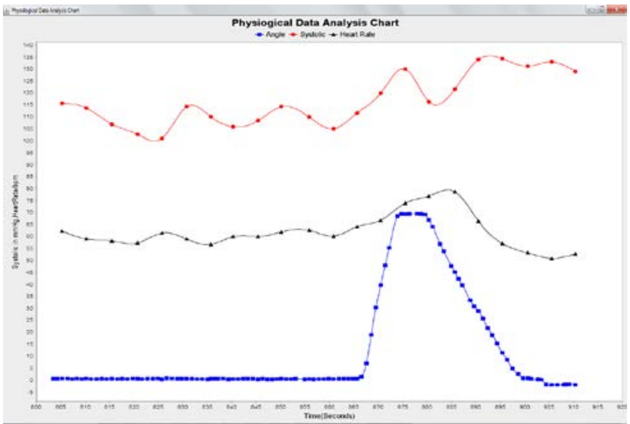


Figure 5. Training dataset (top to bottom): blood pressure, heart rate, tilt angle.

The fuzzy logic solution has two input variables and one output variable. Using the mean and standard deviation as a tolerance band for the input variables, three states (Low, Normal, High) are defined. The two input variables are combined by the AND (i.e. MAX) operator and valid states inferred from the values for the tilt angle, as represented in the decision matrix shown in Table 1.

Table 1. The decision matrix for the training data.

		Input Variable 1: Systolic Blood Pressure		
		Low	**Normal**	**High**
Input Variable 2: Heart Rate	**Low**			Static
	Normal	Static	Static	Lowering
	High		Lowering	Raising

The following rules based on this table were derived:

RULE 1: IF systolic IS low AND heart_rate IS low THEN physiological_status IS Unclassified;

RULE 2: IF systolic IS low AND heart_rate IS normal THEN physiological_status IS Static;

RULE 3: IF systolic IS low AND heart_rate IS high THEN physiological_status IS Unclassified;

RULE 4: IF systolic IS normal AND heart_rate IS low THEN physiological_status IS Unclassified;

RULE 5: IF systolic IS normal AND heart_rate IS normal THEN physiological_status IS Static;

RULE 6: IF systolic IS normal AND heart_rate IS high THEN physiological_status IS Lowering;

RULE 7: IF systolic IS high AND heart_rate IS low THEN physiological_status IS Static;

RULE 8: IF systolic IS high AND heart_rate IS normal THEN physiological_status IS Lowering;

RULE 9: IF systolic IS high AND heart_rate IS high THEN physiological_status IS Raising;

The derived fuzzy rules were applied to the smoothed data of the test set for the second and third cycles, to determine the physiological status. By applying fuzzy logic to these two cycles of testing data, different regions in the data were classified into predicted statuses of Static, Raising and Lowering. Figure 6 shows the results with yellow indicating static status, grey indicating lowering status and green indicating raising status.

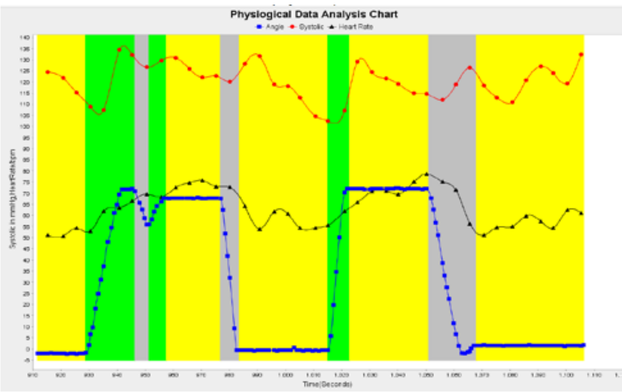


Figure 6. Classifying status using the trained rules.

In order to compare the fuzzy logic output to the gold standard, statuses needed to be inferred from the angle of the tilt table. The following protocol was established to determine three different states categorised as: Static, Raising and Lowering. Only changes of one or more smoothing period timesteps (i.ee >4 sec) were considered. The protocol used was as follows:

1. If the change of angle is < 5° and timestamp interval >4 sec, then the tilting table is in static state.
2. If the change of angle (upward) is: 25°< angle<90° and timestamp interval >4 sec, then the tilting table is in abnormal state and in the raising state.
3. If the change of angle (downward) is: 25°< angle<90° and timestamp interval >4 sec, then the tilting table is in the lowering state.

The results using these steps are summarised in Table 2, and the overall rate of positive and negative outcomes is shown in Table 3. These outcomes were used to analyse classifier performance using the following indicators:

Sensitivity = TP/(TP+FN) = Prob(+ve test)

Specificity = TN/(TN+FP) = Prob(-ve test)

Accuracy = (TP+TN)/total obs = Prob(correct)

Error = (FP+FN)/total obs = Prob(wrong)

Table 2. Matching actual states and predicted states.

		Predicted State (Computed)			
Actual State (Gold standard)		Static	Raising	Lowering	Total
	Static	24	1	2	27
	Raising	2	2	2	6
	Lowering	1	0	5	6
	Total	27	3	9	39

Table 3. Classifier positive and negative outcomes.

	Test Outcome (Static case)	
Gold Standard Set (Static case)	True Positive (24)	False Positive (3)
	False Negative (3)	True Negative (9)
	Test Outcome (Raising case)	
Gold Standard Set (Raising case)	True Positive (2)	False Positive (4)
	False Negative (1)	True Negative (32)
	Test Outcome (Lowering case)	
Gold Standard Set (Lowering case)	True Positive (5)	False Positive (1)
	False Negative (4)	True Negative (29)

The resulting indicator values were calculated as follows:

Sensitivity (Static) = 24 / (24+3) = 24 / 27 = 0.89

Specificity (Static) = 9 / (9+3) = 9 / 12 = 0.75

Sensitivity (Raising) = 2 / (2+1) = 2 / 3 = 0.67

Specificity (Raising) = 32 / (32+4) = 32 / 36 = 0.89

Sensitivity (Lowering) = 5 / (5+4) = 5 / 9 = 0.56

Specificity (Lowering) = 29 / (29+1) = 29 / 30 = 0.97

Accuracy (Static) = (24+9) / 39 = 33 / 39 = 0.85

Error (Static) = (3+3) / 39 = 6 / 39 = 0.15

Accuracy (Raising) = (2+32) / 39 = 34 / 39 = 0.87

Error (Raising) = (4+1) / 39 = 5 / 39 = 0.13

Accuracy (Lowering) = (5+29) / 39 = 34 / 39 = 0.87

Error (Lowering) = (1+4) / 39 = 5 / 39 = 0.13

Across the three states, Sensitivity values ranged from 0.56 to 0.89, and Specificity values ranged from 0.75 to 0.97. The low Sensitivity values are related to the smaller sample sizes for the Raising and Lowering states. Accuracy rates ranged from 0.85 to 0.87, and Error rates ranged from 0.13 to 0.15, indicating good performance.

In considering the performance of this approach, several drawbacks affected the achievable accuracy negatively. The first issue was the time lag in the dropping of the vital sign values when changing the angle of the tilting table. While the tilting table was moved rapidly, it took several seconds for the physiological status of the human body to adapt accordingly. As a result, this problem has affected accuracy in determining the physiological status of a person in FastUp or in FastDown status.

Another problem was related to the error rate associated with using the vital signs measurement equipment. When the position of the tilt table was changed, small movements of the body affected accurate measuring of the physiological data by the monitoring devices. For example, the blood pressure measuring device was attached with the finger and due to the movement of the body and fingers it sometimes gave erroneous readings. The smoothing function that was applied was intended to damp out such errors but there is some residual effect.

6 Conclusion and Future Work

We have described an efficient computational approach to the problem of personal monitoring of vital signs, to provide alerts under well defined abnormal health status conditions which are caused by a known or anticipated health situation. The purpose of such alerts is to provide decision support inputs to carers, to prompt closer observations or direct interventions to be performed to help the subjects of care. This could be useful over a wide range of situations such as elderly or disabled living alone, or patients with chronic diseases or multiple co-morbidities.

Fuzzy logic was chosen as an appropriate computational approach due to its simplicity and ease of tuning to suit relatively smoothly changing vital signs values. Then the approach was implemented in software, providing a multistage process for classifying the condition of a subject using fuzzy functions for each of several observed vital signs, and then combining these using rules to determine the overall health status.

Using this approach, a fuzzy logic rule-based decision support system could, for example, be used to monitor daily activities of living and detection of falls for smart home residents, in combination with other technologies that have more sensitivity in detecting sudden change of body posture such as tri-axial accelerometers. Further research is required to find out the usefulness of such a fuzzy logic rule-based decision support system when a combination of vital signs and acceleration data is used to detect sudden changes in body posture.

On the basis of this foundation work, fuzzy logic has been shown to provide a plausible approach to the general problem of classifying health status in situations of abnormalities in vital signs patterns. It is anticipated that a more extensive system could be built by including further parameters and more complex rules, using the same fundamental algorithm. The implementation methodology using an SQL database and fixed form parameter labelling functions for the fuzzy assignments, provides a robust implementation environment and a sufficiently simple rule specification mechanism to allow users who are not IT experts to reconfigure the system to suit a given vital signs classification problem.

A worthwhile extension of this work would be to improve the level of sophistication and automation of the threshold values for the fuzzy logic classification process. Instead of a simple statistical approach using a set of “normal” observations, actual patterns could be captured and stored which could be tested with greater severity than smooth fuzzy functions. The work offers scope to increase the amount of ambient intelligence which could be provided in the “smart home” of the future, to help sustain occupants’ health circumstances.

7 References

Chan M, Campo E, Estève D, Fourniols JY., *Smart homes — Current features and future perspectives*, Maturitas, 2009, 64(2), p: 90-97.

Vincent R, Norbert N, Lionel B, Jacques D., *Health "Smart" home: information technology for patients at home.* Telemedicine Journal and eHealth, 2002. 8(4), p: 395-409.

LoPresti EF, Bodine C, Lewis C., *Assistive technology for cognition Understanding the Needs of Persons with Disabilities.* Engineering in Medicine and Biology Magazine, IEEE, 2008. 27(2), p: 29-39.

Lockwood C, Tiffany CH, Page T., *Vital Signs.* JBI Reports, Wiley Online Library, 2004. 2(6) p: 207-230.

Norris PR., *Toward New Vital Signs: Tools and Methods for Physiological Data Capture, Analysis, and Decision Support in Critical Care,* PhD Thesis, Graduate School of Vanderbilt University, 2006.

Harries AD, Rony Z, Kapur A, Jahn A, Enarson DA., *The Vital Signs of Chronic Disease Management.* Transactions of The Royal Society of Tropical Medicine and Hygiene, 2009. 103(6), p: 537-540.

Bentzen N., *WONCA dictionary of general/family practice,* Mental Health in Family Medicine, 2009, 6(1), p: 57–59.

Jang JSR., *ANFIS: Adaptive-Network based Fuzzy Inference Systems*, IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(3), p: 665-685.

Zadeh LA., *The birth and evolution of fuzzy logic*, International Journal of General Systems, 1990, 17(2-3), p: 95-105.

Schuh CJ., *Monitoring the fuzziness of human vital Parameters*, Annual Meeting of the North American Fuzzy Information Processing Society, IEEE 2008, p:1-6.

Pandey B, Mishra R., *Knowledge and intelligent computing system in medicine.* Computeras in Biology and Medicine, 2009, 39(3), p: 215-30.

Cicília RM L, Glaucia RS, Adrião DDN, Ricardo AMV, Ana MG G., *A fuzzy model for processing and monitoring vital signs in ICU patients,* BioMedical Engineering OnLine, 2011, 10(68).

A Novel Approach for Improving Chronic Disease Outcomes using Intelligent Personal Health Records in a Collaborative Care Framework

Amol Wagholikar[1]

1: The Australian e-Health Research Centre, CSIRO Computational Informatics

Abstract.
Background- Effective management of chronic diseases is highly important to improve health outcomes of chronic disease patients. Emerging initiatives on online personal health records (PHR) have provided an opportunity to empower patients living with a chronic disease to take control of their own health data management. Online PHR solutions also provide data-driven intelligent analytics capability that can provide an effective view of the patient's health data to the patients themselves as well as to their consented clinicians and carers such as family members fully engaged in their routine care. Research suggests a tendency among chronic disease patients of using self-managed care without as well as with some support to monitor and manage their chronic disease. The rising usage of online solutions enable chronic disease patients store their physical as well as mental health information in self-managed online PHR. There are a variety of such online PHR mechanisms that are available via desktop computers, mobile smartphones, smart TVs as well as biometric devices. However, the main problem of disparate data sources and lack of a universal view of patient's health data still exists. These problems needs a novel way of integrating various types of PHRs in an efficient way and provide effective insights about the patient's health to empower and engage the patients in active management of their chronic condition. **Objective-** To describe a framework to integrate various online PHRs for providing effective self-managed and collaborative care. **Methods-** Comprehensive research was conducted to analyse current trends of various PHR mechanisms. A series of discussions were held with the clinical as well as non-clinical end users of online PHRs to identify the current problems with accessing PHRs and their expectations about usage of PHR in managing chronic disease condition. The requirements analysis and emerging technology trends were utilized to develop a framework that provides intelligent capabilities for a collaborative online platform. **Results-** The requirements analysis and discussions with the end-user representatives showed that the proposed framework is considered novel and intuitive by the stakeholders thus confirming our findings. **Conclusion-** The results of this investigation specified a novel framework that can enhance the value of PHRs and thus may address usability challenges identified by the PHR developers as well as the end-users.

Keywords: Personal Health Record (PHR), Collaborative Care, Self-managed care

1 Introduction

A personal health record (PHR) is a record in a tangible document format (e.g. information recorded on a piece of paper and/or in an electronic document); in which an individual patient creates, maintains and controls his/her health related data [1]. The patient may access, modify as well as control the individual health information before using it for specific purposes such as self-assessment and sharing it with care providers through a consent process. The patients may also store a copy of data collected by their clinicians in their personal health record. The patient's PHR is a component of complete set of the patient's health related information as some information is also created, stored and managed in hospital and clinic health information systems.

Personal Health Records exist in paper-based (offline) format as well as electronic format (online). Some percentage of patients may regularly use offline PHR's to store and access their chronic disease specific health information. A certain section of patients may use various online PHR mechanisms (such as website-based tools and mobile applications provided by private vendors) to manage their health related information. The online PHR is an electronic record of an individual's personal health information stored securely in a central repository that can be accessed by an individual patient for self-managed care, self-monitoring of health conditions and can be shared with the clinicians for clinical use. The patients may choose to share the online electronic health record with clinical information systems used by their care provider to provide an accurate and a complete set of information required for providing point-of-care health services at various geographic as well as clinical settings [2, 3, 4].

1.1 Current State-of-the-art

The online personal health record is an emerging discipline of research. The current research in this discipline makes an attempt to improve value of the personal health records through application of innovative intelligent data processing methods [5]. The concept of Personal Health Record (PHR) has been evolved along with the advancements in web-based technologies. The current state-of-the-art suggests that intelligent PHRs are evolving to add more features such as data exchange, data sharing with clinicians, and family members. There are certain proprietary PHR solutions as well as open source PHR solutions. Each of the PHR solution offers common functions and features that can be accessed by the patients through a web-enabled device with a web browser. These solutions are evolving to add more features such as data exchange, data sharing with clinicians, families and carers. However, there are open issues in the intelligent PR especially in the area of functions provided by these solutions. Our review of the existing PHR solutions indicates that there is a lack of collaborative functions in the PHR [6]. This work attempts to address this gap by proposing a collaborative PHR platform.

1.2 Main Problem

There are certainly growing efforts both in private as well as public domains for adopting online PHR as a data recording and analysis tool for self-managed care, self-monitoring of disease conditions, as a preventative health intervention and clinical use. There is growing number and various types of private online PHR solutions that are available as a web-application and/or mobile application storing patient health data. The growing number of mobile health applications for tracking and monitoring exercise is a good example of this evolution (E.g. Nike+ app or the Fitbit app with optional weight scale and wrist fitness band). The various types of PHR's can be also categorised as per the devices that can be used to access the PHR. The types of online PHR are shown in table 1.

Table 1. Types of Online PHR

No	Online PHR Type	Access Devices
1	Web-based PHR	Desktop Computer, Smart phone, tablet device
2	PHR in Mobile App	Smart phone, tablet device
3	PHR in wireless monitoring device	Smart phone + portable device such as blood glucose monitor

The advances in web-based and mobile apps online technologies as well as rising use of online data recording and analysis solutions has led to the development and launch of private and public online PHRs. The broad categories of online PHR systems are illustrated in Figure 1 which links with the relationship illustrated in Figure 1.

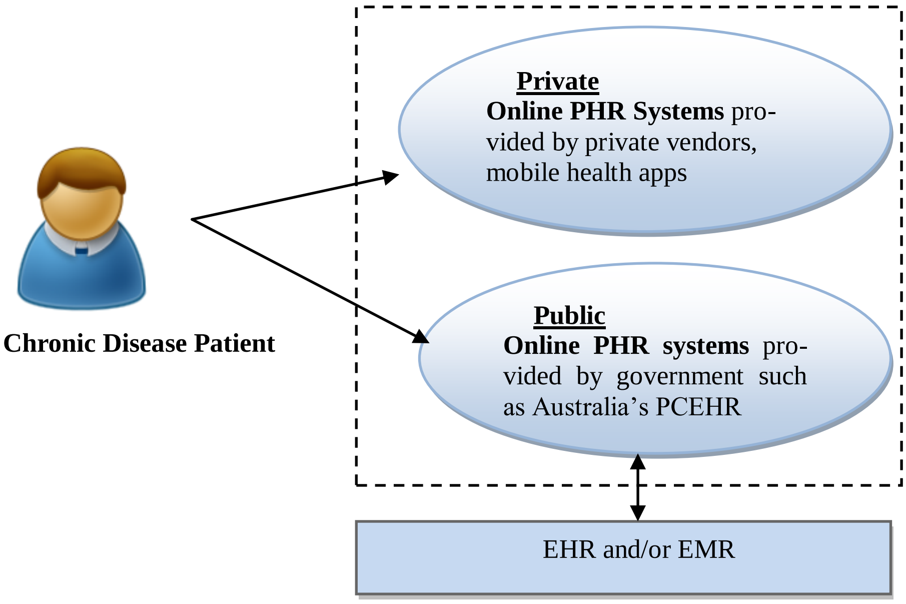


Fig. 1. Types of Online PHR Systems

Due to the widespread use of web-based PHR solutions, the patient's health data is stored in various disperse data sources. Thus, despite advances in online solutions, the core issue of "Health Information SilOs" (HISO) still exists and thus the issue of accurate information for self-managed care as personalised decision support is still largely unresolved. We propose an approach in the form of a framework that attempts to address this issue with a new design proposal for an intelligent PHR framework.

2 Methods

This research was undertaken for an initiative that aims to improve journey of chronic disease survivors. The following main steps were undertaken for our research.

- **Expert Interviews:** A group of clinical experts, patient representatives as well as technology experts was engaged to understand the real-life issues of chronic disease survivors. A series of expert interviews were conducted to understand the main drivers as well as requirements for developing an online technology approach to leverage advances in PHR, established as well as emerging industry trends in health information technology. A key challenge of providing a seamless experience for the patients to manage their own health data was identified during these interviews. The challenges of adoption by the chronic disease patients with limited health and information technology literacy were also identified. The inputs from the interviews were used to specify requirements of our proposed platform. The specifications aimed to propose an innovative design of a PHR-based solution using a vendor-provided PHR platform with customization.
- **Online Solution Investigation:** A comprehensive research was conducted to investigate global landscape of emerging online PHR solutions including desktop as well as smart devices (smart phones and tablets) based solutions. The results of the comprehensive research are summarized in the table below-

Table 2: Online PHR status around the world

Country	PHR Solution	Roll Out	PHR Standards	Current Status
Australia	PCEHR [7]	July 2012	NEHTA PCEHR	Active Adoption in progress
UK	NHS Healthspace [8,9]	2010	HL7 and others	Closed in December 2012
Canada And US	Various Private Online PHRs[10], Big blue button[11], blue button+ as a public PHR in US	Since 2009	Proprietary and open source	Active Adoption in progress

- **Proposed Approach Development:** The investigation resulted into recommendation of our technology platform that can address the issues of HISO in an online PHR context.

3 Proposed Approach for an Integrated PHR

Our investigation resulted into identifying key challenges and critical needs for providing a single platform that can provide a holistic view of the patient's PHR. We propose a platform that can not only record critical data but also provide intelligent analysis of patient's health data to patients as well as their carers. A schematic representation of our approach is shown in figure 2.

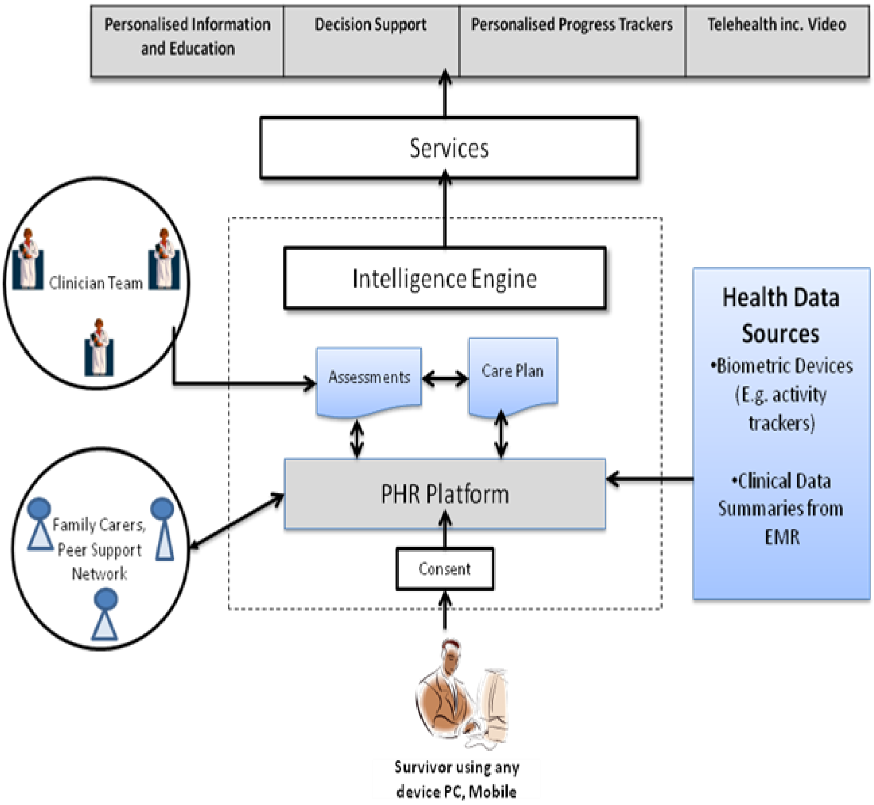


Fig. 2. Schematic Representation of Our Proposed Approach

One of the central components of our proposed platform is the intelligence engine. The intelligence engine component will executive algorithms driven by the personal health data stored in the PHR. The specific details of the algorithms will be reported as the research progresses. Our proposed platform also provides intelligent analytics of the patient's health data which improves patient's own understanding of their health data. The intelligent analytics also improves monitoring of key health indicators such as weight, mood, and nutrition in a personalized visual dashboard.

Our proposed approach is in the form of a collaborative PHR platform addresses the issue of disperse health data sources. It allows the patients as well as their care providers a universal view of patient's health data governed by the consent process. The clinicians also have the ability to view as well as add clinical notes to the patient's PHR. The patient's can collaborate with their clinicians as well as similar patients through the platform which can integrate with patient-driven social network. The patients can connect with their carers through video, voice as well as text communication methods enabling different communication mediums. Our platform also improves patient engagement as it enables collection of their physical health data through wearable health monitoring devices that seamlessly integrate with our platform. This integration with biometric devices improves efficiency in data recording.

3.1 Challenges

Our proposed approach has the potential to support care models that deliver better health outcomes. However, the successful execution requires understanding of the challenges involved in online PHR integration. The challenges are -

- **Adoption:** The adoption rate for mobile health applications for health monitoring and health tracking for self-management among healthy population is increasing over the last decade. [12]. However, the Australian and worldwide research shows that the adoption rate for online PHR based solution among adults with chronic disease conditions is less than expected [13, 14, 15]. Our proposed solution aims to improve the uptake rate by providing a simple and yet highly effective solution.

The challenges in adoption of online PHR mechanisms are not only applicable to the patients but clinicians and carers as well.

Table 3. Adoption Challenges by online PHR users

Online PHR User Category	**Challenges**
Chronic Disease Patients	IT Literacy, Technology device preferences, Ability to record and understand own data (health literacy)

Online PHR User Category	Challenges
Families	IT Literacy, Interpretation of data
Primary and Specialist Clinicians	Quality of concern Patient recorded data for treatment decisions, Time to access patient reported data in online PHR
Nursing Staff	Data Quality, Time to access patient reported data in online PHR
Allied Health clinicians (E.g. Physiotherapists)	Quality concern of Patient recorded data, Lack of instant physical interaction with men
Care coordinators	Quality concern of Patient recorded data for care plan development and implementation

- **Data Quality**: The data quality in the PHR should be endorsed by the clinicians.
- **Cost-effectiveness:** The evidence suggesting online PHR as a cost-effective tool to manage personal health information is not clearly established [16, 17, 18, and 19].
- **Health Outcomes:** There is also no clear evidence about better health outcomes due to online PHR [20].

The above challenges can be addressed through a careful implementation of the proposed platform. The implementation of the proposed platform is under progress and the evaluation of the platform will be undertaken in a randomised clinical trial settings. The proposed approach has received positive feedback from the patient as well as clinical community representatives.

4 Conclusion

This research has made an attempt to address the current issues of disperse personal health data sources. The proposed platform aims to provide a single view of the patient's PHR that can empower chronic disease patients and improve collaboration with the clinicians for self-managed care. The proposed platform will be implemented and evaluated as this research progresses in future.

Limitations

The author acknowledges that the research is a work-in-progress. The proposed platform is not evaluated with a sample data yet. The concepts proposed in this paper are specified strategic perspective. The research does not provide real-life evaluation of the proposed PHR platform as well as core design of the intelligent PHR.

Acknowledgement

This research is mainly supported by Movember Foundation as a part of its survivorship initiative. This research is also supported by the Australian e-Health Research Centre, CSIRO as a project collaborator. The author would like to thank Movember Foundation and the Australian e-Health Research Centre, CSIRO for their support and this opportunity.

References

1. Tang, Paul; Ash, Joan; Bates, David; Overhage, J.; Sands, Daniel (2006). "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption". JAMIA 13 (2): 121–126. doi:10.1197/jamia.M2025
2. National e-Health Transition Authority (NEHTA), http://www.nehta.gov.au/ehealth-implementation/what-is-a-pcehr, Viewed 19 February 2013.
3. Markle Foundation Personal Health Working Group, Connecting for Health: A Public-Private Collaborative. New York, NY Markle Foundation2003;
4. Yamin CK, Emani S, Williams DH, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011;171:568–74
5. Irini Genitsaridi, Haridimos Kondylakis, Lefteris Koumakis, Kostas Marias, Manolis Tsiknakis, Towards Intelligent Personal Health Record Systems: Review, Criteria and Extensions, Procedia Computer Science, Volume 21, 2013, Pages 327-334.
6. Luo G. Open issues in intelligent personal health record - an updated status report for 2012. J Med Syst. 2013 Jun;37(3):9943.
7. National e-Health Transition Authority (NEHTA), http://www.nehta.gov.au/ehealth-implementation/what-is-a-pcehr, Viewed 19 February 2013.
8. NHS Healthspace, http://www.connectingforhealth.nhs.uk/systemsandservices/healthspace, Viewed 20 February 2013.
9. NHS Information Strategy, http://informationstrategy.dh.gov.uk/, Viewed 20 February 2013.
10. US MYPHR, http://www.myphr.com/resources/choose.aspx, Viewed 21 February 2013
11. Big Blue Button, http://www.va.gov/BLUEBUTTON/index.asp, Viewed March 2013, Viewed 21 February 2013.
12. Big Button Plus, http://bluebuttonplus.org/index.html, Viewed 2 April 2013.
13. NEHTA Compliant Product Registers, https://epipregister.nehta.gov.au/registers, Viewed 25 February 2013.

14. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online. 2011; 10:24.
15. Australian PCEHR Registration Targets, Pulse IT Magazine Article, http://www.pulseitmagazine.com.au/index.php?option=com_content&view=article&id=1317:half-a-million-pcehr-registrations-still-achievable-doha&catid=16:australian-ehealth&Itemid=328\\, Viewed 22 February 2013
16. A. Sunyaev, D. Chomyi, C. Mauro, and H. Krcmar, "Evaluation Framework for Personal Health Records: Microsoft HealthVault vs. Google Health," in Proceedings of the Hawaii International Conference on System Sciences (HICSS 43), Kauai, Hawaii, 2010.
17. Executive Summary Integrated Cancer Care Programme, Report by UnitedHealth Europe.
18. Evaluation of Phase 1 of the One-to-One, Support Implementation Project, Baseline Report Macmillan Cancer Support
19. Imison H, Gorman P, Woods S, et al. Barriers and Drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep/Technol Assess; 175 (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290–02-0024). AHRQ Publication No. 09-E004. Rockville, MD: Agency for Healthcare Research and Quality. Nov 2008.
20. Kim Eung-Hun, Stolyar Anna, Lober William B, Herbaugh Anne L, Shinstrom Sally E, Zierler Brenda K, Soh Cheong B, Kim Yongmin. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res. 2009; 11(4):e44.

Partially automated literature screening for systematic reviews by modelling non-relevant articles

Henry Petersen[1] and Josiah Poon[1] Simon Poon[1] Clement Loy[2] Mariska Leeflang[3]

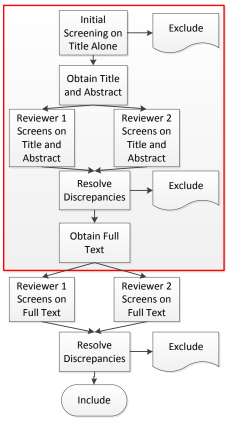
[1] School of Information Technologies, University of Sydney, Australia
[2] School of Public Health, University of Sydney, Australia
[3] Academic Medical Center, University of Amsterdam, Netherlands
hpet9515@uni.sydney.edu.au, {josiah.poon,simon.poon}@sydney.edu.au,
clement.loy@sydney.edu.au, m.m.leeflang@amc.uva.nl

Systematic reviews are widely considered as the highest form of medical evidence, since they aim to be a repeatable, comprehensive, and unbiased summary of the existing literature. Because of the high cost of missing relevant studies, review authors go to great lengths to ensure all relevant literature is included. It is not atypical for a single review to be conducted over the course of months or years, with multiple authors screening thousands of articles in a multi-stage triage process; first on title, then on title and abstract, and finally on full text. Figure 1a shows a typical literature screening process for systematic reviews.

In the last decade, the information retrieval (IR) and machine learning (ML) communities have shown increasing interest in literature searches for systematic reviews [1–3]. Literature screening for systematic reviews can be characterised as a classification task with two defining features; a requirement for near perfect recall on the class of relevant studies (the high cost of missing relevant evidence), and highly imbalanced training data (review authors are often willing to screen thousands of citations to find less than 100 relevant articles). Previous attempts at automating literature screening for systematic reviews have primarily focused on two questions; how to build a suitably high recall model for the target class in a given review under the conditions of highly imbalanced training data [1, 3], and how best to integrate classification into the literature screening process [2].

When screening articles, reviewers exclude studies for a number of reasons (animal populations, incorrect disease etc.). Additionally, in any given triage stage a study may not be relevant but still progress to the next stage as the authors have insufficient information to exclude it (i.e. the title may not indicate a study was performed with an animal population, however this may become apparent upon reading the abstract). We meet the requirement for near perfect recall on relevant studies by inverting the classification task and identifying subsets of irrelevant studies with near perfect precision. We attempt to identify such studies by training the classifier using the labels assigned at the previous triage stage (see Figure 1c). The seamless integration with the existing manual screening process is an advantage of our approach.

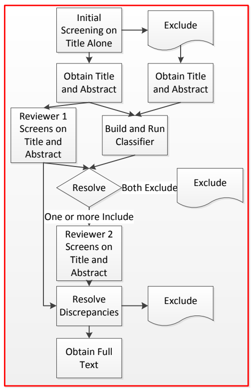
The classifier is built by first selecting terms from the title and abstracts with the greatest information gain on labels assigned in the first triage stage. Articles



(a)

- 'neutropenia', but not 'infection' or 'thorax'
- 'skin' but not 'thorax'
- 'immunoglobulin_g'
- 'animals'
- 'drug_therapy', but not 'risk' or 'infection'

(b)



(c)

Fig. 1: Typical literature screening process for systematic reviews, sample rules generated by our classifier, and the proposed modified screening process.

are then represented as Boolean statements over these terms, and interpretable rules are then generated using Boolean minimisation (examples of rules are given in 1b Review authors can then refine the classifier by selecting only those rules most likely to describe non-relevant studies, maximising overall precision.

Preliminary experiments simulating the process outlined in Figure 1c on a previously conducted systematic review indicate that as many as 25% of articles can be safely eliminated without the need for screening by a second reviewer. The evaluation does assume that all false positives (studies erroneously excluded by the generated rules) were included by the first reviewer. Such an assumption is reasonable; the reason for multiple reviewers is that even human experts make mistakes. A study comparing the precision of our classifier to human reviewers is planned. In addition, future work will focus on improving the quality of the generated rules by trying to better capture reasons for excluding studies matching those used by human reviewers.

References

1. Aaron M. Cohen, Kyle H. Ambert, and Marian McDonagh. Research paper: Cross-topic learning for work prioritization in systematic review creation and update. *JAMIA*, 16(5):690–704, 2009.
2. Oana Frunza, Diana Inkpen, Stan Matwin, William Klement, and Peter OBlenis. Exploiting the systematic review protocol for classification of medical abstracts. *Artificial Intelligence in Medicine*, 51(1):17 – 25, 2011.
3. Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, and Peter O'Blenis. A new algorithm for reducing the workload of experts in performing systematic reviews. *JAMIA*, 17(4):446–453, 2010.

Optimizing Shiftable Appliance Schedules across Residential Neighbourhoods for Lower Energy Costs and Fair Billing

Salma Bakr and Stephen Cranefield

Department of Information Science, University of Otago, Dunedin, New Zealand
salma.bakr@postgrad.otago.ac.nz, scranefield@infoscience.otago.ac.nz

Abstract. This early stage interdisciplinary research contributes to smart grid advancements by integrating information and communications technology and electric power systems. It aims at tackling the drawbacks of current demand-side energy management schemes by developing an agent-based energy management system that coordinates and optimizes neighbourhood-level aggregate power load. In this paper, we report on the implementation of an energy consumption scheduler for rescheduling "shiftable" household appliances at the household-level; the scheduler takes into account the consumer's time preferences, the total hourly power consumption across neighbouring households, and a fair electricity billing mechanism. This scheduler is to be deployed in an autonomous and distributed residential energy management system to avoid load synchronization, reduce utility energy costs, and improve the load factor of the aggregate power load.

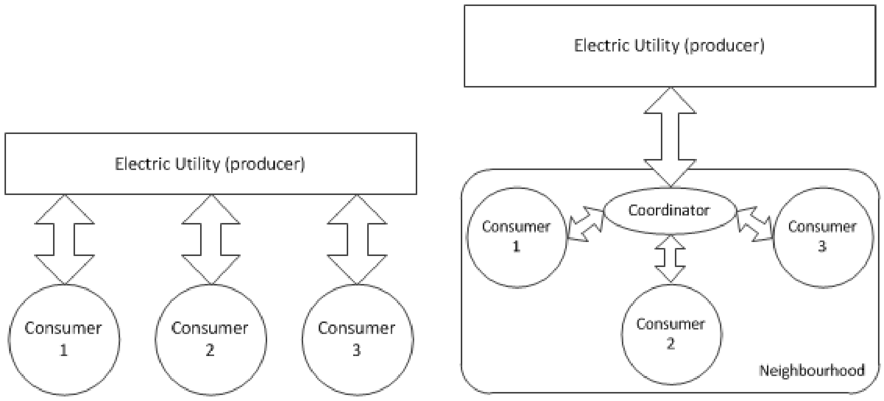
1 Introduction

Electric utilities tend to meet growing consumer energy demand by expanding their generation capacities, especially capital-intensive peak power plants (also known as "peakers"), which are much more costly to operate than base load power plants. As this strategy results in highly inefficient consumption behaviours and under-utilized power systems, demand-side energy management schemes aiming to optimally match power supply and demand have emerged.

Currently deployed demand-side energy management schemes are based on the interactions between the electric utility and a single household [18], as in Fig.1(a). As this approach lacks coordination among neighbouring households sharing the same low-voltage distribution network, it may cause load synchronization problems where new peaks arise in off-peak hours [15]. Thus, it is essential to develop flexible and scalable energy management systems that coordinate energy usage between neighbouring households, as in Fig.1(b).

2 Background

The smart grid, or the modernized electric grid, is a complex system comprising a number of heterogeneous control, communication, computation, and electric



(a) (b)

Fig. 1. The interactions between the utility and the consumers in demand-side energy management schemes are either: (a) individual interactions, or (b) neighbourhood-level interactions

.

power components. It also integrates humans in decision making. To verify the states of smart grid components in a simultaneous manner and take human intervention into account, it is necessary to adopt autonomous distributed system architectures whose functionality can be modelled and verified using agent-based modelling and simulation.

Multi-agent systems (MAS) provide the properties required to coordinate the interactions between smart grid components and solve complex problems in a flexible approach. In the context of a smart grid, agents represent producers, consumers, and aggregators at different scales of operation, e.g. wholesale and retail energy traders [7]. MAS have been deployed in a number of smart grid applications, with a more recent focus on micro-grid control [6, 17] and energy management [10, 12] especially due to the emerging trend of integrating distributed energy resources (DER), storage capacities, and plug-in hybrid electric vehicles (PHEV) into consumer premises.

In agent-based energy management systems, agents may aim at achieving a single objective or a multitude of objectives; typical objectives include: balancing energy supply and demand [4]; reducing peak power demand [13, 16]; reducing utility energy costs [8, 16] and consumer bills [16]; improving grid efficiency [4]; and increasing the share of renewable energy sources [1, 12] which consequently reduces the carbon footprint of the power grid. Agent objectives can be achieved using evolutionary algorithms [8] or a number of optimization techniques such as integer, quadratic [5, 13], stochastic [4] and dynamic programming [5]. As for the interactions among agents, game theory provides a conceptual and a formal analytical framework that enables the study of those complex interactions [19].

3 Research Objectives

This research aims at optimizing the energy demand of a group of neighbouring households, to reduce utility costs by using energy at off-peak periods, avoid load synchronization that may occur due to rescheduling appliance usage, and improve the load factor (i.e. the ratio between average and peak power) of the aggregate load. A number of energy consumption schedulers have been proposed in the literature [14, 16, 21]; however, those schedulers do not leverage an accurately quantified and fair billing mechanism that charges consumers based on the shape of their power load profiles and their actual contribution in reducing energy generation costs for electric utilities [3]. In this paper, we implement and evaluate an energy consumption scheduler that optimizes the operation times of three wet home appliances and a PHEV at the household-level based on the total hourly power consumption across neighbouring households, consumer time preferences, and a fair electricity billing mechanism.

4 Methodology

We use the findings of Baharlouei et al. [3] to resolve a gap in the findings of Mohsenian-Rad et al. [16]. Game-theoretic analysis is used by Mohsenian-Rad et al. [16] to propose an incentive-based *energy consumption game* that schedules "shiftable" home appliances (e.g. washing machine, tumble dryer, dish washer, and PHEV) for residential consumers (players) according to their daily time preferences (strategies); at the Nash equilibrium of the proposed non-cooperative game, it is shown that the energy costs of the system are minimized. However, this game charges consumers based on their total daily electric energy consumption rather than their hourly energy consumption. In other words, two consumers having the same total daily energy consumption are charged equally even if their hourly load profiles are different. This unfair billing mechanism may thus discourage consumer participation as it does not take consumer rescheduling flexibility into consideration. With this in mind, we propose leveraging the fair billing mechanism recently proposed by Baharlouei et al. [3] to encourage consumer participation in the energy consumption game.

5 Energy Consumption Scheduler

5.1 Formulation

Assuming a multi-agent system for managing electric energy consumption at the neighbourhood-level, where agents represent consumers, each agent locally and optimally schedules his "shiftable" home appliances to minimize his electricity bill taking into account the following inputs: appliance load profiles, consumer time preferences, grid limitations (if any), aggregate scheduled hourly energy consumption of all the other agents in the neighbourhood, and the deployed electricity billing scheme. If the energy cost function is non-linear, knowing the aggregate scheduled load is required for optimization.

After this optimization, each agent sends out his updated appliance schedule to an aggregator agent, which then forwards the aggregated load to the other agents to optimize their schedules accordingly. By starting with random initial schedules, convergence of the distributed algorithm is guaranteed if household-level energy schedule updates are asynchronous [16]. The electric utility may coordinate such updates according to any turn-taking scenario.

We assume electricity distributed to the neighbourhood is generated by a thermal power generator having a quadratic hourly cost function [23] given by (1); as this equation is convex, quadratic, and has linear constraints, it can be solved using mixed integer quadratic programming.

$$C_h\left(L_h\right) = a_h L_h^2 + b_h L_h + c_h, \tag{1}$$

where $a_h > 0$, and b_h, $c_h \geq 0$ at each hour $h \in H = [1, ..., 24]$. In (2), L_h and x_m^h denote the total hourly load of N consumers and consumer m, respectively [16].

$$L_h = \sum_{m=1}^{N} x_m^h, \tag{2}$$

To encourage participation in energy management programmes, it is essential to reward consumers with fair incentives. By rescheduling appliances to off-peak hours where electricity tariffs are cheaper, we save on utility energy costs and consequently impose monetary incentives for consumers in the form of savings on electricity bills. The optimization problem therefore targets the appliance schedule x_n^h that results in the minimum bill B_n for consumer (agent) n. The billing equation proposed by Baharlouei et al. [3], which fairly maps a consumer's bill to energy costs (1), is given by (3).

$$B_n = \sum_{h=1}^{H} \frac{x_n^h}{\sum_{m=1}^{N} x_m^h} C_h\left(\sum_{m=1}^{N} x_m^h\right), \tag{3}$$

5.2 Set-up

To model the optimization problem such that each agent n individually and iteratively minimizes (3), we use YALMIP — an open-source modelling language that integrates with MATLAB. We consider a system of three households (agents) and investigate the behaviour of one of those schedulers with respect to fair billing, lower energy costs, and improved load factor. To model consumer flexibility in scheduling, we consider two scenarios for the same household where the consumer's acceptance of rescheduling flexibility differ. We investigate the two scenarios for four days in December, March, June and September.

To test our energy consumption scheduler, we choose to schedule a PHEV and three wet appliances: a clothes washer, a tumble dryer, and a dish washer. Wet appliance power load profiles are based on survey EUP14-07b [22], which was conducted with around 2500 consumers from 10 European countries. For the PHEV load, we use the power load profile of a mid-size sedan at 240V–30A [9].

We choose a budget-balanced billing system and calibrate the coefficients of the hourly energy cost function (1) against a three-level time-of-use pricing scheme used by London Hydro [11], where the kilowatt-hour is charged at 12.4, 10.4, and 6.7 cents for on-, mid-, and off-peak hours, respectively. Energy consumption of neighbouring households and non-shiftable loads of the household investigated are taken from a publicly available household electric power consumption data set [2], for the period from December 2006 to September 2007.

5.3 Scenario 1

In this scenario, we assume the consumer is not flexible about appliance scheduling and use common startup times: clothes washing starts at 7 a.m., drying starts two hours directly after washing starts, dish washing starts at 6 p.m. [22], and PHEV recharging starts at 6 p.m. [20].

5.4 Scenario 2

The consumer is assumed to be flexible about appliance scheduling in Scenario 2; clothes washing starts any time between 6 a.m. and 9 a.m., drying any time after washing but before 11 p.m., washing dishes any time after 7 p.m, but before 11 p.m., and PHEV recharging any time after 1 a.m. but before 5 a.m.

6 Results

6.1 Fair Billing

Results indicate that the consumer's electricity bill for operating household "shiftable" appliances in Scenario 2 is lower by 70%, 57%, 32%, and 65% compared to that in Scenario 1 for the days chosen in December, March, June, and September, respectively. This clearly indicates that flexibility is awarded fairly through the deployed billing mechanism. Figures 2 and 3 depict the appliance schedules resulting in the minimum bill for the household under investigation and the aggregate non-shiftable load of neighbouring households, for Scenario 1 and 2 in December, respectively.

6.2 Lower Energy Costs

As we chose a budget-balanced billing system and since appliances are rescheduled to cheaper off-peak hours, utility energy costs are lower in Scenario 2 by 70%, 57%, 32%, and 65% compared to that in Scenario 1, for the days chosen across the four seasons, respectively.

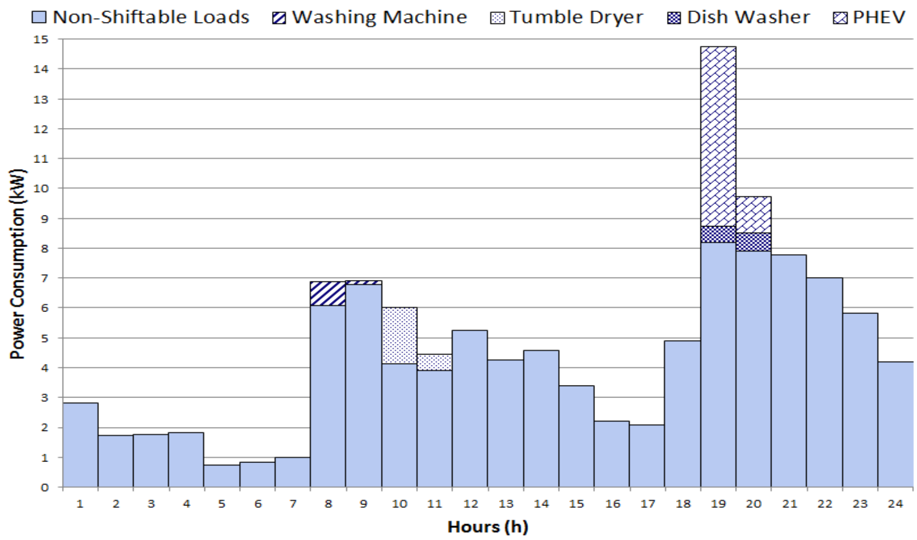


Fig. 2. Scenario 1: the unscheduled “shiftable” appliance loads of the consumer under investigation and the aggregate “non-shiftable” neighbourhood-level loads (December)
1

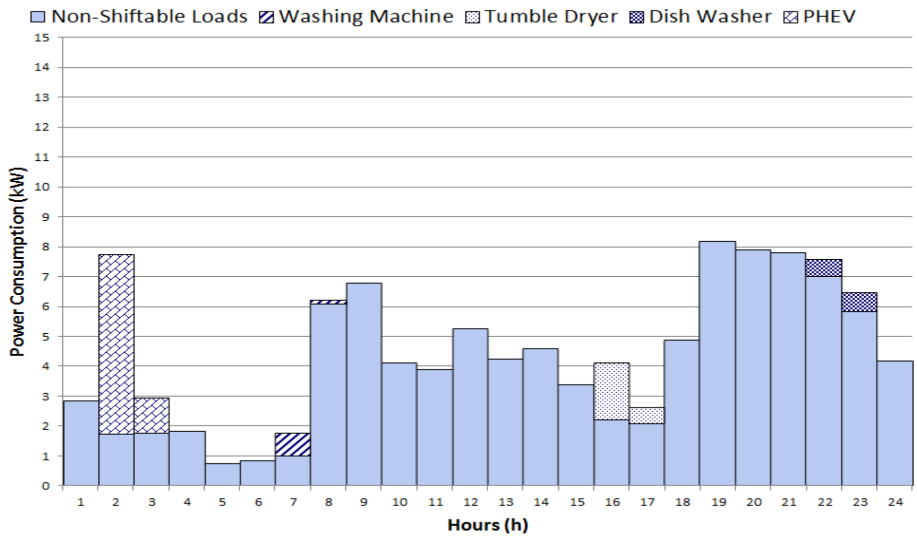


Fig. 3. Scenario 2: the scheduled “shiftable” appliance loads of the consumer under investigation and the aggregate “non-shiftable” neighbourhood-level loads (December)

6.3 Improved Load Factor

As the “shiftable” appliances of the household under investigation are rescheduled to operate during off-peak hours instead of peak hours, the load factor of the aggregate load in Scenario 2 is improved by 44%, 13%, 19%, and 28% compared to that in Scenario 1, for the days chosen across the four seasons, respectively. This indicates improved resource allocation in the power grid.

7 Conclusion

In this paper, we leverage the fair billing mechanism proposed by Baharlouei et al. [3] to evaluate the energy consumption scheduling game proposed by Mohsenian-Rad et al. [16]. We have implemented and evaluated a scheduler that optimally allocates the operation of "shiftable" appliances for a consumer based on his time preferences, the aggregate hourly "non-shiftable" load at the neighbourhood-level, and a fair billing mechanism. As the deployed billing mechanism takes advantage of cheaper off-peak electricity prices, we show that it helps in lowering utility energy costs and electricity bills, and improving the load factor of the aggregate neighbourhood-level power load. We also conclude that consumer flexibility in rescheduling appliances is rewarded fairly based on the shape of his power load profile rather than his total energy consumption.

8 Future Work

Eventually, we intend to investigate an appliance scheduler that coordinates electric energy consumption among a large number of households (agents).

References

1. Baños, R., Manzano-Agugliaro, F., Montoya, F.G., Gil, C., Alcayde, A., Gómez, J.: Optimization Methods Applied to Renewable and Sustainable Energy: A Review. Renewable and Sustainable Energy Reviews 15(4), 1753–1766 (2011)
2. Bache, K., Lichman, M.: UCI machine learning repository (2013), http://archive.ics.uci.edu/ml
3. Baharlouei, Z., Hashemi, M., Narimani, H., Mohsenian-Rad, A.H.: Achieving Optimality and Fairness in Autonomous Demand Response: Benchmarks and Billing Mechanisms. IEEE Transactions on Smart Grid 4(2), 968–975 (2013)
4. Clement, K., Haesen, E., Driesen, J.: Coordinated Charging of Multiple Plug-in Hybrid Electric Vehicles in Residential Distribution Grids. In: IEEE PES Power Systems Conference and Exposition (2009), http://dx.doi.org/10.1109/PSCE.2009.4839973
5. Clement-Nyns, K., Haesen, E., Driesen, J.: The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid. IEEE Transactions on Power Systems 25(1), 371–380 (2010)
6. Colson, C.M., Nehrir, M.H.: Algorithms for Distributed Decision Making for Multi Agent Microgrid Power Management. In: IEEE Power and Energy Society General Meeting (2011), http://dx.doi.org/10.1109/PES.2011.6039764
7. Gammon, R.: CASCADE Overview. In: Sustainable Energy, Complexity Science and the Smart Grid, Satellite Workshop in the European Conference on Complex Systems (2012)

8. Guo, Y., Li, J., James, G.: Evolutionary Optimisation of Distributed Energy Resources. In: 18th Australian Joint Conference on Advances in Artificial Intelligence. vol. 3809, pp. 1086–1091. Springer Berlin Heidelberg (2005)
9. Hadley, S.W.: Impact of Plug-in Hybrid Vehicles on the Electric Grid. Tech. Rep. ORNL/TM-2006/554, Oak Ridge National Laboratory (2006), http://web.ornl.gov/info/ornlreview/v40_2_07/2007_plug-in_paper.pdf
10. Lamparter, S., Becher, S., Fischer, J.: An Agent Based Market Platform for Smart Grids. In: International Conference on Autonomous Agents and Multiagent Systems. pp. 1689–1696. IFAAMAS (2010)
11. London Hydro: Regulated Price Plan Time-of-Use Rates (2013), http://www.londonhydro.com/residential/electricityrates/
12. Mets, K., Strobbe, M., Verschueren, T., Roelens, T., De Turck, F., Develder, C.: Distributed Multi-agent Algorithm for Residential Energy Management in Smart Grids. In: IEEE Network Operations and Management Symposium. pp. 435–443 (2012), http://dx.doi.org/10.1109/NOMS.2012.6211928
13. Mets, K., Verschueren, T., Haerick, W., Develder, C., Turck, F.D.: Optimizing Smart Energy Control Strategies for Plug-in Hybrid Electric Vehicle Charging. In: IEEE/IFIP Network Operations and Management Symposium Workshops. pp. 293–299 (2010), http://dx.doi.org/10.1109/NOMSW.2010.5486561
14. Mohsenian-Rad, A.H., Wong, V., Jatskevich, J., Schober, R.: Optimal and Autonomous Incentive-based Energy Consumption Scheduling Algorithm for Smart Grid. In: IEEE PES Conference on Innovative Smart Grid Technologies (2010), http://dx.doi.org/10.1109/ISGT.2010.5434752
15. Mohsenian-Rad, A.H., Leon-Garcia, A.: Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environments. IEEE Transactions on Smart Grid 1(2), 120–133 (2010)
16. Mohsenian-Rad, A.H., Wong, V.W.S., Jatskevich, J., Schober, R., Leon-Garcia, A.: Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid. IEEE Transactions on Smart Grid 1(3), 320–331 (2010)
17. Pipattanasomporn, M., Feroze, H., Rahman, S.: Multi-Agent Systems in a Distributed Smart Grid: Design and Implementation. In: IEEE PES Power Systems Conference and Exposition (2009), http://dx.doi.org/10.1109/PSCE.2009.4840087
18. Ruiz, N., Cobelo, I., Oyarzabal, J.: A Direct Load Control Model for Virtual Power Plant Management,. IEEE Transactions on Power Systems 24(2), 959–966 (2009)
19. Saad, W., Han, Z., Poor, H.V., Baar, T.: Game-Theoretic Methods for the Smart Grid: An Overview of Microgrid Systems, Demand-Side Management, and Smart Grid Communications. IEEE Signal Processing Magazine 29(5), 86–105 (2012)
20. Shao, S., Pipattanasomporn, M., Rahman, S.: Challenges of PHEV Penetration to the Residential Distribution Network. In: IEEE Power and Energy Society General Meeting (2009), http://dx.doi.org/10.1109/PES.2009.5275806
21. Shinwari, M., Youssef, A., Hamouda, W.: A Water-Filling Based Scheduling Algorithm for the Smart Grid. IEEE Transactions on Smart Grid 3(2), 710–719 (2012)
22. Stamminger, R., Friedrich-Wilhelms, R.: Synergy Potential of Smart Appliances. Deliverable DP2.3 of WP2, EIE Project on Smart Domestic Appliances in Sustainable Energy Systems (2008), http://www.smart-a.org/D2.3_Synergy_Potential_of_Smart_Appliances_4.00.pdf
23. Wood, A., Wollenberg, B.: Power Generation, Operation, and Control. Wiley-Interscience, 2 edn. (1996)

Proposal of information provision to probe vehicles based on distribution of link travel time that tends to have two peaks

Keita Mizuno, Ryo Kanamori, and Takayuki Ito

Nagoya Institute of Technology,
Gokiso, Showa, Nagoya 466-8555, JAPAN
mizuno.keita@itolab.nitech.ac.jp,
kanamori.ryo@nitech.ac.jp,
ito.takayuki@nitech.ac.jp
http://www.itolab.nitech.ac.jp/itl2/page_en/

Abstract. In most cities, traffic congestion is a primary problem that must be tackled. Traffic control/operation systems based on information gathered from probe vehicles have attracted a lot of attention. In this paper, we examine provision of travel information to eliminate traffic jams. Although it is conventional to provide the mean of historical accumulated data, we introduce the 25th percentile and 75th percentile values because a distribution of link travel time tends to have two peaks. As a result, the proposed method reduced travel time of vehicles compared with the conventional method.

Keywords: Traffic management, Probe car, Intelligent Transport System

1 Introduction

Automobile traffic jams have become a major problem in many cities of the world. In Japan, an increase in vehicle emissions and time loss due to traffic congestion have also become significant problems. As a solution to these problems, information collected from probe vehicles is attracting attention. In this research, we assume an environment in which information of the travel time of a vehicle in the past can be obtained, vehicles can communicate mutually, and vehicles can share traffic conditions to reduce the travel time of all vehicles. Thus, we propose a method of providing information to a probe vehicle for reducing travel time of regular vehicles, and show the effectiveness of the proposed method by simulation experiments.

In this research, we focus on how a distribution of link travel time tends to have two peaks for historical accumulated data of travel time of the vehicle. In addition to the mean of historical accumulated data of the link travel time, using the 25th percentile value and 75th percentile value of historical accumulated data, we perform path finding and give information to the probe vehicle. Furthermore, to demonstrate that the proposed method of this research is effective, we implement traffic flow simulation based on the cell transmission model[1][2], and we perform vehicle movement simulation of the conventional method and proposed method. We use travel time of the vehicle, which has also been used in conventional research, for the effect analysis of information provided to the probe vehicle. In addition, we examine the difference between the time taken to move in the simulation and travel time to the destination that is expected from the historical accumulated data of the vehicle.

The remainder of this paper is organized as follows. Background and purpose of this research are presented in chapter 2, and the distribution of link travel time having two peaks is discussed in chapter 3. We describe the proposed information provision method in chapter 4, the vehicle simulation in chapter 5, and the effectiveness of the proposed method, along with future work in chapter 6.

2 Background and purpose

In this chapter, we describe the background and purpose of this research. Personal vehicles have become an essential means of transportation for many people. However, there are many problems we must solve; for example, decline in economic efficiency due to traffic congestion, global environmental degradation such as global warming and air pollution, and many traffic accidents. Transportation and traffic account for about 20% of carbon dioxide emissions in Japan, and of that, vehicles account for about 90%[3]. Figure 1 is a diagram showing the relationship between carbon dioxide emissions and the running speed of a vehicle. Because we can see that the carbon dioxide emissions from the vehicle decrease when running speed of the vehicle increases, we must decrease carbon dioxide emissions by eliminating traffic congestion, and increasing the running speed of the vehicle. Also, there are approximately 5 billion hours per year in time lost to congestion in Japan, and the economic loss is 11 trillion yen. Problems caused by traffic congestion have clearly become serious in Japan, as in many other parts of the world, and it is necessary to resolve these issues.

In addition to the promotion of next-generation vehicles such as electric cars as a way to solve these problems, traffic operation and management measures by Intelligent Transport Systems (ITS), such as providing path information and road pricing, have attracted attention. The number of vehicles with vehicle perception and navigation systems (probe vehicles) is increasing, and technology of information collection and provision has also advanced in route search information. Further, from the historical accumulated data collected from the probe vehicle, it is observed that a distribution of link travel time tends to have two peaks.

About providing information to the probe vehicle, Kanamori et al.[4] simulated providing information to a probe vehicle using not only the historical accumulated data collected from the probe vehicle but also predicting the traffic situation. Morikawa et al.[5] simulated providing information to a probe vehicle using the number of right and left turns in the path to the destination, in addition to the historical accumulated data collected from the probe vehicle.

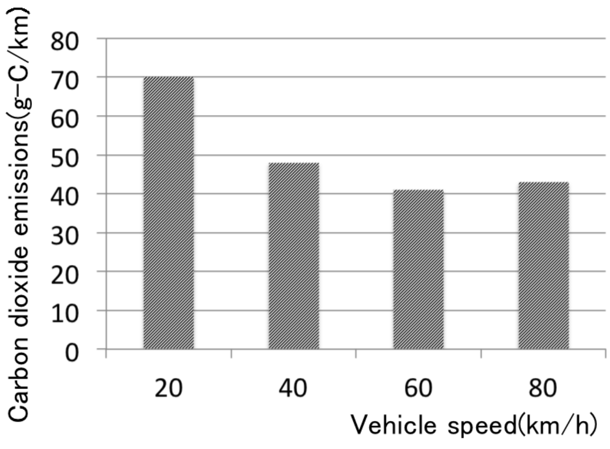


Fig. 1. Relationship between carbon dioxide emissions and running speed of vehicle

In researches of Kanamori et al. and Morikawa et al., they simulated providing information that uses the mean of historical accumulated data collected from probe vehicles, and searches for a route to a destination.

The purpose of this research is to propose a method to use historical accumulated data focusing on the distribution of link travel time, which tends to have two peaks, and reducing travel link time of vehicles in the simulation.

3 Distribution of link travel time

In this section, we discuss how a distribution of link travel time tends to have two peaks. Link travel time of the vehicle described in this research is the time to travel from one intersection to another.

Figure 2 shows example of distribution of link travel time. It is observed that a distribution of link travel time tends to have two peaks when the vehicles pass through the intersection, and simulations that reproduce a distribution of link travel time have been researched[6].

The cause of the link travel time of the vehicle having two peaks is, for example, a traffic signal. When the vehicle passes through an intersection, a considerable difference occurs because the vehicle stops at the signal or doesn't stop. In previous research, they didn't consider that a distribution of link travel time tends to have two peaks; instead, they used the mean value of the link travel time collected from the probe vehicle.

4 Information provision to probe vehicles

In this chapter, we provide a detailed description of the method of information provision to the probe vehicle in this research. As usage of the historical accu-

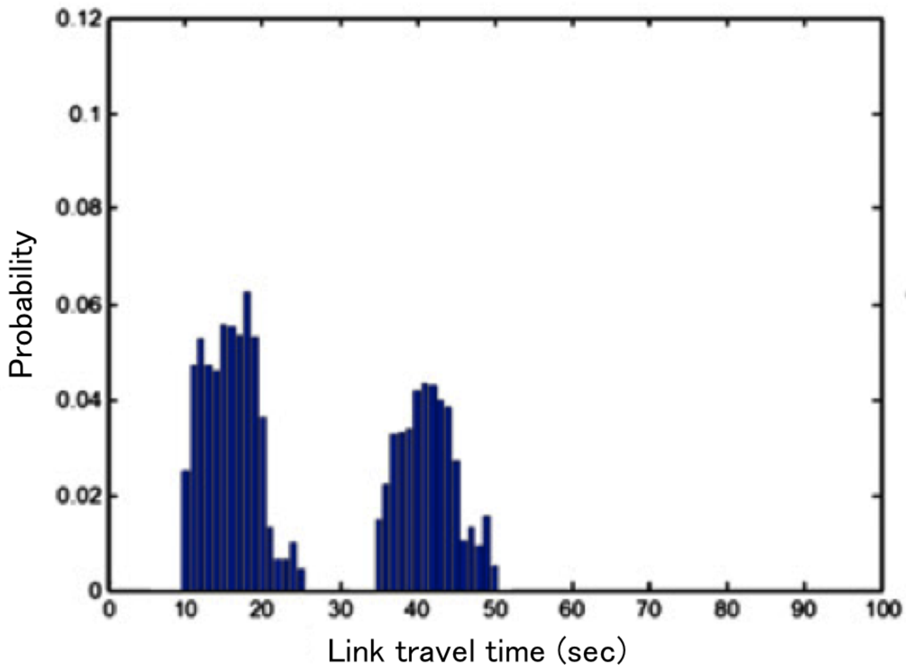


Fig. 2. Distribution of link travel time

mulated data of link travel time for searching the route to the destination, in addition to a conventional method to provide the mean of historical accumulated data of the travel time, we introduce provisions of the 25th percentile value and 75th percentile value of historical accumulated data of the travel time in this research.

Probe vehicle assumed in this paper is sending information of link travel time and receiving information of path to destination with least travel time. Information of path to destination with least travel time is predicted by link travel time collected from probe vehicle.

In this experiment, we use the data of the 25th percentile and 75th percentile values of the historical accumulated data of link travel time. To decide which value we will use in this research, we conduct a preliminary experiment. First, we used only the 25th percentile value of the historical accumulated data in the information-providing simulation. Second, we used only the 75th percentile value of the historical accumulated data in the information-providing simulation. We compared the mean of historical accumulated data of the link travel time with 25th percentile and 75th percentile values regarding the travel time of the vehicle. In this research, assuming the differences of factors such as the number of intersections passed through depending on the travel distance of the vehicle, we compare the mean value, 25th percentile and 75th percentile values by travel distance of each vehicle.

We set the travel distance of vehicles using the 25th percentile or 75th percentile values in the simulation, and conduct information provision simulation using the 25th percentile and 75th percentile values for searching the route to the destination.

5 Simulation for evaluation

5.1 Settings of simulation

We use the data of Kichijoji and Mitaka that are provided in the traffic simulation clearing house as a road network used for the evaluation experiment in this research. The traffic simulation clearing house[7] is an institution providing various data for validation. The network is composed of 57 nodes and 137 links. Vehicles in the simulation number about 17,000 units, and approximately 50% are probe vehicles in this experiment. Further, in order to accumulate link travel time for the vehicles to be used for route search, the simulation was repeated about 30 times. Figure 3 is a network diagram from Kichijoji and Mitaka that is used for the simulation in this research.

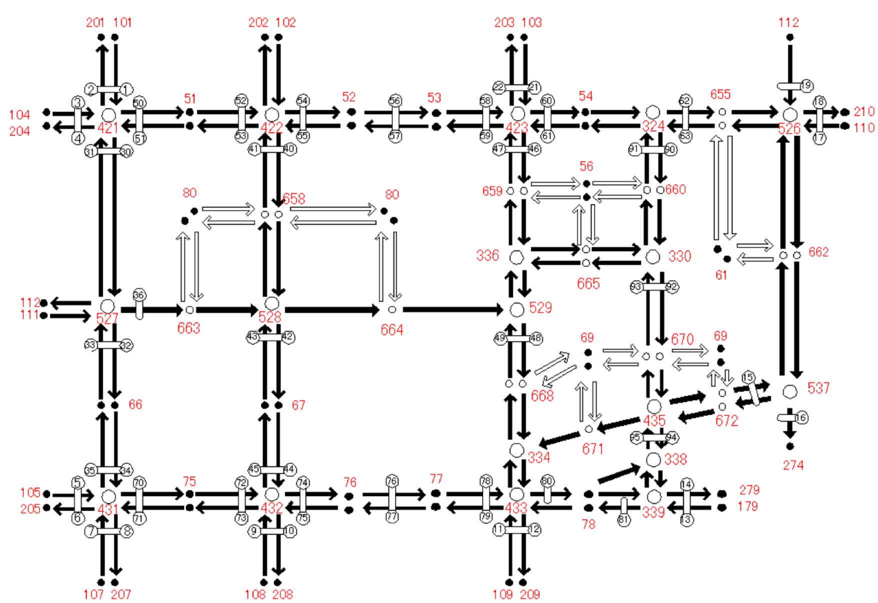


Fig. 3. Network of Kichijoji and Mitaka

Table 1 shows survey contents collected by Kichijoji and Mitaka. Investigation time is set to a high-traffic period.

Since the investigation data contain the times each vehicle entered and exited the network, we can obtain the travel time to the destination of each vehicle.

Table 1. Details of survey of Kichijoji and Mitaka

investigation time	AM 7:00 AM 10:00
target area	Mitaka and Musashino, Tokyo
observation points	70
target vehicle	four wheel vehicles
survey contents	passage time
	vehicle number
	car model(bus, taxi, and other)

5.2 Traffic flow simulation

In this research, we implemented a traffic flow simulation based on the cell transmission model, in which the repeatability of travel time is high and we can control the route choice of the vehicle in the simulation. The cell transmission model is a model that divides the network links into cells and controls the movement of vehicles by the density of vehicles in a cell.

$$y_i(t) = min\{ n_{i-1}(t), Q_i(t), N_i(t) - n_i(t)\} \tag{1}$$

- $y_i(t)$: number of vehicles moving to the cell of index i at time t
- $Q_i(t)$: maximum number of vehicles that can flow into the cell of index i at time t
- $N_i(t)$: maximum number of vehicles in the cell of index i at time t
- $n_i(t)$: number of vehicles in the cell of index i at time t

Equation (1) represents the number of vehicles to move between cells on the cell transmission model. The number of vehicles that can move to the next cell is determined by the smallest number of the following: number of vehicles in the present cell, the amount of empty space in the next cell, or maximum number of vehicles that can flow into the next cell. Equation (2) represents traffic flow rate.

$$q = k * v \tag{2}$$

- q: traffic flow rate in the cell.
- k: vehicle density in the cell.

– v: vehicle speed in the cell.

Traffic flow rate can be calculated from the vehicle speed and vehicle density in the cell. There are many equations that can calculate the vehicle speed from the density. In this research, we use the formula of Green Shields[8] to calculate the traffic flow rate.

The traffic flow simulation implemented in this research uses a data set of network and departure time, departure point, destination point, and whether the vehicle is a probe vehicle. To verify the reproducibility of the traffic flow simulation, we compare ours with the traffic flow simulation based on the cellular automata model[9][10] regarding a coefficient of simple linear regression and root mean square of the travel time of the vehicle. The cellular automata model is a discrete model and is easy to implement. In the experiments, root mean square being close to 0 and a coefficient of simple linear regression being close to 1 represents that the reproducibility of vehicle travel time is high.

Table 2. Comparison of cellular automata model and cell transmission model for reproducibility of travel time

model	root mean square	coefficient of simple linear regression
cell transmission	2.029	0.835
cellular automata	3.502	0.339

Table 2 shows the results of a comparison of the coefficient of simple linear regression and the root mean square regarding the simulation based on the cellular automata model and the cell transmission model. Table 2 shows that the reproducibility of the travel time in the simulation based on the cell transmission model is greater than that of the cellular automata model from the values of both the coefficient of simple linear regression and the root mean square.

Traffic flow simulation that reproduces a distribution of link travel time tending to have two peaks is required for information provision and shows the effectiveness of proposed method.

Figure 4 shows that the passage number and travel times of the vehicles on one link in the network when we simulated movement of the vehicles using the Kichijoji and Mitaka data set on traffic flow simulation. As Figure 4 shows, it was confirmed that it is possible to reproduce a distribution of link travel time tending to have two peaks in the traffic flow simulation implemented in this research.

5.3 Experimental result

Difference of the travel time for each distance of vehicles We show the comparison results regarding the travel time of vehicles between using the

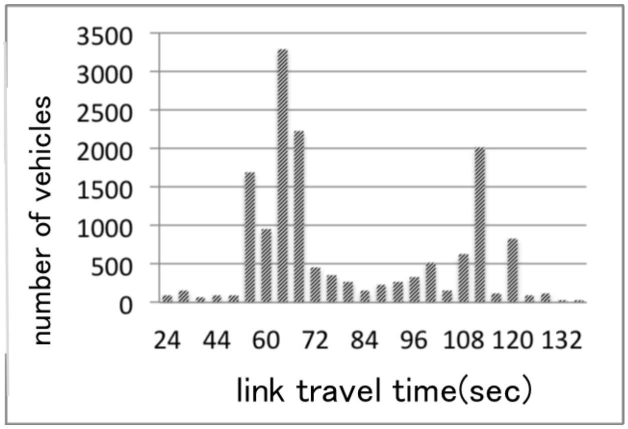


Fig. 4. Traffic volumes and travel time of the vehicles at a certain link in the simulation

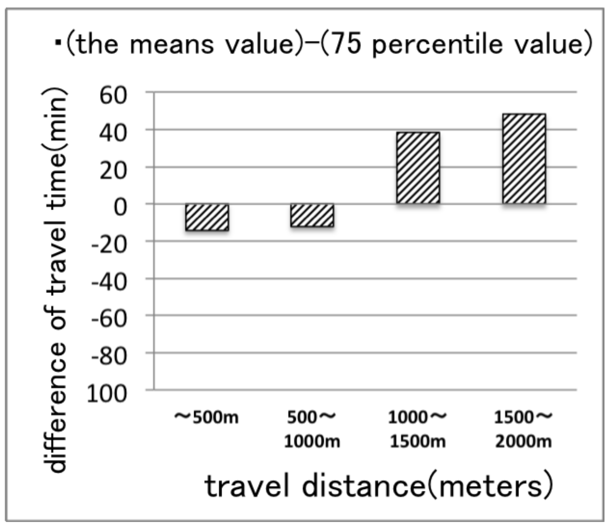


Fig. 5. Difference in travel time of vehicles between using the mean value and 75th percentile value for route search by travel distance of vehicles

mean value, 25th percentile value and 75th percentile value of the historical accumulated data of the link travel time.

Figures 5 and 6 show difference of travel time between using the mean, 25th percentile value, and 75th percentile value for route search by travel distance of vehicle. The value of the graph subtracts the travel time when using 75th percentile and 25th percentile values from the travel time in case of using the mean value. As the value of the graph is large, it represents that the travel time

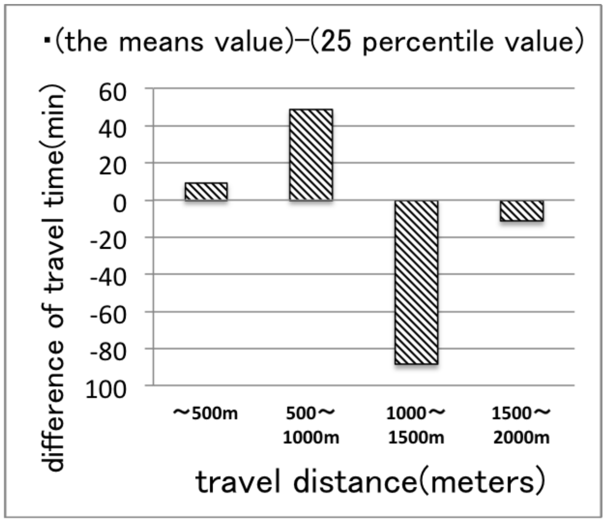


Fig. 6. Difference in travel time of vehicles between using the mean value and 25th percentile value for route search by travel distance of vehicles

of vehicles using the mean value is more than the travel time of vehicles using the 25th percentile value and 75th percentile value. In Figure 5, the travel time of vehicles using the 75th percentile value is less than that using the mean value regarding vehicles that travel distances of 1,000 meters or more. On the other hand, in Figure 6, the travel time of vehicles using the 25th percentile value is less than that using the mean value regarding vehicles that travel distances of 1,000 meters or less.

Proposed method and evaluation In this research, we proposed that vehicles whose travel distance is 1,000 meters or less perform a route search using the 25th percentile value of historical accumulated data, and vehicles whose travel distance is 1,000 meters or more perform a route search using the 75th percentile value of historical accumulated data. The effect analysis is the total travel time of all vehicles in the simulation.

Figure 7 shows the result of the simulation experiment in each case. Values in the graph of Figure 7 show the total travel time of all vehicles in each case. We describe setting of each case. There is no probe vehicle in case 1; that is, vehicles do not change their routes in repetition. The probe vehicles search for the route using mean value in case 2, 25th percentile value in case 3, and 75th percentile value in case 4 as link cost. We use the proposed method in case 5.

As shown in the graph of Figure 7, using both 25th percentile value and 75th percentile value of historical accumulated data reduced the travel time of all vehicles most.

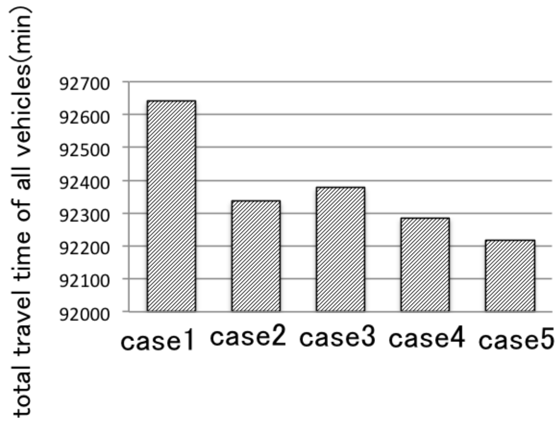


Fig. 7. Total travel time of all vehicles in each information provision

6 Conclusion and future work

In this research, we presented background information about the problems caused by the increasing number of vehicles on the road, such as economic losses and environmental degradation. Also, the number of probe vehicles has increased in recent years, and the distribution of link travel time tends to have two peaks. Next, we proposed information provision based on a distribution of link travel time tending to have two peaks. In the experimental simulation, as the information provision to the probe vehicle, we proposed using both the 25th percentile and 75th percentile values as a function of travel distance of a vehicle. We demonstrated that the proposed method reduced the travel time of all vehicles compared with the conventional method.

In future work, we will simulate a large network. In this experiment, since we used a small network data set, it is necessary to test a larger network to confirm that the proposed method is effective.

The information method proposed in this research used travel distance of the vehicles; it is also necessary to use such factors as the departure time of the vehicles in future research.

References

1. Carlos F. Daganzo: The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, Transportation Research B, 28(4):269-287(1994)
2. Carlos F. Daganzo: The cell transmission model, part II: Network traffic, Transportation Research Part B: Methodological, Volume 29, Issue 2, Pages 79-93, (1995)

3. Ministry of Land, Infrastructure, Transport and Tourism: amount of carbon dioxide emissions: http://www.mlit.go.jp/sogoseisaku/environment/sosei_environment_tk_000007.html, Ministry of Land, Infrastructure, Transport and Tourism
4. Ryo Kanamori: Jun Takahashi, Takayuki Ito: A proposal of Route Information Provision with Anticipatory Stigmergy for Traffic Management, IPSJ SIG Technical Report. ICS, 2012-ICS-169(1), 1-6, (2012)
5. Takayuki Morikawa: Toshiyuki Yamamoto, Tomio Miwa, Lixiao Wang: Development and Performance Evaluation of Dynamic Route Guidance System "PRONAVI" , Traffic Engineering, Vol.42, No.3, pp.65-75(2007)
6. Mohsen Ramezani, Nikolas Geroliminis: On the estimation of arterial route travel time distribution with Markov chains, Elsevier, vol. 46, pages 1576-1590(2012)
7. Traffic Simulation Clearing House: http://www.jste.or.jp/sim/index.html
8. B.D.Greenshields: A Study in Highway Capacity, H. R. B., Proceedings, Vol.14, p.468(1935)
9. A. Schadschneider, M. Schreckenberg: Cellular automaton models and traffic flow, J. Phys. A26, L679 (1993)
10. Shinichi Tadaki, Makoto Kikuchi: Stability of congestion phase in a two-dimensional cellular automata traffic model, Bussei Kenkyu 61(5), 453-457(1994)
11. Hirokazu Akahane, Takashi Ooguchi, Hiroyuki Oneyama: Genealogy of the development of traffic simulation model, Traffic Engineering 37(5), 47-55(2002)