G.F. Cooper

Center for Biomedical Informatics
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Synopsis

Knowledge Processing and Decision Support Systems

The section on knowledge processing and decision support systems contains four papers. One paper highlights methodological issues in developing and evaluating predictive clinical models. Although the emphasis is on artificial neural networks, most of the points apply to all types of predictive models. The remaining three papers in the section are about guidelines. Two evaluate guidelines that are integrated into the physicians' workflow; they show increased physician compliance with the use of computer-based guidelines. The other describes a novel method for converting decision models into guidelines; this paper describes a pilot study that shows relatively high user subjective satisfaction with the generated guidelines. The remainder of this synopsis discusses each paper in turn.

A clinical decision support system for prevention of venous thromboembolism

The first paper, which is by Durieux et al. [1], describes the evaluation of physician compliance with a guideline-based clinical decision-support system. The system recommends anticoagulant therapy for surgical patients as a prophylactic measure to prevent venous thromboembolism.

The clinical guideline for thromboembolism prophylaxis was developed locally. Clinicians complied with the guideline about 95% of the time in the intervention period, during which the decision-support system presented the guideline. When the system was removed, clinician practice reverted back to near a pre-intervention level of about 84%. On the whole, this pattern is consistent with the results of previously published studies of reminder systems. It further shows that even *locally* developed guidelines are not heeded automatically, but rather, require continual reminding to be highly effective.

The decision-support system improved the clinical practice of all clinicians except one, whose guideline compliance was close to 100% during the non-intervention (control) period. The greatest improvement in guideline compliance was observed among the five surgeons who operated on fewer than 40 patients.

The greatest effect of the decision-support system was for patients at moderate risk for venous thromboembolism. It is understandable that a decision-support system might have the most influence on decision making for those patients in the "gray area" (e.g., moderate risk patients). For those patients at low or at high risk, the best decision may be more obvious.

On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology

The second paper, which is by Schwarzer et al.[2], discusses problems with how artificial neural networks (ANN) are derived and evaluated. Although the paper focuses on ANN models, the caveats are applicable for most predictive modeling techniques. Indeed, it is not clear that more traditional statistical models (e.g., logistic regression models) are being derived and evaluated in a manner that is any better than forANNs.

The paper describes how feed-forward neural network models are a generalization of logistic regression models. As pointed out, a more expressive modeling methodology may in theory be able to generate better (i.e., more predictive) models, but any given application of the methodology may not actually result in a better model. Models that are highly expressive (and therefore highly "tunable") are particularly subject to overfitting, which can inhibit predictive performance. Overfitting occurs when the model is tuned too closely to the training (derivation) data in ways that do not correspond well to the process that is generating the data. The paper illustrates the problem and describes how penalty terms can attenuate it. Although not mentioned, Bayesian approaches to learning neural networks intrinsically include such penalty terms [3]. These and other advanced approaches for deriving ANNs are described in papers at [4]. On the whole, current ANN researchers are aware of the problems that were reviewed by Schwarzer et al. and are avoiding those problems. As this paper indicates, however, recent application of those methods by clinical researchers is lagging behind.

Schwarzer et al. describe problems in evaluating ANNs. It is worth noting that the evaluation of most predictive models is subject to these same problems. For example, validating a model only on the dataset that was used to derive the model is usually a bad idea, regardless of the type of model being validated. The paper mentions the use of cross-validation methods as a way of efficiently using the data, while avoiding a biased evaluation. The computer-intensive nature of these methods is rarely a problem nowadays, and these methods are being used routinely in machine-learning research.

Beyond ANNs, other statistical and machine-learning methodologies continue to be explored. For example, in recent years, increasing attention has been focused on Support Vector Machines (SVM) and related methodologies [5, 6]. Hopefully the current paper by Schwarzer et al. [2] will lessen the chance that a similar paper will need to be written in ten years about the misuses of SVMs for prognostic and diagnostic classification.

Assessment of decision support for blood test ordering in primary care

The paper by van Wijk et al. [7] studies physician compliance with a computer-based guideline for blood-test ordering, which was integrated into the physicians' workflow. The physician control group received an electronically displayed list of possible blood tests, which were organized to highlight 15 tests that cover most of the clinical situations in primary care.

Seventy four percent of the eligible practices participated in the study. Among the physicians receiving the guideline, 71% of their orders were generated by using the decision-support system.

Overall, there was 20% more reduction in test ordering for those physicians who used the guideline versus those in the control group. Interestingly, the reduction was due primarily to a decrease in the ordering of some specific tests. For example, there was a marked reduction in the number of tests for creatinine, but little reduction in the number of tests for sodium. Like the Durieux et al. [1] study discussed above, this investigation by van Wijk et al. indicates that the impact of a guideline on clinician behavior can be selective in interesting and at times non-obvious ways.

Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models

The paper by Sanders et al. [8] introduces a method for converting a decision model (e.g., a decision tree) into a clinical guideline that is made available on the web. By modifying the parameters in the decision model, the guideline can be tailored to a clinical site or even a particular patient. Doing so may make decision support more accessible, relevant, and helpful.

The paper describes a pilot study that involved participants from the authors' department and program who used the decision-model-generated guideline versus a control guideline from the literature. The domain area was the treatment of non-small cell lung cancer.

Results show that the decision-model-generated guideline was rated relatively highly by the participants. As the paper points out, however, the participants were not blinded to the source of the guideline they used, and thus the possibility of a rating bias is a concern. Therefore, the results are quite preliminary. Nonetheless, the approach seems sound and promising.

Discussion

The four papers in this section provide solid contributions in the area of knowledge processing and decision support systems. Several trends are suggested, which this author believes will continue for many years to come.

Traditionally, a clinical guideline or model has been developed from knowledge and data at one or a few study sites and then applied (perhaps with ad hoc modification) at many local clinical sites. Increasingly, these guidelines and models are likely to be automatically or semi-automatically tailored based on local knowledge and data. Taken to its logical conclusion, this trend leads to patient-specific computer-based guidelines and models.

A closely related issue is the use of experimental and observational data in the development of guidelines and models. Experimental studies, such as randomized controlled trials, often provide the most trustworthy methods we have for establishing causal relationships from clinical data. Such studies, while potentially highly informative, may not always be safe, ethical, logistically feasible, or financially worthwhile. Observational data are passively observed. Such data are more readily available than experimental data and, indeed, most clinical databases are observational. As clinical data become more and more routinely recorded electronically, the opportunities for using such local observational data increase. Interestingly, recent analyses of the literature indicate that experimental and observational studies often yield similar predictions about the direction and magnitude of causal relationships in clinical medicine [9, 10]. These results suggest that both experimental and observational data have powerful roles to play in deriving clinical knowledge of cause and effect. Increasingly, it is likely that the data from (typically) *non-local* experimental studies will be tailored using data from *local* observational databases in order to develop clinical guidelines and models that are applied locally. Such approaches will leverage the power of both experimental and observational data [11].

Finding a bridge between the development of sophisticated diagnostic, predictive, and therapeutic models and the presentation of such models to clinical users is likely to remain an important issue. Although at the present time the targeted users are usually clinicians, increasingly there are likely to be other users (particularly patients) of the underlying computer-based clinical models. Each user will be able to provide their own input into the models and receive tailored output. As one example, direct computer-based assessment of patient symptoms, preferences and subjective outcomes is likely to grow significantly in the years ahead [12].

The evaluation of computer-based clinical guidelines and models will continue to be critically important. Such studies are usually expensive in terms of time and money. As our understanding of such evaluations develops, it seems likely we will see sophisticated computer-based tools that assist in designing, performing, and analyzing such experiments in a manner that is both efficient and sound.

References

1. Durieux P, Nizard R, Ravaud P, Mounier N, Lepage E. A clinical decision support system for prevention of venous thromboembolism: Effect on physician behavior. JAMA 2000;283:2816-21.
2. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. Stat Med 2000;19:541-61.
3. MacKay DJC. Bayesian methods for supervised neural networks. In: Arbib MA, editor. The Handbook of Brain Theory and Neural Networks. Cambridge, MA: M.I.T. Press; 1995. p. 144-9.
4. NIPS Online website at http://nips.djvuzone.org/
5. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines. Cambridge, U.K.: Cambridge University Press; 2000.
6. Herbrich R, Graepel T, Campbell C. Bayes point machines. Journal of Machine Learning Research 2001;1:245-279 (Available at http://www.ai.mit.edu/projects/jmlr/papers/volume1/herbrich01a/herbrich01a.pdf).
7. vanWijk MAM, van der Lei J, Mosseveld M, Bohnen AM, Van Bemmel JH. Assessment of decision support for blood test ordering in primary care: A randomized trial. Ann Intern Med 2001;134:274-81.
8. Sanders GD, Nease RF, Owens DK. Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models. Med Decis Making 2000;20:145-59.
9. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000;342: 1878-86.
10. Ioannidis JPA, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821-30.
11. Cooper GF, Yoo C. Causal discovery from a mixture of experimental and observational data. In: Proceedings of the Conference on Uncertainty in Artificial Intelligence; Morgan Kaufmann; 1999. p. 116-25.
12. Treadwell JR, Soetikno RM, Lenert LA. Feasibility of quality-of-life research on the Internet: A follow-up study. Qual Life Res 1999;8:743-7.

Address of the author
Gregory F. Cooper, M.D., Ph.D.
Center for Biomedical Informatics
University of Pittsburgh
Suite 8084 Forbes Tower
200 Lothrop Street
Pittsburgh, PA 15213-2582
U.S.A.
E-mail: gfc@cbmi.upmc.edu