Decisions at Hand: A Decision Support System on Handhelds

Blaz Zupan[a,b,c], Ales Porenta[a], Gaj Vidmar[d], Noriaki Aoki[c], Ivan Bratko[a,b], J. Robert Beck[c]

[a] Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[b] Jozef Stefan Institute, Ljubljana, Slovenia
[c] Baylor College of Medicine, Houston, TX, USA
[d] Faculty of Arts, University of Ljubljana, Slovenia

Abstract

One of the applications of clinical information systems is decision support. Although the advantages of utilizing such aids have never been theoretically disputed, they have been rarely used in practice. The factor that probably often limits the utility of clinical decision support systems is the need for computing power at the very site of decision making - at the place where the patient is interviewed, in discussion rooms, etc. The paper reports on a possible solution to this problem. A decision-support shell LogReg is presented, which runs on a handheld computer. A general schema for handheld-based decision support is also proposed, where decision models are developed on personal computers/workstations, encoded in XML and then transferred to handhelds, where the models are used within a decision support shell. A use case where LogReg has been applied to clinical outcome prediction in crush injury is presented.

Keywords:

Decision support; Handhelds; Logistic regression

Introduction

In the past clinical information systems have mainly focused on the management of clinical data. Besides the data management, modern clinical information systems are more ambitious and many include tools for decision support. Providing physicians with such tools can increase the quality of clinical decisions and decrease costs [1]. Clinical decision support systems can help in prognosis, diagnosis, treatment planning and other frequent and non-trivial tasks physicians have to face in daily practice.

Decision support systems in medicine can be viewed as intelligent advisors, or sources of second opinion. Their typical life cycle often consists of defining a problem on which to focus, gathering the corresponding retrospective data, and constructing the predictive model. The predictive models form a core of decision support systems: a physician would enter the data on the patient under consideration, and then invoke the decision support model to compute the corresponding probabilities of, say, diagnosis or prognosis, success of treatment plans.

For medical decision support systems to be accepted and used by physicians, they have to be available wherever decisions take place: in physicians' offices, during visits at patient's beds, in meeting rooms when discussing their use with colleague physicians, … In other words, the decision support system has to be immediately accessible – best, the technology has to reside and be readily available in the physician's pocket.

Recent technological advances can help us to solve the abovementioned problem. Namely, handheld computers are more and more becoming a standard physician's accessory to track dates, store contact information, or even manage other data on patients. They are computationally sufficient to carry decision support tasks of reasonable complexity. While there are an increasing number of medical applications available for handhelds, their use in clinical decision support is at best rare.

Of the few systems that exist and are in routine use, we here mention the Prostogram by M. Kattan and P. Fearn [4,5,6] from Memorial Sloan Kettering Hospital in New York, NY. Using Prostogram, expert urologist can enter the preoperative or postoperative data of a patient with prostate cancer, and compute the probability of cancer recurrence after radical prostatectomy. The program is easy to use and is in routine use among physicians in the Department of Urology at the aforementioned institution. Its major weakness, however, is that the prognostic model in the Prostogram is hardcoded in the program. Any alternation of the prognostic model, like changing the coefficients used to compute probabilities, or adding other predictive variables, requires changes of the underlying program, its recompilation and reinstallation on the handheld – not a trivial task that requires involvement of programmer with special skills.

The work described in this paper was inspired by the approach of Kattan and Fearn, by extending their idea to the point where a physician that has a decision support model of a particular type available can use it on handhelds without the need to devise a special program. For this purpose, we have developed a decision support shell called LogReg that runs on Palm™ handhelds, an often used handheld platform. LogReg takes a decision support model encoded in XML and provides for data entry and model-based reasoning on a handheld. The current version of LogReg supports models obtained by logistic regression, a modern statistical technique that is often used for medical data analysis and modeling.

In the paper, we first introduce basic concepts of logistic regression and present some of its aspects important for the proposed decision support shell. The discussion of LogReg's architecture is given next. We use the model that predicts the probability of crush syndrome, and present a series of LogReg's snapshots when using this model. Several ongoing developments that are using the proposed framework are discussed in the Conclusion.

Logistic Regression

The Logistic Model

Logistic regression is one of the most frequently used multivariate statistical methods in the biomedical sciences and it is extensively treated in various textbooks [7,3]. It is use for predicting a binary dependent variable, i.e., an outcome coded as either 0 (e.g., NO or disease absence) or 1 (e.g., YES or disease presence). It is based on the logistic function, which is of sigmoid shape and asymptotically approaches 1 or -1 when z approaches positive or negative infinity, respectively:

$$f(z) = \frac{1}{1+e^{-z}} \qquad (1)$$

The linear combination of the independent variables (the components of the predictors' vector **x**)

$$z = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k \qquad (2)$$

is included in the logistic model

$$P(Y=1 \mid X_1,\ldots,X_k) = P(\mathbf{x}) = f(z) \qquad (3)$$

An increase in the sum of $\beta_i X_i$ thus implies an increase in the conditional probability of the outcome 1. The model is basically a linear one, as the relationship between the predictors and the predicted value, subject to the logit transformation (i.e., taking the natural logarithm of the odds), is linear. The regression equation is

$$z = \operatorname{logit}(P) = \ln\frac{P}{1-P} \qquad (4)$$

As always within the general linear model, one can include binary, categorical and/or quantitative predictors in the logistic regression model, as well as higher-order terms and interactions, whereby each nominal (categorical) variable with n categories is encoded with n-1 binary ('dummy') variables. Due to its prototype nature and in order to assure simplicity of display and interpretation, LogReg is presently limited to models without higher-order terms and interactions.

The usual linear regression is less suitable for predicting a binary outcome (in which case the second 'classical' alternative – linear discriminant analysis – is equivalent to linear regression) than the logistic model, even though for P between 0.2 and 0.8 the two models do not differ much. One reason is face validity, since linear regression can predict dependent-variable values above 1 or below 0, while the major advantage of the logistic model lies in its robustness (its assumptions are weaker, the linear model's assumption of normally distributed residuals can not be met).

Logistic regression procedures are built into numerous modern statistical software packages. The maximum likelihood method is almost exclusively used for estimating model parameters. It requires somewhat larger samples, but with a sufficiently high ratio of the number of cases compared vs. the number of variables, it produces much more reliable results than the least-squares method.

A problem, which is particularly relevant for our application, is estimation of the baseline odds, i.e., the parameter α which equals logit P (0,…,0). In order to be able to estimate it, one must collect the data with a follow-up study, or carry out a case-control study while knowing the sampling fractions of the population within cases and controls [3]. Relative influence (importance) of individual predictors is usually expressed in terms of odds ratio (*OR*), with which there is no estimation problem with case-control and cross-sectional data. The general formula

$$OR = e^{-\beta_i (X_{i1} - X_{i2})} \qquad (5)$$

reduces to e^{β} in case of binary predictors. For numeric predictors, the formula represents the change in the odds if X_i increases by one unit holding all other predictor values constant.

Estimating Confidence Interval for Predicted Value

The predicted value P for a case that was not present in the training set is obtained by plugging its X values into the logistic formula with the estimated regression coefficients. Henceforth we will assume that P>0.5 always implies classifying the case into the class 1 and P<0.5 always implies class 0, because determination of optimal cut-off point for classification far exceeds the scope of our application, so as does the whole area of measures of predictive accuracy of the logistic regression model.

The systems for supporting medical decision-making should provide the user at least with some idea of the reliability (certainty) of the decision. Therefore we implemented calculation of confidence interval (*CI*) for predicted value in LogReg. Approximating the binomial distribution of proportion p with the normal distribution, i.e., obtaining *CI* by calculating the standard error as

$$SE(p) = \sqrt{\frac{p(1-p)}{N}} \qquad (6)$$

would be a feasible procedure for sample size over 100 (which is a rule when developing logistic regression models) for P between 0.2 and 0.8, but such a solution does not work for extreme proportions (where the upper bound of the 95%-*CI* can exceed 1, and the lower bound can fall below 0) and it also completely ignores the information provided by the values of independent variables for the classified case. Hence, a modern procedure [9] was implemented, which is based on the linear nature of the logistic model: the standard inferential apparatus of linear regression is used to assess $g(\mathbf{x}) = \text{logit}(P)$, then the obtained CI bounds for $g(\mathbf{x})$ are transformed into probabilities with the inverse transformation

$$P = \frac{1}{1+e^{-g(\mathbf{x})}} \qquad (7)$$

To implement this in LogReg, two problems had to be solved, arising from the standard expression for 95% (or – with proper adjustment of the multiplicative factor applied to the standard error of the estimate – any other narrower or wider) *CI* for predicted valued in linear regression [8]:

$$CI = E(Y) \pm \sqrt{MSE(1+h_{ii})} \qquad (8)$$

The first problem is of statistical nature and regards the mean squared error (*MSE*): there is no real equivalent of this quantity in logistic regression, so in the definition

$$MSE = \frac{\sum (Y - E(Y))}{1 - \text{number of predictors} - 1} \qquad (9)$$

instead of using the difference between actual and predicted value, the difference between actual class (0 or 1) and predicted probability P can be used, regardless of the fact that the calculation of standard error of estimate is carried out 'within the world of' logit values.

The leverages (h_{ii}), i.e., the diagonal values of the matrix $\mathbf{H}=\mathbf{X}\ (\mathbf{X}^T\ \mathbf{X})^{-1}\ \mathbf{X}^T$, represent a technical problem. Transferring the data matrix $\mathbf{X}$ (which has values 1 in the first column, corresponding to the constant term in the regression equation) which has been used to develop the logistic regression model into the handheld, then transposing, inverting and multiplying matrices in the handheld is out of consideration because of the handheld's memory and speed limitations. An efficient solution is to transfer previously calculated matrix $(\mathbf{X}^T\ \mathbf{X})^{-1}$ into the handheld and calculate leverage of a new case as

$$h_{ii} = \mathbf{x}_i^{\ T} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{x}_i \qquad (10)$$

Explaining Decisions with Risk Odds Ratio

In addition to the presentation of uncertainty of prediction, explaining the system's decision increases the quality of support for the user's decision. The issue of inter-relations and visualization of different parameters and characteristics of logistic regression analysis demands extensive and in-depth coverage, such as provided by Harrell [2], while here we only mention the display of risk odds ratio (*ROR*), provided by LogReg for each of the predictors.

Generally speaking, *ROR* provides a comparison between any two predictor configurations regarding the predicted value P, but it is usually calculated for a pair of configurations (cases) differing only in one predictor value. If the chosen predictor is binary, the expression is either computationally equivalent to the odds ratio (in the case of LogReg this applies to the cases with the value 1 of the predictor of interest), or it equals 1 (if the predictor value is 0). With numeric predictors, an individual case is compared with a value representative for the population from which the analyzed sample has been taken, which is usually the sample arithmetic mean of predictor (this is also the implementation in LogReg). If the predictor is a categorical one, the calculation (which, as already stressed, only applies to models without interaction terms) can best be illustrated by the following example. Four categories are coded as a group of three binary variables, whereby the baseline (first) category has been coded as X_{1a}=1, X_{2a}=0 and X_{3a}=0, and the individual for whom we have predicted P by means of logistic regression belongs to the category coded as third (X_{1b}=X_{2b}=0, X_{3b}=1):

$$ROR_{3:1} = e^{(X_{1a}-X_{1b})\beta_1+(X_{2a}-X_{2b})\beta_2+(X_{3a}-X_{3b})\beta_3} = e^{-\beta_1+\beta_3} \qquad (11)$$

Architecture of Handheld-Based Decision Support System

To use LogReg on a Palm™ handheld, one has first to develop a decision support model on a personal computer or a workstation (Figure 1). The model has than to be encoded in XML and (with a help of a special program) the user has to mark the corresponding XML file to be transferred to the Palm™ handheld. The model is then loaded to Palm™ handheld with its first synchronization with a PC. Multiple models can be prepared and selected to be used on Palm™ handheld in this way. One can prepare and choose any number of such models. Once models are transferred to the Palm™ handheld, they are ready to be used.

When implementing LogReg, we needed to solve three problems: (1) the transfer of XML-encoded models to the Palm™ handheld, (2) the data entry on a Palm™ handheld, and (3) the utility of the decision models. For synchronization, a special program is automatically run on a PC that opens the XML files that have been marked by the user, parses the XML files, and sends the data on the models to handheld's data base. When invoked, LogReg consults the database to see which models reside on the Palm™ handheld. The most recently used model is selected, and an entry form with predictor variables is dynamically generated. After the user specifies the data, LogReg consults the database again to retrieve the logistic model's coefficients that are used for computation of the probabilities.

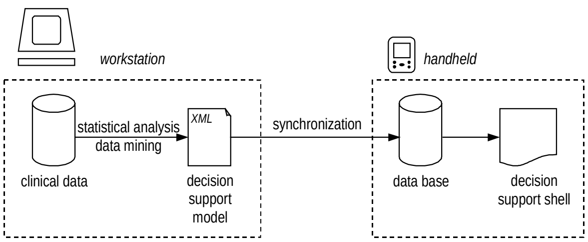

Figure 1 Handheld-based decision support system

Notice that the content of the LogReg's screen depends on the information encoded in XML files. The construction of a decision support system for some specific problem thus consists of defining the corresponding XML file that encodes the decision support model, and does not involve any programming.

Use Case: Clinical Outcome Prediction in Crush Injury

Figure 2 offers a series of snapshots that show how LogReg can be used in practice. For demonstration purposes, we here use the model developed by co-author Aoki and his colleagues. The model is used for clinical outcome prediction in crush injury and consists of three prognostic variables (delay of response by emergency medical service personnel, patient's urine color and pulse rate), and based on their values the model computes the probability the patient has severe crush syndrome.

To use LogReg, the user has to press on its icon (Figure 2.A). The program displays the entry form of the model that was most recently used (B). The user can also see which other models were loaded into Palm™ handheld (D) and display additional information about each model (E). LogReg supports different modes for data entry (B, C), including pull-down menus and check boxes for nominal and entry fields for continuous variables. To engage the decision support model, the user has to press the "Compute" button (F): LogReg displays the probability of outcome and its confidence intervals (G). The influence of each of the prognostic variables on the outcome's probability can be observed through risk odds ratios (H). The XML file that encodes the crush injury syndrome model and that was used in our example can be obtained from the web site at http://magix.fri.uni-lj.si/palm.

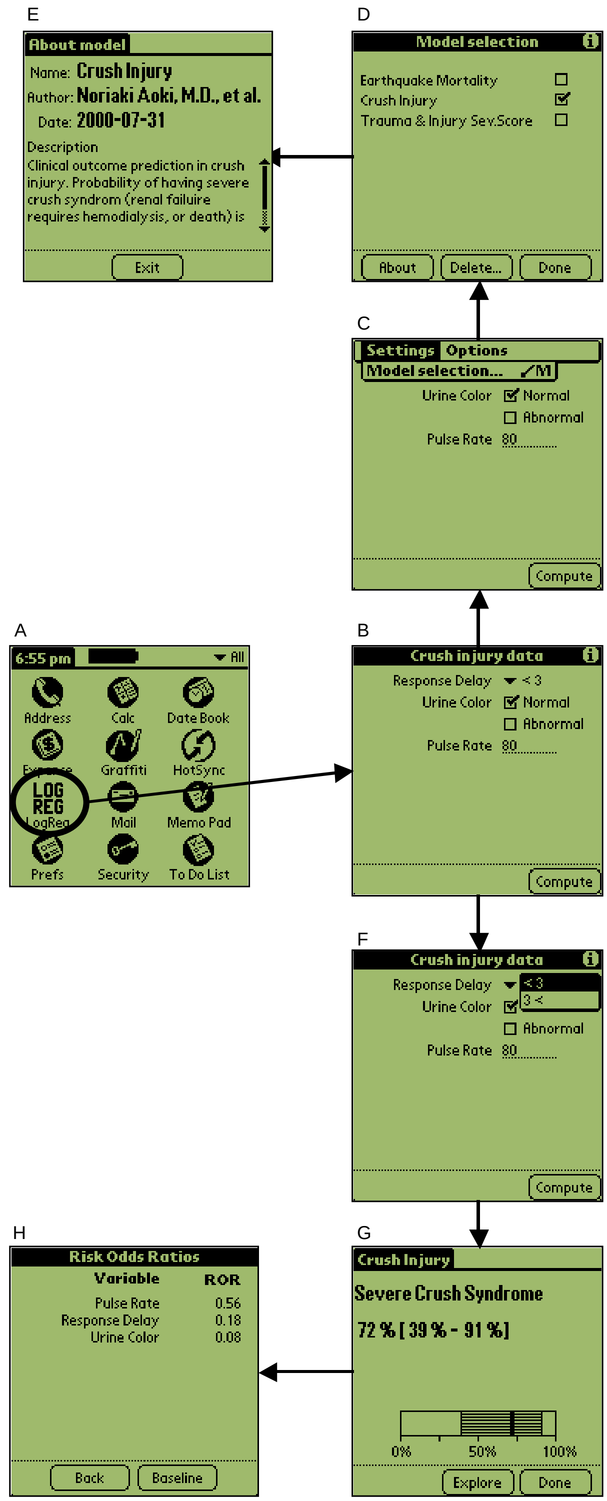

Figure 2 Snapshots from LogReg

Conclusion

We have described the framework and the decision support shell LogReg that enable the use of logistic regression models on handheld computers. Our main assumption that such systems will be useful in clinical practice is the observation that handhelds, and in particular Palm™ handhelds, are increasingly used by physicians, who are becoming familiar with such devices and their easy-to-use operating systems. While LogReg at present only supports the reasoning with logistic regression models, the concept that we have described in this paper is general and can be extended to other modeling techniques.

At the time of writing this paper, several applications of LogReg are being developed. With Baylor College of Medicine in Houston, TX, we are developing decision support systems for management of trauma patients. A treatment planning decision support system for patients with hip or wrist injury is being developed in collaboration with the Clinical Center in Ljubljana, Slovenia. With Memorial Sloan Kettering Hospital in New York we are constructing LogReg-compatible predictive models for cancer recurrence prognosis and treatment selection.

Acknowledgments

We would like to acknowledge the financial support of Ministry of Science and Technology of Republic of Slovenia and of American Cancer Society (grant no. 9610000501). CHS, a company from Ljubljana, Slovenia, sponsored some of the Palm™ devices that were used in the project described. We are thankful to Michael W. Kattan, Begoña Campos Bonilla, Clint Cummins and Janez Stare for valuable suggestions and advice.

References

[1] Anderson JD: Increasing the acceptance of clinical information systems. *MD Computing* 1999; 16(1):62-5.

[2] Harrell FE: *Predicting outcomes: applied survival analysis and logistic regression*. Charlotesville 1999: School of Medicine, University of Virginia.

[3] Hosmer DW, Lemeshow S: *Applied logistic regression*. New York 1989: Wiley.

[4] Kattan MW, Zelefsky MJ, Kupelian PA, Scardino PT, Fuks Z, Leibel SA. Pretreatment nomogram for predicting the outcome of three-dimensional conformal radiotherapy in prostate cancer. *J Clin Oncol* 2000; 18(19):3252-3259.

[5] Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol* 1999; 17(5):1499-1507.

[6] Kattan MW, Fearn P: Nomograms Web Page. http://www.nomograms.org, 2000.

[7] Kleinbaum DG: *Logistic regression: a self-learning text*. New York 1994: Springer.

[8] Myers JL, Well AD: *Research design and statistical analysis*. New York 1991: Harper Collins.

[9] Piergorsch WW, Casella G: Confidence bands for logistic regression with restricted predictor variables. *Biometrics* 1988; 44(3):739-750.

Address for correspondence

Blaz Zupan, Faculty of Computer and Information Science, University of Ljubljana, Trzaska 25, SI-1000 Ljubljana, Slovenia

e-mail: blaz.zupan@fri.uni-lj.si
url: http://magix.fri.uni-lj.si/blaz