

Learning from uncertain image data using granular fuzzy sets and bio-mimetic applicability functions

Jonathan Rossiter¹ and Toshiharu Mukai

Bio-mimetic Control Research Center

RIKEN (The Institute of Physical and Chemical Research)

2271-130, Anagahora, Shimoshidami, Moriyama-ku, Nagoya, 463-0003, JAPAN

{rossiter@bmc.riken.jp, tosh@bmc.riken.jp}

Abstract

In this paper we present a new method for machine learning from images using uncertain information granules which has been inspired by bio-mimetic study of human perceptual processing. We present a method for generating labelled image data using the rapid application of rough labelled regions to the image under study. Over each region is defined an applicability function which acts as a centre of focus for the uncertain information contained within the region. We present a number of alternative applicability functions inspired by the human visual system and particularly by the centre-weighted effect of the fovea within the retina. We also show how this uncertain data can be used to directly train a granular fuzzy machine learner.

Keywords: machine learning with words, image processing, bio-mimetic, fovea-like functions.

1 Introduction

In this paper we are concerned with learning uncertain class models from images. Where our approach differs from other uncertainty-based approaches in the image processing field is the ability to deal with uncertain image information and the bio-mimetic nature of this approach at both the low, visual processing, level and the high, computing with words, level.

Where conventional machine learning from images generally takes a crisp dataset, i.e. a dataset consisting of tuples of crisp feature values and their associated class labels, here we permit the data tuples to

¹The first author is also a lecturer at the University of Bristol, UK, and can be contacted via email to: Jonathan.Rossiter@bris.ac.uk

also possess an uncertainty value, referred to here as applicability. Where conventionally the process is,

crisp image data $\xrightarrow{\text{induction}}$ **uncertainty model**

the process outlined in this paper becomes,

uncertain image data $\xrightarrow{\text{induction}}$ **uncertainty model**

This is extremely important for the set of applications where a labelled dataset of examples is not available. In these cases one method of obtaining training data is to ask an expert to label some suitable images. Unfortunately this is where we hit a number of serious problems, as outlined in the next section.

This paper presents two important contributions to solving these problems: a rapid bio-mimetic method for the generation of uncertain image data, and a method for the induction of fuzzy granular models from this uncertain image data.

2 Extracting uncertain image data

Where labelled examples for use in training do not exist we must ask a domain expert to label images by hand. Unfortunately labelling at the pixel level is not without problems:

- Experts have no time.
- Experts are expensive.
- The effects of bias, tiredness, frustration and satisfaction of search [4] soon come into play.

To combat these problems one may be tempted to pre-process the images to make it easier for the expert to label large areas more efficiently and more

quickly. Unfortunately this very pre-processing raises a further problem: pre-processing distorts the problem space. Since there is no “correct” segmentation or pre-processing method, the very method chosen can severely bias the expert’s perceptions of the images. This image pre-processing can effectively “lead” the expert.

To enable more natural and unbiased labelling we adopt a bio-mimetic method that is both quick and intuitive for the experts. In this paper we present a labelling method that employs “quick and dirty” labelling using regions of interest over which uncertainty is implicitly distributed. An applicability function defines this uncertainty as a function of position within the region of interest. In effect we are shifting the image processing from the pre-labelling stage, where it can bias the expert, to the post-labelling stage.

We present a number of possible functions that can describe this applicability. We take inspiration from the human visual system to define functions analogous to the distribution of light sensitive cells in the human retina, and specifically the region of high spatial and colour sensitivity in the center of the retina called the *fovea*.

2.1 The uncertain image labelling process

We now present a method for extracting labelled data from images using roughly labelled regions of interest.

It is interesting at this point to note that this process also has analogies with the imprecise language used by domain experts when they are called upon to describe image and class features within the problem at hand. For example, a medical expert might summarize a medical image with a single sentence such as “there is extensive emphysema in the upper left lung and some fibrosis around the periphery”. Here we have vague information about both the disease features and the position of these features in the image. If we can contain such vague information in a suitable data model we can allow experts to use similar vague concepts when labelling images.

To this end we present a method for labelling images and generating suitable data from which to learn. The general process is as follows,

1. Display the image to the expert.

2. The expert then applies rough patches, or “blobs”, to the image each of which has an associated class label. These patches can be any shape and can overlap.
3. These patches and their class labels are processed along with an applicability function to generate image sub-regions, each of which has a membership in one or more classes.

Each expert-labelled region can be interpreted as a single linguistic sentence along the lines of “this region is a X example of class Y” where X may be some linguistic term such as *poor*, *good* etc. and Y is a class label. Using this interpretation we can link the low-level image to the high-level descriptions using words. Figure 1 shows four such tagged regions on an image of a fibre composite material. The dark regions indicate the class “resin” and the light regions are the class “matrix”.

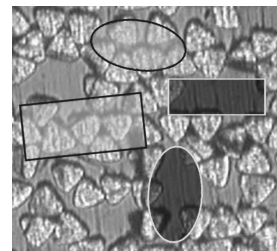


Figure 1: Examples of labelled regions on an image

By using this technique we have moved away from the requirement of conventional image processing approaches that data be crisp examples of the problem domain towards a more free and natural generation of uncertain training data. Although the uncertain data is, by its very nature, not completely correct or crisp, it *is* natural and easy for the experts to both generate and understand.

[3] describes the labelling process in more detail and the extraction of sub-region data from the larger expert-labelled image regions. In short, for every expert-labelled region R in an image, we sub-sample across it conditioned on the applicability function A , the class label L associated with that region and the chosen set of image processing features F to generate a set of n feature values $\{f_1, \dots, f_n\}$ with applicability value a and associated class L . This is performed in two stages,

$$[R, A, L, F] \rightarrow \{[r_i, a_i, L, F] \mid i = 1 \dots j\}$$

$$[r, a, L, F] \rightarrow [\{f_1, \dots, f_n\}, a, L,] \quad (1)$$

The first stage performs sub-sampling of region R into sub-samples r_1, \dots, r_j and calculation of corresponding applicability values a_1, \dots, a_j . The second stage calculates image feature values for each feature in each sub-sample.

Figure 2 shows an elliptical region R sub-sampled into smaller regions r .

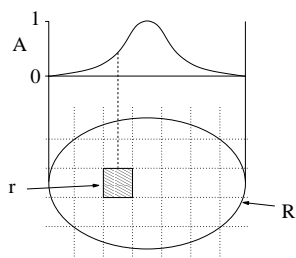


Figure 2: applicability across the region of interest

Applying this method to all expert-labelled regions we thus generate a training database of image processing feature values, along with their associated class labels and applicability values. This data is then used to train an uncertain granular fuzzy set model as outlined in section 4.

3 The applicability function A

The data generated in the previous section is reliant on an applicability function A defined over the expert-labelled region, as shown in figure 2a. This function defines the suitability, or applicability, of the class label to each part of the region. We use the term applicability to distinguish this function from a membership function, which may also be present as a general measure of the membership of the whole region (not just a part) in the class label.

This applicability function is crucial to both the data generation process described in the previous section and the interpretability of the expert-derived information. Indeed, the choice of function may have deep impact on the confidence the expert has in the whole machine learning process. That is, if the applicability function is not intuitive to the expert he may have

problems labelling the image and also in interpreting his own labels.

In previous research ([3]) we have chosen a simple linear applicability function. In this paper we study the human visual perception system in a little more detail and take inspiration from it in generating suitable applicability functions. Again we take a bio-mimetic approach in order to mimic aspects of human vision to improve our artificial system.

3.1 Aspects of human visual perception

Recall that the expert is trying to label the image by drawing rough areas of interest and associating each one with a class label. It is now helpful to examine how the human visual system may impact upon both the shape and the size of the region chosen by the expert and also the applicability function we select.

Figure 3 shows a simple model of the human eye and it illustrates how light from an observed scene enters the eye and strikes the retina. The retina itself is not homogeneous but is made of two types of light sensitive cells, the *cones* and the *rods*. Cones are active most in bright conditions and have high colour sensitivity while rods are better at much lower light levels but are less colour sensitive. The distribution of cones and rods are non-uniform, with many more cones present at the center 5° of the retina in a region called the *fovea*. Moving away from the fovea the concentration of cones rapidly falls and rods dominate.

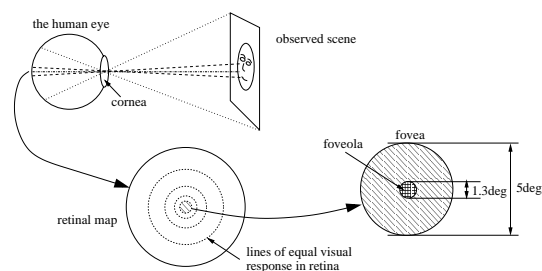


Figure 3: Simple diagram of visual perception

Figure 4 shows the densities of cones and rods at angular distance from the centre of the fovea, or *foveola*. See [2] for more background on human vision. Note the shape of the curve for cones. We now make the following point about how this physiology impacts on the expert labelling of images,

- For the purposes of this paper, the fovea is the

most crucial and most used region of the retina.

- This is justified in part by eye movements as humans survey a scene. These movements may also include the class of movement called saccade. If the above assumption were not the case then eye movement and especially saccades would be less prevalent and less important to scene observations.
- This leads us to focus on the fovea as inspiration for a human-like applicability function. That is, the fovea is a natural regioning mechanism in the retina and we suggest that such a regioning mechanism may be a suitable model for our applicability function.

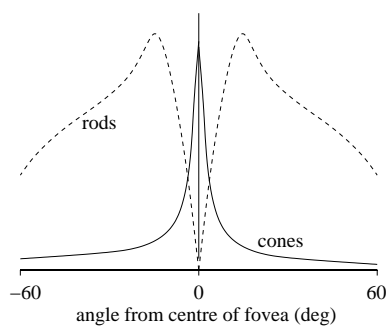


Figure 4: Cone/rod density distribution

To this end we now suggest some suitable applicability functions based on bio-mimicry of the phenomenon illustrated in figure 4

3.2 Fovea-like functions

While figure 4 shows the densities of both cones and rods, we must be a little careful in our interpretation of this figure. The points listed above hints at the fact that for our typical application (daylight/bright classification of images and scenes) the cones (and hence the fovea) contribute most. Yet the rods will also contribute to some degree. We can combine the contributing factors of both cones and rods into a rough class of applicability function. In this paper we assume the function is smooth and monotonic with maximum at the centre of the region of interest and minimum at the edge.

For simplicity we implement two families of fovea functions and an additional spatial scale option:

- **Trapezoidal functions.** Figure 5 shows a trapezoidal function defined across the distance from

the centre of the expert-defined region to the edge (i.e. the interval $[0, D]$). The point x can have any value in the interval $[0, D]$. Figure 5 shows a functional symmetry about $x = 0$ for ease of visualisation. By varying x we can change applicability from a smoothly graded function ($x = 0$) to a function having no practical effect ($x = D$).

- **Gaussian-like functions.** Here we approximate the centre-focus of the fovea with a simple Gaussian with variance σ . Note again that we are not trying to exactly match the distributions of cones in the fovea, rather the centre-weighting, or focus, of the retina. Figure 5, shows such a Gaussian applicability function. The function can be optionally normalized such that $A(0) = 1$ and/or $A(D) = 0$.

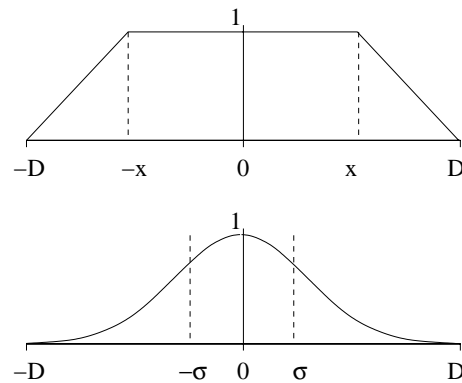


Figure 5: Possible applicability functions

- **Spatial scale.** Whilst the expert is able to define and label image regions of any size or shape the human fovea is constant in size. We therefore may be tempted to change the spatial scaling of our applicability functions to an *absolute* scale, that is, in proportion to the size of the complete scene, rather than *relative*, that is, in proportion to the size of each individual image region. In the context of this paper both of these scalings have their own benefits. Fixed or absolute scaling of the applicability function most closely matches the physiology of the fovea, but may distort the information contained in either very small or very large image regions. Scaling relative to the region geometry, on the other hand, is more likely to preserve the information from both large and small regions, but the function then becomes less like the human fovea.

If we consider saccades and eye movements, the relative scaling scheme becomes more attractive. This is because the visual focus can roam over a reasonably large region and this seems to simulate a larger, more general, applicability function than just the fovea alone. Thus the applicability function becomes one which is inspired by both the fovea *and* the internal regioning ability of the higher level visual perception in the brain.

4 A granule model and learning regime

In this paper we use a linguistic learning method based on the induction of fuzzy sets over a universe of cross-product (Cartesian) labels ([1],[5]). Each label corresponds to the conjunction of a set of labels which have previously been defined over each individual feature domain. A granule is thus a discrete fuzzy set G over the universe of cross-product labels L :

$$G = \{l \in L : m \in [0, 1]\} \quad (2)$$

where

$$L = \times (K_i \mid i = 1, \dots, n) \quad (3)$$

and K_i is the label universe for feature i .

A cross product fuzzy set can be used to represent a single point in feature space (an instance) or a collection of more than one instance (a model). The learning regime involves generating cross product fuzzy set representations for each training instance and then aggregating these instance representations to generate a model representation. Testing of the model involves calculating the similarity between the model cross product fuzzy set and the cross product fuzzy set representation generated from a single test instance.

In this paper the aggregation operation used to turn training instances G_j into the model G_M is simply the summation operation followed by normalisation of the resulting fuzzy set:

$$G_M = Norm(\sum_j G_j) \quad (4)$$

The similarity measure used in comparing a trained granular model with a cross product representation of a test instance is the sum of products operation on membership values.

4.1 Learning with applicability measures

Where our approach differs from the more conventional cross product fuzzy set methods is the inclusion of the applicability measure. This uncertainty measure affects how the training instance is handled in the training of the granular model. In this paper the instance cross product fuzzy set G_j is first scaled by the applicability measure a_j before it is aggregated with the other (also scaled) instances. In this way an instance with low applicability (i.e. taken from the edge of an image region) has little effect on the final granular model while an instance with high applicability (i.e. taken from the centre of an image region) has greater effect. The training operation thus becomes:

$$G_M = Norm(\sum_j G_j \times a_j) \quad (5)$$

5 Examples

Figure 6 shows example of uncertain labelling and the granular machine learner in operation. The application here is the navigation of an autonomous robot in a new hazardous environment. Although this image shows a safe (office) environment, the intention is to operate the robot remotely and for an operator (domain expert) to train the robot into semi-autonomous operation. This form of remote on-line training is required of rescue robots that cannot be completely trained to deal with unknown hazardous environments in advance.

The image in figure 6a is rapidly labelled by the expert using rough geometric regions into the four classes *body*, *clothes*, *wall*, and *ceiling* as shown in figure 6b. Figure 6c is used in the testing of the model and shows regions corresponding to the four classes. 6d shows the results from testing the learnt model on this image when it was trained using a unity applicability function, 6e shows results after training with a relative-scale Gaussian function with $\sigma = 0.3$, and 6f shows results with an absolute-scale Gaussian with $\sigma = 0.4$ defined over 5% of the image. Percentage values in figure 6 indicate the percentage of pixels correctly classified.

Figure 7 shows results from training the learner to detect emphysema disease in CT scans of the lung. Here the class *healthy* is labelled with black regions and

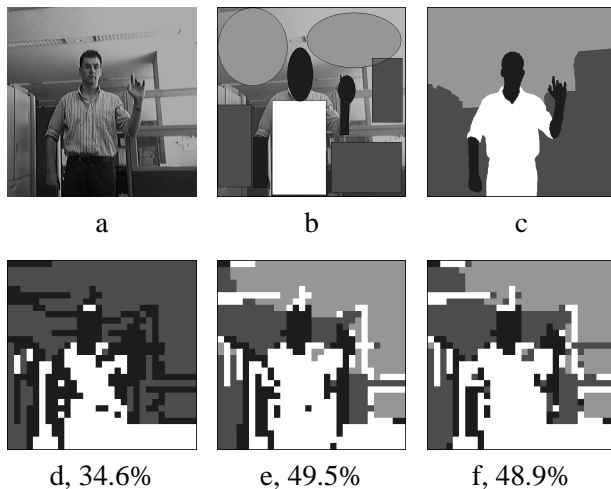


Figure 6: Robot vision example

the class *diseased* is labelled with white regions, as shown in figure 7b. Figure 7c is used in the testing of the model and shows regions corresponding to the two classes. 7d shows results from training using a trapezoid applicability function with $x = 0$, 7e shows results from training using a relative-scale Gaussian with $\sigma = 0.2$, and 7f shows result where the applicability function was unity.

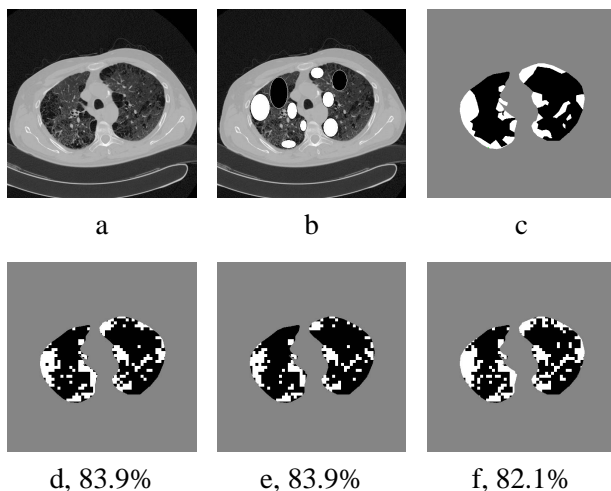


Figure 7: Diagnosing disease from lung scans

It is clear from both the above examples that a graded applicability function improves classification accuracy above a unity function (figures 6d and 7f.) What is less clear is the impact of the type of graded function on results. With these relatively small image sizes (480^2 and 512^2 respectively) and the small number of image features used, both the cone-like trapezoid ($x = 0$) and the low-variance Gaussian ($\sigma = 0.2$) per-

formed equally well. What is clear from our tests is that the relative-scale applicability function is much more convenient to use since we do not need to determine a suitable absolute scale over which to define the applicability function.

6 Conclusion

We have presented a new approach to machine learning from images that encompasses a bio-mimetic method for eliciting uncertain image training data from domain experts and a machine learner that can learn a fuzzy granular model directly from this uncertain training data. We have presented a number of applicability functions for use in the uncertain labelling of images which are based on a simplified model of the human visual system and perceptual process.

Further research will involve optimizing the choice of applicability function and incorporating such a system into tools to aid medical diagnosis and into vision systems for rescue robots operating in hazardous environments.

Acknowledgments

This research is funded by a Japanese Society for the Promotion of Science and Royal Society research fellowship, Ref. P03857

References

- [1] J. F. Baldwin, T. P. Martin, and J. G. Shanahan, "Modelling with words using Cartesian granule features," Proc. of Int. Conf. on Fuzzy Systems, FuzzIEEE 1997, Barcelona, Spain, pp1295-1300, 1997.
- [2] P. Buser and M. Imbert, Vision, MIT press, 1992.
- [3] J. Rossiter and T. Mukai, "The Rapid Elicitation of Knowledge About Images using Fuzzy Information Granules," Proc. Int. Conf. on Fuzzy Systems, FuzzIEEE 2004, Budapest, August 2004.
- [4] S. Samuels, H. L. Kundel, C. F. Nodine, and L. C. Toto, "Mechanism of satisfaction of search : eye position recordings in the reading of chest radiographs," Radiology, 94:895-902, 1995.
- [5] J. G. Shanahan, Soft Computing for Knowledge Discovery: Introducing Cartesian Granule Features, Kluwer Academic, 2000.
- [6] S. Ullman, High-Level Vision: object recognition and visual cognition, MIT Press, 1996.