Detecting Bright Band using AI techniques in radar hydrology

DANIEL R. McCULLOCH, JONATHAN LAWRY, MIGUEL A. RICO-RAMIREZ & IAN D. CLUCKIE

Water and Environmental Management Research Centre, University of Bristol, Bristol BS8 1TR, UK

daniel.mcculloch@bristol.ac.uk

Abstract An important tool in flood prediction is quantitative precipitation estimation using weather radars. This study focuses on the VRPs (vertical reflectivity profiles) obtained from S-Band RHI scans from the Chilbolton radar in the UK. The RHI scan consists of several beams, each one elevated at an incremental angle. In an RHI scan pixels represent reflectivity with a corresponding range and height above the Earth's surface. The VRP is extracted by a polar to rectangular transformation. During stratiform rainfall a region of enhanced reflectivity associated with echoes from melting snowflakes is commonly observed. This region is known as the Bright Band. It is necessary to detect the Bright Band region and correct for its presence accordingly. However, previous work has shown Bright Band classification to be rather problematic. Various AI algorithms are introduced, namely: Naive Bayes, Fuzzy Naive Bayes, LID3 (Linguistic ID3), and applied to the Chilbolton data set in order to classify Bright Band regions pixel by pixel. Once trained, these algorithms have the added value of being computationally very cheap compared with current detection models. To estimate class probabilities we attempt to discretize data using uniform discretization and use two more sophisticated discretization methods—Entropy Minimization Partitioning and K-Means Clustering. The algorithms are shown to perform well, especially when incorporated with these more sophisticated discretization techniques, which result in a reduction in the number of partitions required for discretization, which in turn reduces the computational requirements of the algorithms. The Naive Bayes algorithms were effective, but attention was focused in particular on LID3, a linguistic version of the decision tree induction algorithm ID3. LID3 incorporates uncertainty and fuzziness into its input variables, in an attempt to infer a more robust model. In addition LID3 generates linguistic rules for the classification of Bright Band at the decision making level, which can improve our understanding of the underlying relationships between measurements and classes. This work is representative of a number of new approaches currently being applied in the Radar Hydrology area.

Key words Bright Band; entropy-minimization partitioning; fuzzy naive Bayes; K-means clustering; LID3; naive Bayes

INTRODUCTION

The quantitative use of radar-based precipitation estimations in hydrological modelling for flood forecasting has been limited due to different sources of uncertainty in the rainfall estimation process. Some of the factors affecting radar rainfall estimations include radar calibration, signal attenuation, clutter and anomalous propagation, variation of the Vertical Reflectivity of Precipitation (VPR), range effects, Z-R relationships, variation of the drop size distribution, vertical air motions, beam overshooting the shallow precipitation and sampling issues among others (Battan, 1973; Austin, 1987; Doviak & Zrnic, 1993; Collier, 1996).

The VPR is an important source of uncertainty in the estimation of precipitation using radars. The variation is largely due to factors such as the growth or evaporation of precipitation, the thermodynamic phase of the hydrometeors, or melting and wind effects. As the range increases from the radar, the radar beam is at some height above the ground, while the radar sampling volume increases and is unlikely to be homogeneously filled by hydrometeors. As an example, the lower part of the volume could be in rain, whereas the upper part of the same volume could be filled with snow, or even be without an echo. This variability affects reflectivity measurements and the estimation of precipitation may not represent the rainfall rate at the ground. Snowflakes are generally low-density aggregates and when they start to melt they look like big raindrops to the radar, resulting in larger values of reflectivities compared to the expected reflectivity below the melting layer (Battan, 1973). This phenomenon is called “Bright Band” and the interception of the radar beam with melting snowflakes can cause significant overestimates of precipitation up to a factor of 5, and when the radar beam is above the Bright Band can cause underestimates of precipitation up to a factor of 4 per kilometre above the Bright Band (Joss & Waldvogel, 1990).

The Bright Band can be seen as the very dark region in RHI (Fig. 1(a)) and PPI (Fig. 1(b)) scans. The power reflected back to the radar is related to the rainfall intensity and therefore radar beams striking this melting layer of snow causes overestimation of precipitation (Rico-Ramirez, 2004). Therefore the Bright Band needs to be detected and corrected for. In addition to this, when estimating precipitation intensity, determining which hydrometeors the beam intersects is crucial to the calculation.

This paper applies three AI algorithms to the Bright Band problem. It is hoped that these algorithms that learn from example will be able to learn the nonlinear and uncertain relationship between the input attributes and their class, and accurately classify images pixel by pixel in real time. Naive Bayes, is a very simple and well known AI algorithm that is based on Bayes theorem, as well as assuming conditional independence between attributes to classify instances. Label Semantics, proposed by Lawry (2005), is a framework for modelling with linguistic expressions. Label Semantics incorporates the notion of overlapping fuzzy sets and partial membership of labels. Label semantics is applied to Naive Bayes and the commonly known decision tree ID3. These algorithms are called Fuzzy Naive Bayes (Discrete Version) and LID3 (Linguistic ID3).

DATA

RHI scans from the Chilbolton weather radar have been used for this analysis. The Chilbolton radar is operated by the Radio Communications Research Unit (RCRU). It is an S-band (9.75 cm wavelength) weather radar developed to study the effects of rain on communication systems (Goddard *et al.*, 1994). It is currently the largest steerable meteorological radar in the world, with a 25 m diameter antenna, allowing very high

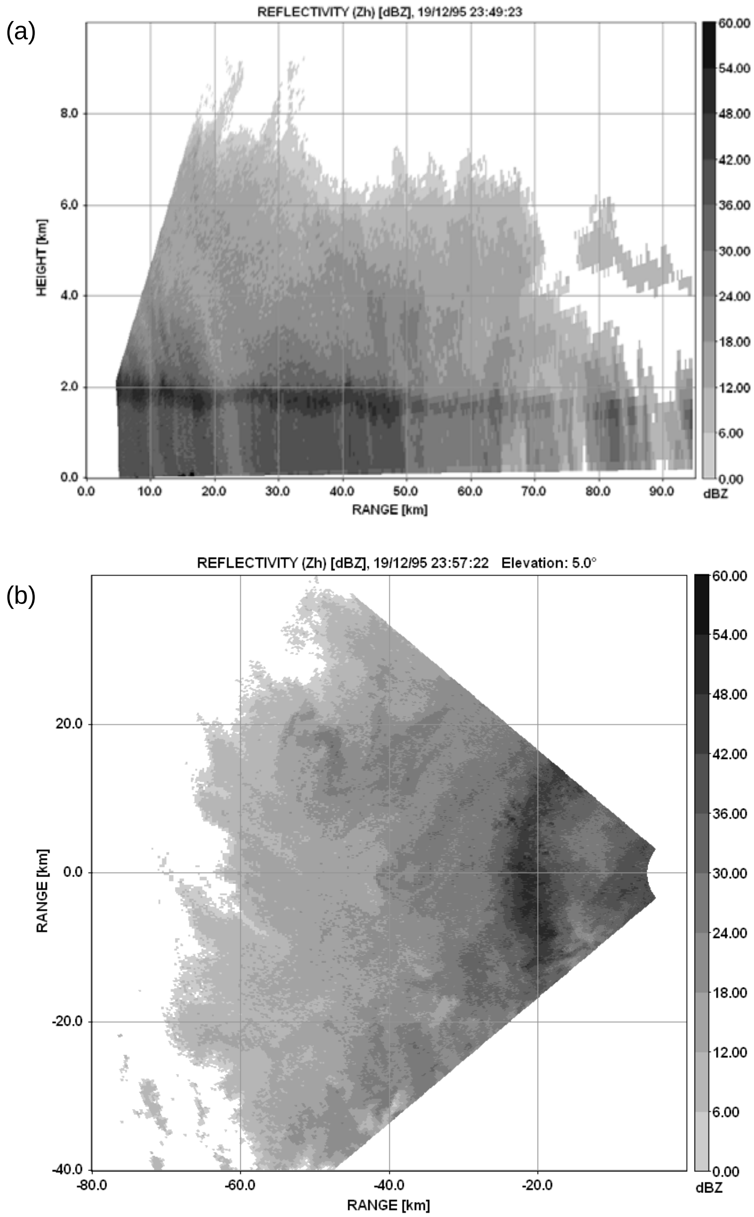

Fig. 1 A Typical RHI (a) and PPI (b) scans for a stratiform event.

resolution measurements from precipitation particles with a very narrow beam width of 0.25 degrees and 300 m gate sizes. The Chilbolton radar has a dual-polarization capability, which allows the study of the size, shape, phase and orientation of the hydrometeors.

Our interests are with the vertical reflectivity profiles obtained from S-Band RHI scans from the Chilbolton radar (see Rico-Ramirez, 2004; Rico-Ramirez *et al*., 2005, for further details). The measurements obtained are the Reflectivity Factor (Z_h), the Differential Reflectivity (Z_{dr}), the Linear Depolarisation Ratio (Z_{dr}) and the height of the measurement (H_0). The estimated Bright Band boundaries can then be determined by performing a vertical search for the largest differential in reflectivity. The image cannot be classified in real time as computations can only be performed once the whole image is presented. Ramirez's rotation algorithm (Rico-Ramirez & Cluckie, 2007) performs this maximum reflectivity differential search in order to determine the estimated boundaries of the Bright Band.

THE CLASSIFIERS

The Naive Bayes classifier

Naive Bayes is one of the most simple and well known algorithms in AI. It simply uses Bayes theorem and assumes conditional independence between attributes to obtain a conditional probability for a class given some input variables. A brief summary is presented in this paper. Given the continuous input variables $X = x_1, \ldots, x_n$, suppose that the output variables are partitioned into classes $C = C_1, \ldots, C_t, \ldots, C_T$, where T is the number of classes and n is the number of attributes.

Now let us consider Bayes' theorem to determine the probability of the (C_t) given an input (X):

$$P(C_t \mid X) = P(C_t)\frac{P(X \mid C_t)}{P(X)} \tag{1}$$

Now, assuming that each input variable is independent of one another given class (C_t) we have an estimation of $P(C_t \mid X)$ given by:

$$P(C_t \mid X) = \frac{P(C_t)\prod_{j=1}^{n} P(x_j \mid C_t)}{P(X)} \tag{2}$$

For our need, classification, we merely need to choose the class C_t with the highest value of $P(C_t \mid X)$. Note that because we are only interested in the relative size of $P(C_t \mid X)$ for each class we can completely disregard the denominator $P(X)$ as it is the same value for all classes.

The Naive Bayes classifier—continuous or discrete?

As discussed, we will need to evaluate the numerator of equation (2) in order to classify instances using Naive Bayes. When estimating $P(x_j \mid C_t)$ there are two general approaches, depending on whether the data has been discretized or left continuous. For discrete attributes we determine $P_r(x_j \mid C_t)$ for each interval and class by simply determining the frequency within each discretization given the class and dividing by the number of examples belonging to that class. The continuous method assumes the numerical values to a have a Gaussian probability distribution. Classification is estimated by a probability density function with mean μ and standard deviation θ given by:

$$f(x) = \frac{1}{\sqrt{2\pi\theta}} e^{\frac{(x-\mu)^2}{2\theta^2}} \tag{3}$$

When analysing the Chilbolton Bright Band data set the probability distribution of all four attributes is not strictly Gaussian. In particular, the attribute H_0 (height) seems to be multimodal. This clearly violates the Gaussian assumption of the density function (3) and could introduce inaccuracies if the attribute is important to the overall classifier. Therefore the discretization method is used in place of the continuous method, for all of the classifiers in this paper.

Discretization

There are numerous ways to discretize an attribute space, the simplest being uniform discretization. In uniform discretization each variable is split into a predetermined number of bins equally spaced. However, in data sets that are clustered and distributed unevenly throughout the input space, we obtain some bins containing a great deal of data while others contain almost none. For this reason, in addition to uniform discretization, two more sophisticated algorithms are considered, one supervised and one unsupervised discretization. Unsupervised discretization methods consider only the values of given attributes. Whereas, supervised discretization methods consider the values of an attribute space as well as their corresponding class, thereby containing discretizations dominated by a particular class, theoretically aiding learning. The unsupervised discretization method we will consider is K-Means Clustering and the supervised discretization method is Entropy Minimization Partitioning.

Label semantics

Label semantics, proposed by Lawry (2005) is a framework for modelling with linguistic expressions, or labels such as *small medium large*. Such labels are defined by overlapping fuzzy sets which cover the universe of continuous variables (Qin & Lawry, 2004).

Consider an example *x*, in a continuous universe Ω, which is represented by a set of linguistic labels $LA = \{L_1, .., L_b\}$. Fuzzy set theory introduced by Zadeh (1965), considers these labels to overlap and allows *x* to have partial membership in more than one label. In our case we will only allow an element to have partial membership of two labels, with 50% overlap of the continuous universe (Fig. 2).

Since the labels overlap we cannot define probability distributions for $LA = \{L_1, .., L_b\}$, as more than one label can be appropriate, for an example *x* we define a set of focal elements corresponding to atomic expressions which are both exclusive and exhaustive (see Fig. 3) For example, if $LA = \{\{s\},\{m\},\{l\}\}$ then there are 8 possible atoms of the form $s\wedge\neg m\wedge\neg l$, $s\wedge m\wedge\neg l$, $\neg s\wedge m\wedge\neg l$ etc. These are represented by focal set $\{\{s\},\{s,m\},\{m\},\{m,l\},\{l\}\}$ (see Lawry, 2005, for more details).

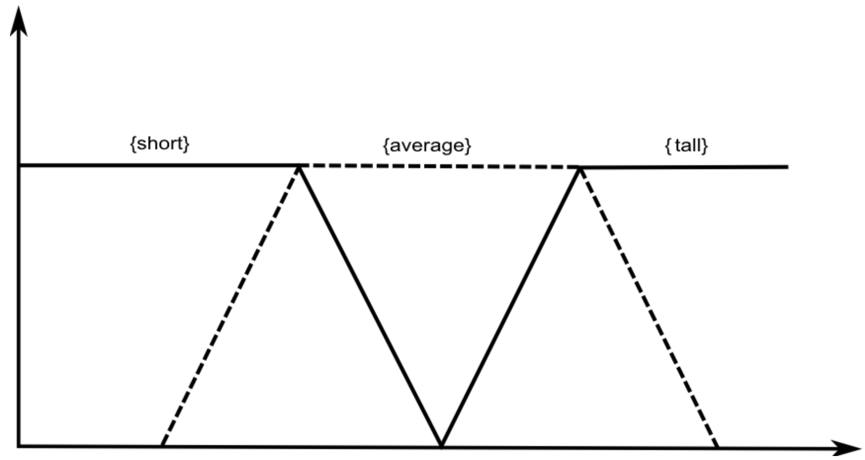

Fig. 2 An example of three uniformly distributed trapezoidal fuzzy sets with 50% overlap.

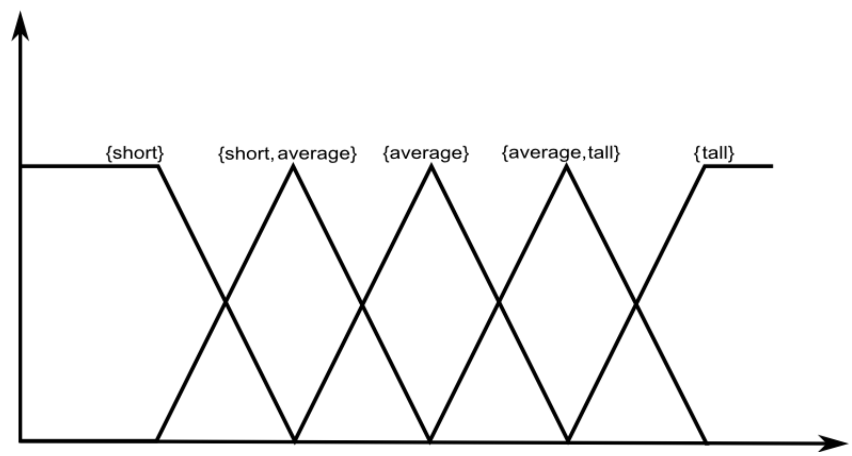

Fig. 3 An example of a focal set derived from Fig. 2 with focal elements *{short}*, *{short,average}*, *{average}*, *{average,tall}*, *{tall}*.

LID3

The ID3 classifier described by Quinlan (1986) is a very well known and widely used decision tree algorithm for data sets with discrete attributes. ID3 builds a decision tree from training examples. As the tree is constructed a decision is made at each new node, about which attribute it should split on next. The attribute that maximizes the information gain (minimizes expected entropy) is chosen. We then iterate to the next node, where this process is repeated until the tree is complete. Qin & Lawry (2004) propose an LID3 classifier that incorporates label semantics, in an attempt to increase the robustness of ID3 leaving it less susceptible to misclassification due to crisp discretization. The full details of this algorithm are given in Qin & Lawry (2004). Figure 4 is an example of a LID3 decision tree with depth 2, where the focal set on attribute L_{dr} contains the focal elements *{low}({l})*, *{low,high}({l,h})* (notice overlap of label low

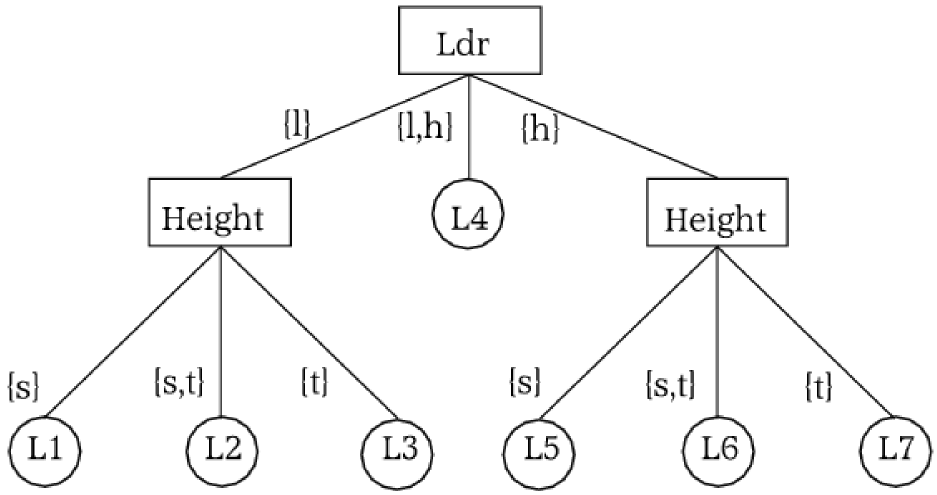

Fig. 4 A typical LID3 depth 2 decision tree illustrating the attributes L_{dr} and Height and their corresponding Focal Elements.

Table 1 Classification accuracies and standard deviations for all models.

Classification Technique	Classification accuracy and standard deviation			
	Rain	Snow	Bright Band	Average
Naïve Bayes (Continuous)	75.4%±0.006	68.5%±0.007	98.6%±0.002	80.8%
Naïve Bayes (Uniform)	78.2%±0.005	81.9%±0.005	97.5%±0.002	85.9%
Naïve Bayes (Entropy Minimization)	78.7%±0.010	82.0%±0.005	97.8%±0.008	86.2%
Naïve Bayes (K-Means)	82.5%±0.005	78.8%±0.004	99.3%±0.001	86.9%
Fuzzy Naïve Bayes (Uniform)	84.9%±0.008	77.4%±0.004	98.9%±0.001	87.1%
Fuzzy Naïve Bayes (Entropy Min.)	85.2%±0.005	80.6%±0.003	98.7%±0.001	88.2%
Fuzzy Naïve Bayes (Uniform)	84.9%±0.004	81.6%±0.002	99.0%±0.001	88.5%
LID3 (Uniform)	89.4%±0.004	85.2%±0.002	99.3%±0.001	91.3%
LID3 (Entropy Minimisation)	92.3%±0.002	87.5%±0.002	99.4%±0.001	93.1%
LID3 (K-Means)	93.5%±0.002	89.0%±0.004	99.4%±0.001	94.0%

and high) and *{high}*(*{h}*). The names of these labels are arbitrary and simply to illustrate the notion of focal elements.

MAIN RESULTS

A Continuous and Discrete Naive Bayes, Discrete Fuzzy Naive Bayes and a LID3 classifier were applied to the Chilbolton data set. The data set was separated into training and testing data sets using 10-fold cross validation. Discretization was performed using Uniform, K-means Clustering, and Entropy-Minimisation Partitioning. The accuracies for each classifier and discretization method are shown in Table 1.

A typical RHI scan is shown (Fig. 5) before hydrometeor classification with LID3 (Fig. 6). The 3 regions seen in Fig. 6 are rain (lower region), snow (upper region) and Bright Band (middle region). Overall LID3 was the leading performer by some margin. When discretized with K-Means Clustering, an overall accuracy of 94% was achieved.

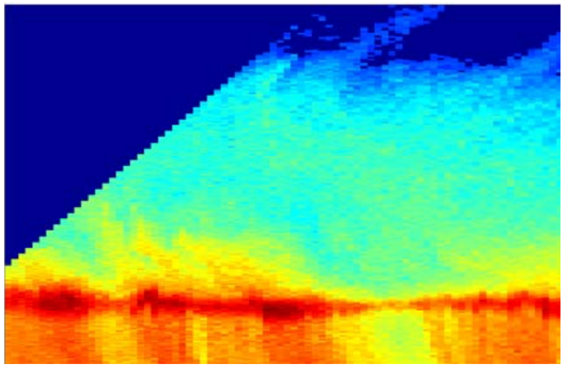

Fig. 5 Typical RHI scan.

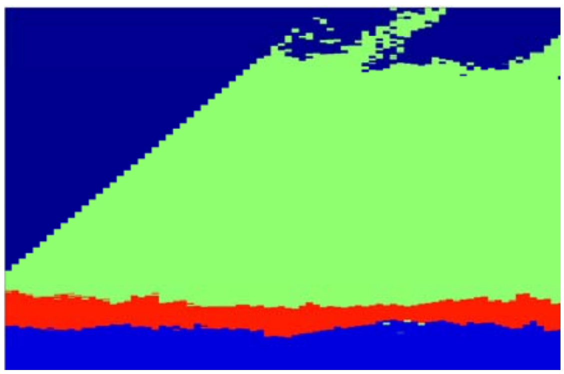

Fig. 6 RHI scan after LID3 classification.

CONCLUSIONS

All three classifiers performed excellently on the Bright Band, regardless of the discretization method used. The classification of snow and rain was rather more challenging and varied between classifier and discretization method. The ranking was LID3 > Fuzzy Naïve Bayes > Discrete Naïve Bayes > Continuous Naïve Bayes. In terms of discretization methods both K-Means and Entropy Minimization Partitioning clearly outperformed Uniform Discretization. In addition to this, Entropy Minimization Partitioning and K-Means requiring 30% fewer discretizations than the uniform method creating a more efficiently partitioned attribute space. This greatly reduces the computational expense as well as increasing the robustness of the model with a lesser chance of over-fitting.

Let us consider our proposal to discretize our data rather than assume a Gaussian distribution. We can see from Table 1 that the continuous version of Naive Bayes certainly lacked performance compared to the three discretization methods considered. This supports our analysis of the data set and demonstrates the downfall of models that often assume a Gaussian distribution.

We can see from Table 1 that in general, Fuzzy Naive Bayes outperforms Naive Bayes. This is because Fuzzy Naive Bayes is less susceptible to misclassification due to crisp partitions, and is more robust to noise and heavily overlapping attribute spaces.

Notice that LID3 is a far more accurate classifier than the Naive Bayes Methods. This is probably due to the Naive Bayes assumption, that each attribute is independent given the class. As Z_h, Z_{dr} and L_{dr} are highly correlated, this assumption is not completely valid and LID3 is a much more suitable model for classifying Bright Band.

Both K-Means and Entropy Minimization Partitioning outperformed Uniform classification by some margin, with K-Means improving slightly on Entropy Minimization Partitioning. Here unsupervised discretization is favoured to supervised discretization and it is worth mentioning that in some cases supervised discretization can lead to slight over fitting. This seems to be the case in this instance. Nevertheless Entropy Minimization Partitioning should be used in favour of the basic Uniform method.

LID3 not only achieved very good classification accuracy, but was able to generate rules for the classification of Bright Band. Let us consider the attributes of the underlying problem, Z_h, Z_{dr}, L_{dr} and height. When assessing the resulting decision tree from LID3, the single most important attribute is L_{dr}, followed by height for classifying Bright Band instances. In many instances the Bright Band is determined solely on the L_{dr} value, where the tree terminates and all other attributes are ignored. It is possible to analyse the importance of particular attributes and how they contribute to the overall classifier. It is the rule induction and attribute ranking that makes decision trees such as LID3 a valuable analytical tool in data mining large databases, such as RHI scans to understand further the main properties of the Bright Band phenomenon. This is a huge advantage over most machine learning algorithms which are often either black box or lack the transparency of LID3.

Overall, all the algorithms discussed in this paper have performed fairly well on the given Bright Band problem. It is also apparent that the way we discretize continuous variables has a very big overall effect on computational expense and more general rule generation, as well as classification accuracy. LID3 stands apart from the other two algorithms in terms of accuracy and the added property of rule induction. It is hoped that this rule generation could be a good analytical tool to learn more about the classification of Bright Band, which is still not completely understood.

FURTHER WORK

In terms of simply increasing prediction accuracy there is a vast amount of possibilities in AI. SVMs (Support Vector Machines), Neural Networks, Bayesian Networks, etc. However, there is always the problem of computational complexity on large databases, which algorithms like SVMs have suffered from in the past. One method of particular interest to the author is machine learning fusion techniques. One example being bagging, whereby many models amalgamate their output into a single classification (Quinlan, 1986). There is also a more sophisticated version of Naive Bayes known as Semi Naive Bayes proposed by Lewis, whereby attributes are grouped so that the conditional independence assumption is not violated.

In these studies the input space was fuzzified for two of the algorithms. Let us consider the output space, classes snow, rain, Bright Band, and our uncertainty on these class boundaries. It seems completely viable to fuzzify these, as these class boundaries clearly overlap. For instance there are instances with a degree of uncertainty, whether they are rain or Bright Band or both. The problem lies in how we overlap these classes. More information is needed about the uncertainty of these classes to begin in such a task.

Acknowledgements The authors would like to express their thanks to the following individuals and organizations without whose cooperation this research could not have been undertaken: the EPSRC Flood Risk Management Research Consortium (GR/S76304/01) and the Radio Communications Research Unit at the Rutherford Appleton Laboratory, in particular to John Goddard and Charles Kilburn, for providing the Chilbolton radar data.

REFERENCES

Austin, P. M. (1987) Relation between measured radar reflectivity and surface rainfall. *Mon. Weath. Rev.* **115**, 1053–1070.

Battan, L. J. (1973) *Radar Observation of the Atmosphere.* The University of Chicago Press, Chicago, USA.

Collier, C. G. (1996) *Applications of Weather Radar Systems*. John Wiley & Sons, Chichester, UK.

Doviak, R. J. & Zrnic, D. S. (1993) *Doppler Radar and Weather Observations.* Academic Press, New York, USA.

Goddard, J. W., Eastment, J. D. & Thurai, M. (1994) The Chilbolton Advanced Meteorological Radar: a tool for multidisciplinary atmospheric research. *Electronics & Commun. Engng J.* **6**, 77–86.

Joss, J. & Waldvogel, A. (1990) Precipitation measurement and hydrology. In: *Radar in Meteorology* (ed. by D. Atlas). (Battan Memorial and 40th Anniversary Radar Meteorology Conference), 577–606. Am. Met. Soc.

Lawry, J. (2001) *Label Semantics: A formal framework for modelling with words, Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. LNAI 2873, Springer-Verlag, USA.

Qin, Z. & Lawry, J. (2005) Linguistic rule induction on a random set semantics. In: *Fuzzy Logic, Soft Computing and Computational Intelligence* (*Proc. IFSA world congress (IFSA-05)*), 1398–1404. Springer-Tsinghua, China.

Quinlan, J. R. (1986) Induction of decision trees. *Machine Learning* **1**, 81–106.

Rico-Ramirez, M. A. & Cluckie, I. D. (2007) Bright Band detection from radar vertical reflectivity profiles. *Int. J. Remote Sens*. (in press).

Rico-Ramirez, M. A. (2004) Quantitative Weather Radar and the Effects of the Vertical Reflectivity Profile. PhD Thesis, University of Bristol, Bristol, UK.

Rico-Ramirez, M. A., Cluckie, I. D. & Han, D. (2005) Correction of the Bright Band using dual-polarisation radar. *Atmos. Sci. Letters* **6**, 40–46.

Witten, I. H. & Eibe, F. (2005) *Data Mining-Practical Machine Learning Tools & Techniques*, 2nd edn. Morgan Kaufmann, California, USA.

Zadeh, L. A. (1965) Fuzzy sets. *Inform. Control* **8**(3), 338–353.

Zengchang, Q. & Lawry, J. (2004) Decision tree learning with fuzzy labels. *Inform. Sci.* **172**, 91–129.