In the proceedings of the Intn'l conference of the North American Fuzzy Information Processing Society, NAFIPS 1999, New York, pp 228-232

Knowledge Discovery using Cartesian Granule Features with Applications

James G. Shanahan[+], James F. Baldwin[#], Trevor P. Martin[#]

[+] Xerox Research Centre Europe (XRCE)
Grenoble Centre, 6 chemin de Maupertuis
38240 Meylan, FRANCE
Email: Jimi.Shanahan@XRCE.Xerox.com

[#] Advanced Computing Research Centre
University of Bristol, Bristol,
BS8 1TR, ENGLAND

Abstract

Current approaches to knowledge discovery can be differentiated based on the discovered models using the following criteria: effectiveness, understandability (to a user or expert in the domain) and evolvability (ability to adapt over time to a changing environment). Most current approaches satisfy understandability or effectiveness, but not simultaneously, while tending to ignore knowledge evolution. Here we show how knowledge representation based upon Cartesian granule features and a corresponding induction algorithm can effectively address these knowledge discovery criteria (in this paper the discussion is limited to understandability and effectiveness) across a wide variety of problem domains including control, image understanding and medical diagnosis.

1. Introduction

Knowledge discovery can be viewed as the non-trivial general process of discovering valid, useful and ultimately understandable knowledge about an application domain from observation data and background knowledge. It is normally an iterative multi-step process involving steps such as knowledge representation (KR) determination, data preparation, learning, evaluation and interpretation. Knowledge representation and machine learning play key roles in this process. Current approaches to knowledge discovery can be differentiated based on the discovered models using the following criteria: effectiveness (accuracy of model on unseen data), understandability (to user or expert in the domain) and evolvability (ability to adapt over time to a changing environment). Most current approaches satisfy understandability or effectiveness, but not simultaneously, while tending to ignore knowledge evolution. For example most approaches that have focussed on the symbolic representation (transparent) of processes have had only mild success in terms of performance accuracies compared to their mathematical (generally opaque) counterparts. In this paper we present many examples that support this, including a diabetes diagnosis system (see Section 3.1 for full details) where a symbolic learning approach such as ID3 [1] is applied to model this diagnosis process. Similarly a mathematical-derived approach such as neural networks is applied, however the mathematical-based approach outperforms (in terms of accuracy) the symbolic approach even though the symbolic approach outperforms in terms of model transparency and understandability. From a knowledge evolution perspective, only few current approaches address this issue; these include instance-based or case-based learning approaches [2], and radial basis function networks [3]. Other approaches to machine learning generally require retraining in order to adapt to a changing environment, for example neural networks.

Here a new form of knowledge representation based upon Cartesian granule features and corresponding induction algorithms are introduced as means of addressing all three of these knowledge discovery criteria. However in this paper the discussion is limited to understandability and effectiveness. Transparency is addressed through representing concepts in terms of additive Cartesian granule feature rules where each feature focuses on modelling the interactions of its constituent subset of input variables; i.e. a divide and conquer strategy to representation. Cartesian granule features, due to their multi-dimensional nature, increase model accuracy through combating decomposition error, a problem that has plagued flat-feature based approaches such as Bayesian and data browser approaches [4]. Furthermore, the proposed induction algorithm (G_DACG), due to the global genetic search technique used, tends to identify near global optima models (in terms of model accuracy and simplicity) avoiding pitfalls of other induction algorithms such as poor feature selection and feature abstraction. These issues can lead to overly complex models and possibly poor generalisation. For example in the case of decision tree induction (using the ID3 algorithm [1]), concepts are iteratively refined by adding or removing features to/from concept definitions based upon entropy measures. This can result in local optima models due to the greedy nature of the search algorithms.

Cartesian granule features provide a very powerful means of representing real word concepts, which in general seem to be vague or uncertain and in some cases unknown. This power derives from the use of fuzzy sets to partition universes thereby alleviating knowledge stability problems which can arise from approaches such as decision trees [1] which rely on crisp partitions. This stability problem arises from the sharp transition from one concept to another that is inherent in the use of crisp partitions. Furthermore Cartesian granule features facilitate a knowledge discovery process for both classification and prediction problems. Section 2 overviews the knowledge discovery process from a Cartesian granule feature perspective, while Section 3 presents the results when Cartesian granule feature modelling and other approaches are applied to a variety of problems. Finally we finish off with some conclusions in Section 4.

2. Overview of the approach

Cartesian granule features [4, 5] are a new type of multidimensional feature defined over the Cartesian product of words drawn from the linguistic partitions of the constituent feature universes. Cartesian granules (characterised by fuzzy sets) provide an abstraction of the multidimensional universe by carving it into regions that are drawn together as result of indistinguishability, similarity, proximity or functionality. Figure 1 gives an illustrative example of how to extract a Cartesian granule fuzzy set corresponding to car positions in images from example car positions where the top left table corresponds to examples of car positions, corresponding linguistic descriptions and least prejudiced distributions (LPDs or probability distributions). Mass assignment theory provides a formal mapping between linguistic descriptions (fuzzy sets) and probability distributions [6]. The top middle graph corresponds to the initial Cartesian granule frequency distribution, where the granule characterisations (i.e. the fuzzy sets) are also shown. The top right graph depicts the Cartesian granule frequency distribution after updating with the LPD corresponding to the value of *40*. The right middle graph shows the Cartesian granule frequency distribution after updating with the LPD corresponding to the value of *60*. The right bottom graph displays the Cartesian granule frequency distribution after counting all the LPDs corresponding to the example car positions. Finally the left bottom graph depicts the corresponding Cartesian granule fuzzy set for car positions in images i.e. a linguistic summary of car positions in images in terms of the words *left, middle* and *right.*

Systems can be quite naturally described in terms of Cartesian granule features incorporated into additive models i.e. if-then-rules with weighted antecedents. Such systems can be discovered automatically from example data using the G_DACG constructive induction algorithm [7]. The G_DACG algorithm discovers good Cartesian granule features (i.e. the feature subsets and the feature universe abstractions). G_DACG is a population-based search algorithm (based on genetic programming), where each node in the search space is a Cartesian granule feature. G_DACG iteratively hones in on good Cartesian granule features based on the evolutionary operations of crossover, mutation and reproduction. Good Cartesian granule features are subsequently incorporated into rule based models (see for example Figure 3). Inference is carried out using evidential reasoning and semantic unification/match of class fuzzy sets and data fuzzy set [6]. Classification corresponds to taking the class corresponding to the maximum of the inferred results. In the case of prediction problems the predicted value corresponds to a weighted sum of the least prejudiced distributions corresponding to the rule output fuzzy regions [8].

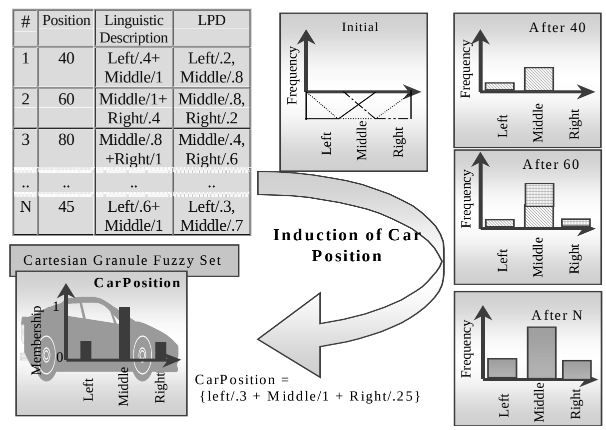

Figure 1: Induction of the Cartesian granule fuzzy set for car positions in images.

3. Results

Modelling with Cartesian granule features is illustrated and compared with other machine learning approaches on a variety of real problem domains.

3.1 Modelling Pima Diabetes Detection Problem

The problem posed here is to predict whether a patient would test positive or negative for diabetes according to the World Health Organisation criteria given eight physiological measurements and medical test results [9]. There are 500 positive diabetes cases and 268 negative cases. The Pima diabetes data set of 768 tuples was split class-wise, approximately as follows: 60% of data allocated to training, 15% to validation and 25% to testing. We applied the G_DACG constructive induction algorithm to the Pima diabetes problem. All eight base features were considered and Cartesian granule features of dimensionality up to five with granularity ranges of [2, 12] were considered (while parsimony was promoted in the form of the fitness function used) thus yielding a multi-million node search space. The G_DACG algorithm iterated for thirty generations. The best discovered model from both a model accuracy and simplicity perspective consists of two four dimensional Cartesian granule features, yielding a model accuracy on test data of 79.7%. A trapezoidal fuzzy set with 70% overlap was determined to be the best granule characterisation.

The Pima diabetes dataset serves as a benchmark problem in the field of machine learning and has been tested on many learning approaches. Table 1 compares some of the results of the more common machine learning techniques with the ACGF modelling approach.

Table 1: Comparison of results for the Pima diabetes detection problem.

Approach.	Accuracy
Additive Cartesian granule feature Model	79.7
Mass Assignment based MATI [10]	79.7
Oblique Decision Trees [11]	78.5
Neural Net (normalised Data)	78
C4.5 [12]	73
Data browser	70

The Pima diabetes problem is a notoriously difficult machine learning problem. Part of this difficulty arises from the fact the dependent output variable is really a binarised form of another variable which itself is highly indicative of certain types of diabetes but does not have a one-to-one correspondence with the condition of being diabetic. To date no machine learning approach has obtained an accuracy higher than 78% [9]. The discovered ACGF models have yielded very high accuracies (79.7%), outperforming other machine learning approaches (see Table 1). The transparency of the induced Cartesian granule feature models suffers a little from the use of four dimensional features.

3.2 Modelling a Dynamical System - The Box-Jenkins Gas Furnace Problem

This example deals with the widely used benchmark problem of modelling a gas furnace (an example of a dynamical process) which was first presented by Box and Jenkins [13]. The modelled system consists of a gas furnace in which air and methane are combined to form a mixture of gases containing CO_2 (carbon dioxide). Air fed to the furnace is kept constant, while the methane feed rate can be varied in any desired manner. The furnace output, the $C0_2$ concentration, is measured in the exhaust gases at the outlet of the furnace.

The dataset here corresponds to a time series consisting of 296 successive pairs of observations of the form $(u(t), y(t))$, where $u(t)$ represents the methane gas feed rate at the time step t and $y(t)$ represents the concentration of CO_2 in the gas outlets. The sampling time interval is nine seconds. Using a time-discrete formulation, the dynamics of the system is represented by a relationship that links the predicted system state $y(t+1)$ to the previous input states $u(t_i)$ and the previous output states $y(t_i)$, that is $y(t+1)$ is a function of the previous input and output states i.e. $y(t+1) = f(u(t_1), u(t_2), \ldots, u(t_n), y(t_1), y(t_2), \ldots, y(t_n))$. Here we have set the value of n to five. Consequently we consider ten input variables and our database reduces to 291 data tuples of the form $(u(t), u(t-1), \ldots, u(t-4), y(t), y(t-1), \ldots, y(t-4), y(t+1))$.

In the case of this problem all data tuples were considered for both training and testing. The main reason for this is provide a comparison with other approaches presented in the literature. We applied the G_DACG constructive induction algorithm to the gas furnace problem. All ten base features were considered and Cartesian granule features of dimensionality up to five with granularity ranges of [2, 12] were considered thus yielding a multi-million node search space. The output universe was uniformly partitioned using eight triangular fuzzy sets. The G_DACG algorithm iterated for fifty generations (or if the stopping criterion was satisfied it halted earlier, arbitrarily set at an mean square error (MSE) of less than 0.05). As a result of the G_DACG process an additive Cartesian granule feature model where each rule consists of two Cartesian granule features was deemed to be the most suitable model. The model consists of eight rules and a trapezoidal fuzzy set with 50% overlap was determined to be the best input feature granule characterisation. The performance accuracy of the model was measured based upon the mean square error (MSE) between the actual data outputs and the model outputs.

The discovered model yields a relatively low MSE of 0.128. In Figure 2 the model performance is compared with the original data. Increasing the granularity of the output universe (and consequently the number of rules) can lead to models with lower MSE, however, this also leads to more complex models. For example, if the granularity of the output universe is increased to ten the MSE of the model drops to 0.11.

The gas furnace problem serves as a benchmark problem in the field of system identification and has been tested on many learning approaches. Table 2 compares some of the results of the more common statistical and fuzzy based techniques with the ACGF modelling approach. Overall the ACGF modelling approach generally outperforms the other fuzzy and statistical based approaches from an accuracy perspective. The Takagi-Sugeno linear model gives the best performance accuracy, however it lacks the transparency provided by the other approaches including that of ACGF modelling. The models generated by the various approaches were evaluated on the same data that was used to generate them. As a result the results provided no information on the generalisation powers of the extracted models. From a model transparency, the extracted ACGF model is relatively easy to interpret since the extracted Cartesian granule fuzzy sets are all two dimensional in nature. The various fuzzy approaches listed in Table 2, differ mainly in the identification algorithms used. In general, they use local hill climbing strategies and treat the steps of input variable selection and abstraction separately, which may subsequently result in models which are only locally optimum.

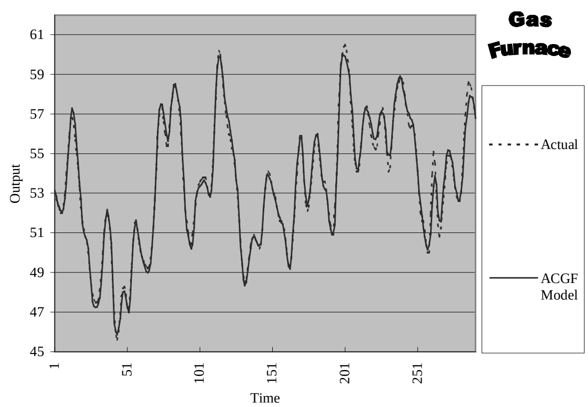

Figure 2: ACGF model predictions versus the actual data for gas furnace problem.

Table 2: Comparison of results for the gas furnace problem.

Approach.	MSE
Box & Jenkins Statistical (1970) Approach[13]	0.710
Tong(1980) Fuzzy Model [14]	0.469
Pedrycz(1984) Fuzzy Model [15]	0.320
Linear Model [16]	0.193
Takagi-Sugeno Linear Model (1993) [16]	0.068
Fuzzy Position gradient model(1993) [16]	0.190
Nakoula et al's. Fuzzy model (1997) [17]	0.175
Additive Cartesian granule feature Model	0.128

3.3 Modelling Image Interpretation

We address the problem of modelling region classification, within the context of digital image library system. The goal is to automatically acquire a system from example data, which can classify image regions. In this case the number of classes considered is two: *Road* and *notRoad.* Eighty images of typical outdoor rural scenes were selected from the Bristol image database [18, 19]. Subsequently these images (characterised by intensity) were segmented into *road* and *non-road* regions using the *k*-means segmentation algorithm, where *k* was set to 4. This resulted in 13,628 regions being generated. Feature values were subsequently generated for each region feature. Non-overlapping training (70%), validation (15%) and test sets (15%) of regions were subsequently generated in a class-wise manner. For the road classification problem each segmented image region was described using a set of over sixty features, comprising of colour, location, orientation, size, shape and texture features. In order to reduce the complexity of the learning process a neural network-based "filter" feature selection algorithm was applied to this feature set [20]. This resulted in ten features been selected as representative features for task of road classification. Table 3 describes the features that were selected for the subsequent induction step. A full description of all features considered is presented in [20]. When the G_DACG algorithm was applied to this problem it resulted in models similar to that in Figure 3. Models such as this provide accuracy levels of 97%, which perform as well or better than neural network (97%) and decision tree approaches (92.5). The resulting Cartesian granule feature models prove to be the most transparent as the decision trees tend to be bushy and the neural networks opaque.

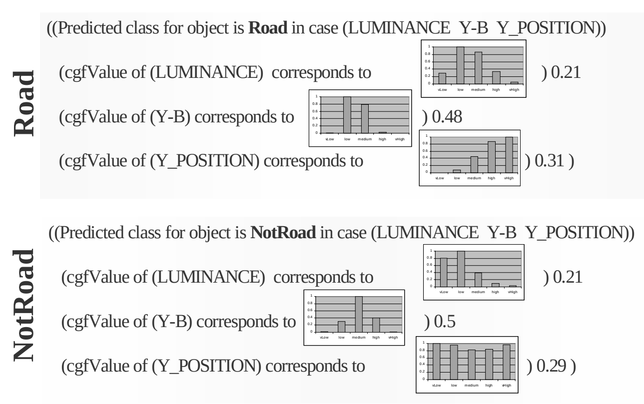

Figure 3: Additive Cartesian granule feature model for road classification.

4. Conclusions

Knowledge discovery from a Cartesian granule feature perspective has been presented. The approach had been illustrated and compared with a variety of more traditional approaches on a variety of real world problems. Overall the proposed approach performed well in terms of accuracy and transparency compared to other algorithms.

Table 3: Selected features for each region that are considered for learning.

10 Selected FEATURES	
No.	**Features**
0	Luminance
1	Red-Green
2	Yellow-Blue
3	Centroid (X, Y)
5	Orientation 1
6	Orientation 2
7	Shape 1 (principle mode)
8	Texture Gabor_{128} – high frequency, isotropic
9	Texture Gabor_{256}– high frequency, isotropic

Acknowledgements

James Shanahan mainly carried out this work while at the University of Bristol on a EU Marie Curie Fellowship.

Bibliography

1. Quinlan, J.R., *Induction of Decision Trees.* Machine Learning, 1986. **1**(1): p. 86-106.
2. Kibler, D. and D.E. Aha. *Learning representative exemplars of concepts*. in *International Workshop on Machine Learning*. 1987.
3. Bishop, C.M., *Neural Networks for Pattern Recognition*. 1995, Oxford: Clarendon Press.
4. Baldwin, J.F., T.P. Martin, and J.G. Shanahan. *Modelling with words using Cartesian granule features*. in *FUZZ-IEEE*. 1997. Barcelona, Spain: pp 1295-1300.
5. Shanahan, J.G., *Cartesian Granule Features: Knowledge Discovery of Additive Models for Classification and Prediction*, . 1998, PhD Thesis, Dept. of Engineering Maths, University of Bristol, Bristol, UK.
6. Baldwin, J.F., T.P. Martin, and B.W. Pilsworth, *FRIL - Fuzzy and Evidential Reasoning in A.I.* 1995: Research Studies Press(Wiley Inc.), ISBN 086380159 5.
7. Baldwin, J.F., T.P. Martin, and J.G. Shanahan, *System Identification of Fuzzy Cartesian Granule Feature Models using Genetic Programming*, in *IJCAI Workshop on Fuzzy Logic in Artificial Intelligence, Lecture notes in Artificial Intelligence (LNAI 1566) - Fuzzy Logic in Artificial Intelligence*, A.L. Ralescu and J.G. Shanahan, Editors. 1998, Springer. p. 26.
8. Baldwin, J.F., T.P. Martin, and J.G. Shanahan, *Controlling with Words using Automatically Identified Fuzzy Cartesian Granule Feature Models.* (To appear) International Journal of Approximate Reasoning (IJAR) dedicated the International Summer School on Fuzzy Logic Control: Advances in Methodology, 1998: p. 37.
9. Merz, C.J. and P.M. Murphy, *UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA*. 1996, Irvine, CA: University of California.
10. Baldwin, J.F., J. Lawry, and T.P. Martin. *Mass assignment fuzzy ID3 with applications*. in *Fuzzy Logic: Applications and Future Directions Workshop*. 1997. London, UK.
11. Murphy, S.K., S. Kasif, and S. Salzburg, *A system for induction of oblique decision trees.* Journal of Artificial Intelligence Research, 1994. **2**: p. 1-33.
12. Michie, D., D.J. Spiegelhalter, and C.C. Taylor, eds. *Machine Learning, Neural and Statistical Classification*. . 1993, Ellis Horwood: New York, USA.
13. Box, G.E. and G.M. Jenkins, *Time series analysis, forecasting and control*. 1970, San Francisco, CA: Holden Day.
14. Tong, R.M., *The evaluation of fuzzy models derived from experimental data.* Fuzzy Sets and Systems, 1980. **4**: p. 1-12.
15. Pedrycz, W., *An identification algorithm in fuzzy relational systems.* Fuzzy Sets and Systems, 1984. **13**: p. 153-167.
16. Sugeno, M. and T. Yasukawa, *A Fuzzy Logic Based Approach to Qualitative Modelling.* IEEE Trans on Fuzzy Systems, 1993. **1**(1): p. 7-31.
17. Nakoula, Y., S. Galichet, and L. Foulloy, *Identification of linguistic fuzzy models based on learning*, in *Fuzzy Model Identification*, H. Helledoorn and D. Driankov, Editors. 1997, Springer: Berlin. p. 281-319.
18. Campbell, N.W., B.T. Thomas, and T. Troscianko, *Automatic segmentation and classification of outdoor images using neural networks.* International Journal of Neural Systems, 1997. **8**(1): p. 137-144.
19. Mackeown, W.P.J., *et al. Contextual Image Labelling with a Neural Network*. in *IEE Vision, Speech and Signal Processing*. 1994.
20. Shanahan, J.G., *et al.*, *Transitioning from recognition to understanding in vision using additive Cartesian granule feature discovery models.* Journal paper in preparation, 1999: p. N/A.