FIONN:
A Framework for Developing CBR Systems

Dónal Doyle, John Loughrey, Conor Nugent,
Lorcan Coyle, Pádraig Cunningham
Trinity College Dublin
Donal.Doyle@cs.tcd.ie

1. Introduction

Case-Based Reasoning (CBR) is a very popular methodology for developing knowledge-based systems [1]. Yet there are few toolkits available for building CBR systems. In this paper we present a framework called Fionn that is specifically designed for the development of CBR systems. Since Fionn was designed specifically for CBR it provides good support for some of the unique characteristics of CBR systems. We will concentrate on three of these in this paper:

1. Feature Selection
2. Learning Similarity Measures
3. Case-Based Explanation

Fionn is implemented in a XML framework for CBR called CBML that has been described elsewhere [11]. Fionn also has a comprehensive evaluation framework that will be described.

1.1 Evaluation Framework

One of the main benefits Fionn offers is the ability to perform Feature Selection and to learn Similarity Measures. These are essentially parameter setting tasks that are achieved by a search through the parameter space. The best parameters are selected by cross-validation on the data that is available. This comes down to a task of comparing different classifiers (i.e. different parameter sets). Recent research in Machine Learning shows that this is anything but straightforward [2, 4, 5, 6, 7]. Problems of overfitting and accuracy estimation with skewed class distributions need to be considered, for instance. To address these issues Fionn supports a variety of validation schemes (hold-out validation, cross-validation, leave-one-out cross-validation) and a selection of error functions (e.g. 0/1 loss, harmonic mean of positive and negative errors).

1.2 Feature Selection

Feature selection in Fionn supports the search for an optimal feature subset using a variety of Wrapper-like [3] algorithms (see Figure 1). The framework is both independent of the type of classifier and of the evaluation technique used.

The algorithms currently implemented in Fionn are;

- Sequential Forward Floating Search (SFFS)
- Forward Selection (FS)
- Backward Elimination (BE)
- Genetic Search
- Wrapper[2] [13]

The FS and BE searches can all be modified so that the hop size can be increased using the Compound Operators algorithm [3]. This is necessary when using datasets where the number of features is large. A Node Caching option is also implemented in the Genetic Algorithm as some nodes can often be revisited. Caching in such situations means that these nodes do not need to be re-evaluated.

The feature selection framework integrates with the evaluation framework so that a variety of validation and error (loss) measures can be used. Reunanen recently showed that intensive search in feature subset selection can lead to overfitting [2]. Fionn has been designed so that the results obtained from the feature selection module can easily be evaluated to investigate whether or not overfitting has occurred.

Figure 1. The output from the Fionn Feature Selection process (using Wrapper[2]).

1.3 Similarity Measures

Fionn supports a number of specifications that describe the components of a CBR system. One of these specifications deals with the representation of a similarity measure. This representation contains two dimensions; a set of weights for each feature in the case, and a representation of the local similarity measure, i.e. the similarity measure between two feature values.

By representing the similarity measure separately from the application code we make it possible to edit the measure without writing a single line of code. This offers considerable advantages in CBR application development, where the domain experts who design similarity measures are unlikely to be coding experts.

We have been developing a multiple user case-based reasoning application, the Personal Travel Assistant [10] within the Fionn framework. Using its similarity measure specifications it is possible to use a separate similarity measure for each user or for individual groups of users. Fionn also allows us to optimize these similarity measures automatically (in a manner similar to that described in [8, 9, 12]). This provides an added level of personalization to the system.

1.4   Case-Based Explanation

A further advantage of CBR is the potential to use retrieved cases to support explanation. This idea is supported in Fionn in a framework we call explanation-guided retrieval. This is motivated by the fact that when two cases are equally similar to the target case one may be more compelling in explaining a prediction than another. Explanation-guided retrieval attempts to assign a higher level of utility to cases that lie between the query case and the decision surface than to cases that don't. We have done some work in the area of predicting blood alcohol levels relative to drink driving limits [14]. In this domain an important feature in terms of explanation is the units of alcohol consumed, e.g. if trying to predict that a query case with units consumed = 5 is over the drink-driving limit, a more convincing retrieval is a case of 4 units that is over the limit than a case with 6 units consumed that is over the limit.

2.   Conclusion

In this paper we introduced some of the features of the Fionn framework for developing CBR systems. We have outlined the evaluation framework that underpins the feature selection and similarity learning process and we have presented a brief description of the Explanation-Guided Retrieval mechanism in Fionn.

References

1.   Aamodt, A., Plaza, E., Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches .*AICom - Artificial Intelligence Communications*, IOS Press, Vol. 7: 1, pp. 39-59. 1994.
2.   Reunanen, J., Overfitting in Making Comparisons Between Variable Selection Methods. *Journal of Machine Learning Research* 3 (2003) 1371-1382.
3.   R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 1997.
4.   Provost, F.; Fawcett, T.; and Kohavi, R. The case against accuracy estimation for comparing classifiers. *In Proc. of the Intl. Conf. on Machine Learning*, 445-453, 1998.

5. Bouckaert, Remco R*., Choosing between two learning algorithms based on calibrated tests, *In Proc. of the Intl. Conf. on Machine Learning*, 51-59, 2003

6. Bi, Jinbo, Bennett and Kristin, P., Regression Error Characteristic Curves, In *Proc. of the Intl. Conf. on Machine Learning,*, 43-51, 2003

7. Provost, F. and Fawcett, T, Robust Classification Analysis for Imprecise Environments, *Machine Learning*, vol. 42, no. 3, pp. 203--231, Mar. 2001.

8. L. Karl Branting, Learning Feature Weights from Customer Return-Set Selections, *The Journal of Knowledge and Information Systems (KAIS)* to appear, 2003.

9. Coyle, L., Cunningham, P., (2003) Exploiting Re-Ranking Information in a Case-Based Personal Travel Assistant, *The ICCBR'03 Workshop on Mixed-Initiative Case-Based Reasoning* (http://www.aic.nrl.navy.mil/~aha/iccbr03-micbrw/), Trondheim, Norway, June 2003.

10. Coyle, Cunningham & Hayes (2002) A Case-Based Personal Travel Assistant for Elaborating User Requirements and Assessing Offers. *Proceedings of the 6th European Conference, ECCBR 2002,* Susan Craw, Alun Preece (eds.). LNAI Vol. 2416 pp. 505-518, Springer-Verlag

11. Coyle, L., Cunningham, P. & Hayes, C. Representing Cases for CBR in XML. In Proceedings of 7th UKCBR Workshop, Peterhouse, Cambridge, UK, 2002.

12. Stahl, A. & Gabel, T., Using Evolution Programs to Learn Local Similarity Measures. *5th International Conference on Case-Based Reasoning.* K. D. Ashley & D. G. Bridge (Eds.). LNAI 2689, pp537-551, Springer Verlag, 2003.

13. Dunne, K., Cunningham, P., Azuaje, F., Solutions to Instability Problems with Sequential Wrapper-based Approaches to Feature Selection, *Trinity College Dublin Computer Science Technical Report,* TCD-CS-2002-28, 2002.

14. Cunningham, P., Doyle, D., Loughrey, J., An Evaluation of the Usefulness of Case-Based Explanation, *5th International Conference on Case-Based Reasoning.* K. D. Ashley & D. G. Bridge (Eds.). LNAI 2689, pp122-130, Springer Verlag, 2003.