

# FUZZY SPATIAL DATA MINING

George Brannon Smith

Susan M. Bridges

Department of Computer Science  
Box 9637

Mississippi State University  
Mississippi State, MS 39762

e-mail: brannonsmith@yahoo.com, bridges@cs.mssstate.edu

## Abstract

A *fuzzy spatial data mining* technique has been developed to extract relationships describing relative position of classes of objects from raster images. Several different rule forms are described which represent different types of directional relationships between classes of objects. The method has been tested with hand-generated, synthetic, and sonar imagery.

**Keywords:** data mining, spatial relations, fuzzy data mining, association rules, spatial data mining

## 1.0 Introduction

Rapid advances in computing power, storage capacity, and collection techniques have caused dramatic increases in the rates at which data are being collected. The research areas of data mining and knowledge discovery in databases have developed in response to the need for methods to automatically analyze this data and extract useful new knowledge implicit in the data. Remote sensing technologies have accounted for an explosive growth in the quantity of spatial data collected including satellite images or sonar images. Many data mining techniques that were originally developed for transaction data have been adapted for spatial data mining. Koperski, Han, and Adhikary [5] define spatial data mining as “the extraction of implicit knowledge, spatial relations, or other patterns not explicitly stored in spatial databases.” This definition assumes that the spatial data will be stored in a spatial database. However, much spatial data remains in raster form in flat files. We extend the previous definition to cover all collections of spatial data, not just spatial data bases.

One aspect of data mining is the extraction of relationships among objects that are typical in certain data sets. We describe a method for extracting rules that describe relative position relationships typically found among classes of objects in images. Relative position is a prototypical vague concept. A statement such as “X is north of Y” may be true to a degree and

is conveniently represented using fuzzy logic. Likewise a statement such as “objects of type A are typically north of objects of class B” is also vague and amenable to fuzzy logic representation. We have developed a fuzzy adaptation of association rule mining for mining descriptions of directional relationships between classes of objects in image data sets.

## 2.0 Related Work

In 1993, Agrawal, Imielinski and Swami presented their theory of association rule mining for finding generalizations in a database [1]. The traditional form of such an association rule is the implication

$$\text{antecedent} \Rightarrow \text{consequent} (c\% \text{ confidence}, s\% \text{ support})$$

where the antecedent consists of one or more items in the transaction base being mined and the consequent consists of an item not in the antecedent. Later interpretations allow for logical forms beyond implication. The support metric is the percentage of all transactions in the database containing both the antecedent and the consequent, and it measures the significance of the rule in the database. The confidence is the percentage of those transactions containing the antecedent that also contain the consequent – the strength of the rule. Subsequent research has extended the technique to categorical attributes and quantitative attributes. Kuok, Fu, and Wong were one of the first groups to adapt the techniques to extract fuzzy association rules from quantitative data [6]. Their method allows the items in both the antecedent and the consequent to be fuzzy sets. An example might be:

$$\text{capacity} = \text{high} \Rightarrow \text{margin} = \text{low} \\ c=90\%, s=21\%$$

Spatial data mining presents additional challenges not encountered in transaction data mining. Shekhar et al. [10] state that “extracting interesting and useful patterns from spatial datasets is more difficult than extracting corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial

autocorrelation.” There is a large body of research in the area of spatial data mining. The previous research efforts most relevant to ours deal with extraction of rules containing spatial predicates from spatial data and with generation of linguistic descriptions of relative position.

In the research area of rule extraction from spatial data, Ordonez and Omiecinski [9] transform image data into traditional Boolean transactions, thus enabling the application of retail-style association rule-mining. They focus on simple object co-occurrence and do not deal with other spatial relationships. Koperski and Han [5] describe methods for spatial association rule discovery from spatial databases. They define a spatial association rule as association which contains at least one spatial predicate. Their rules have the restriction of always using a reference feature. The Co-location Miner of Shekhar et al. [10] mines co-location relationships from Boolean spatial features without the use of a reference feature. Our research differs from previous work because rather than using Boolean spatial features as a starting point for the mining process, we use fuzzy relations—in particular fuzzy relations that describe the relative position of objects.

There is a long history of research in the area of generating descriptions of relative position of objects in images beginning with the seminal work of Winston in 1975[12]. Many families of fuzzy directional relations have been defined and methods for computing these developed. For example, a number of these methods depend on the computation of a histogram of angles [8]. Matsakis, et al. [7]define a family of directional relations that rely on the computation of a histogram of forces. Bloch [2] combines a morphological approach with fuzzy pattern matching to compute the fuzzy relative positions of two objects—the computation is done directly in image space. We have used Bloch’s method of computing fuzzy relative position in our work.

### 3.0 Mining Fuzzy Spatial Association Rules

The problem we are addressing is the following:

Given a segmented, classified raster image, generate rules that describe commonly found directional relationships that exist between classes of objects.

Because of the inherent imprecision in directional relationships, due in part to morphology, the resulting rules will be fuzzy.

We have combined the fuzzy association rule approach of Kuok, Fu, and Wong [6] with the

methods of Bloch for finding fuzzy spatial relationships [2] to develop a method for extracting fuzzy spatial rules from an image. We restrict our attention to raster data, operating in image space as opposed to object space. The use of Bloch’s approach allows implicit consideration of spatial object morphology. When used in conjunction with techniques for fuzzy association rule mining, information can be mined that describes the certainty of the extracted rules in terms of a fuzzy membership interval. In the following sections we describe the fuzzy relations that are used as the starting point for our mining process, the fuzzy rule forms that we have developed, and the mining algorithm.

### 3.1 Fuzzy Spatial Relations

We apply Bloch’s [2] fuzzy spatial techniques for 2D space applied to segmented and classified image data to produce fuzzy spatial relations between pairs of objects. Given an image space  $S$  containing reference object  $R$  and another object  $A$ , and a direction  $\alpha$ , the evaluation of the relative position of  $A$  with respect to  $R$  is given by a function of  $\mu_{\alpha}R(x)$  and  $\mu_A(x)$  for all  $x$  in  $S$ . The possibility distribution  $\mu_{\alpha}R(x)$  defines the “fuzzy landscape” of  $R$  in direction  $\alpha$ . Figure 1 illustrates the fuzzy landscape of an object.



Figure 1. Fuzzy landscape of an object in direction 0 radians

The possibility distribution  $\mu_A(x)$  defines the membership of each pixel in an object A for all  $x$  in  $S$ . In our current work, we have used only crisp objects, but the concepts extend naturally to accommodate fuzzy objects. Fuzzy pattern matching between the two possibility distributions  $\mu_{\alpha}R(x)$  and  $\mu_A(x)$  results in three numbers, a necessity degree  $N$ , a possibility

degree  $I$  and an average measure  $M$ . These three measures are defined by Bloch as follows:

$$\prod_{\alpha_1 \alpha_2}^R(A) = \sup_{x \in S} t[\mu_{\alpha}(R)(x), \mu_A(x)]$$

$$N_{\alpha_1 \alpha_2}^R(A) = \inf_{x \in S} s[\mu_{\alpha}(R)(x), 1 - \mu_A(x)]$$

$$M_{\alpha_1 \alpha_2}^R(A) = \frac{1}{|A|} \sum_{x \in S} \mu_A(x) \mu_{\alpha}(R)(x)$$

where  $t[]$  is a fuzzy t-norm,  $s[]$  is a fuzzy t-conorm, supremum is the least upper bound, and infimum is the greatest lower bound. The *possibility* is an optimistic result, *necessity* is a pessimistic result, and *mean* is the average membership degree over all the points. Together they express the value and interval

$$M \in [N, I].$$

We use Bloch's definitions of fuzzy relative position to define binary relations between objects of different classes. Given the classes of all objects in an image, we can form descriptive relations  $D_{\alpha}$  for pairs of objects as shown in Figure 2. This yields a transaction database of  $(|O| \cdot (|O| - 1) \cdot |D|)$  tuples for each image. These fuzzy relations are used to form fuzzy spatial association rules.

### 3.2 Fuzzy Spatial Association Rules

We have developed a number of rule forms that can express different types of relative position associations between objects of different classes. One variation aggregates all objects of each class into large "pseudo-objects" and mines associations between these pseudo-objects. This approach supports a very efficient mining algorithm, but in general, we have found that it yields less information than rule forms that treat objects individually. We describe two of these rule forms in the following section.

#### 3.2.1 Universally Quantified Rule Form

The first rule form describes situations in which all objects in one class tend to be located in the same direction relative to all objects in a second class.

Rule Form 1:

$$\forall x \forall y \quad class(a) \wedge object(x) \wedge inClass(x, A) \\ \wedge class(B) \wedge object(y) \wedge inClass(y, B) \\ \Rightarrow inDirection(\alpha, x, y)(S, C)$$

or:

Given any class A object and any class B object, it follows that the class B object will be located in direction  $\alpha$  of the class A object with a certain confidence and support in the database. We generally restrict  $\alpha$  to the values  $(0, \pi/2, \pi, \text{ and } 3\pi/2)$  radians.

We have investigated different methods for computing the support and confidence. One method for computing support is in terms of pixels:

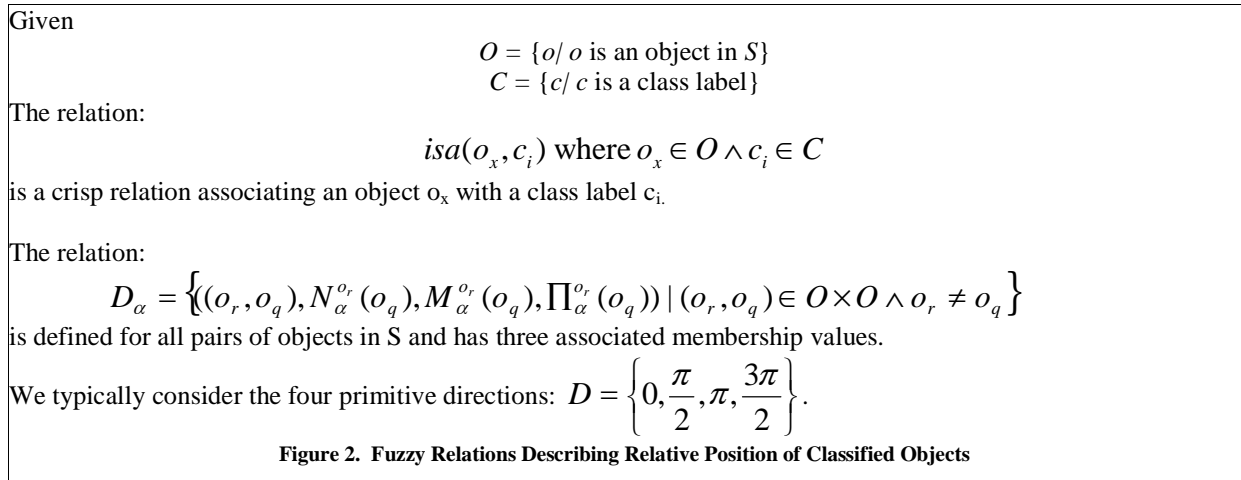
$$S_p = Sum(NP_A, NP_B) / NP_{Obj}$$

where  $NP_A$  and  $NP_B$  are the number of pixels in objects of classes A and B respectively and  $NP_{Obj}$  is the number of pixels in all objects. A second method computes support in terms of objects:

$$S_o = Sum(NO_A, NO_B) / NO$$

The first method seems most applicable when objects differ significantly in morphology. The second method works well for those cases in which the number of objects for which the relationship holds is more important than the size of those objects.

We have adapted Bloch's method for computing necessity, mean, and possibility of directional relationships between individual objects to the computation of similar metrics for association rule confidence:



$$C_i = \frac{\sum_{x \in A} \sum_{y \in B} \text{inDir}_i(\alpha, x, y)}{NO_A + NO_B}$$

where  $i$  corresponds to either N, M, or  $\Pi$  as computed by Bloch and  $NO_A$  and  $NO_B$  are the number of objects in class A and B respectively.

### 3.2.2 Existentially Quantified Rule Form

A second rule form describes the situation in which objects of one class tend to be located in the same relative direction of at least one object of a second class.

Rule Form 2:

$$\forall x \exists y \quad \text{class}(a) \wedge \text{object}(x) \wedge \text{inClass}(x, A) \\ \wedge \text{class}(B) \wedge \text{object}(y) \wedge \text{inClass}(y, B) \\ \wedge \text{inDirection}(\alpha, x, y)(S, C)$$

or:

Given any class A object, we assert that there exists some class B object located in direction  $\alpha$  of the class A object with a certain confidence and support in the database.

Pixel and object support values are computed as for the universally quantified rule form. Confidence is computed as follows:

$$C_i = \frac{\sum_{x \in A} \max_{y \in B} (\text{inDir}_i(\alpha, x, y))}{NO_A}$$

As before, three different confidence values based on necessity, possibility, and mean are computed for each rule. The obvious difference in the existential rule form and the universal form is that the existential rule form is less demanding. The other difference is the absence of the traditional implication in favor of a simple conjunction.

### 3.3 Mining Algorithm

In general, the rules generated with the forms described above are similar to those discussed by Kuok, Fu, and Wong[6]. Their work establishes a convenient framework for interpreting directional attributes in which they describe a set of fuzzy sets

$$F_{i_k} = \{f_{i_k}^1, f_{i_k}^2, f_{i_k}^3, f_{i_k}^4\}$$

that are associated with an attribute  $i_k$  in the set of attributes  $I$ . In our case, the set  $D$  of primitive directions represents a partitioning of the attribute direction ( $D_\alpha$ ) into such fuzzy sets

$$F_{D_\alpha} = \{f_{D_\alpha}^1, f_{D_\alpha}^2, f_{D_\alpha}^3, f_{D_\alpha}^4\} \\ = \{\text{right}, \text{above}, \text{left}, \text{below}\}$$

The direction field of an ordered pair of objects has some degree of membership in each of these four fuzzy sets.

Bloch describes both an exact and an approximation algorithm for generating the fuzzy landscapes used by her method. The approximation method is much more efficient and gives results that are typically visually indistinguishable from the exact algorithm. We use the approximation algorithm for our mining system.

The rule mining algorithm follows directly from the computation of support and confidence for each rule form. Given a set of tuples and the template rule forms that we wish to derive, the support and confidence for each rule are computed and if these are above thresholds that have been established, the rules are retained. This exhaustive approach, although computationally expensive, provides acceptable performance with the highly restricted rule forms we are currently using. When the method is extended to accommodate other spatial predicates and rule forms, it will be necessary to use develop a more sophisticated mining algorithm with high-level pruning capabilities.

## 4.0 Experiments and Results

We have tested the spatial mining algorithms described above on hand-constructed images, randomly generated images, and partitioned, classified sonar images.

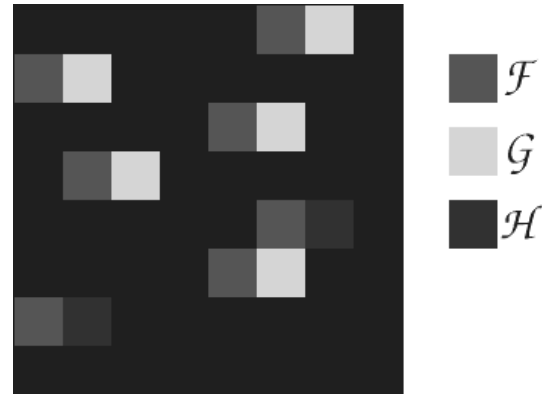


Figure 3. Hand crafted image with 3 classes

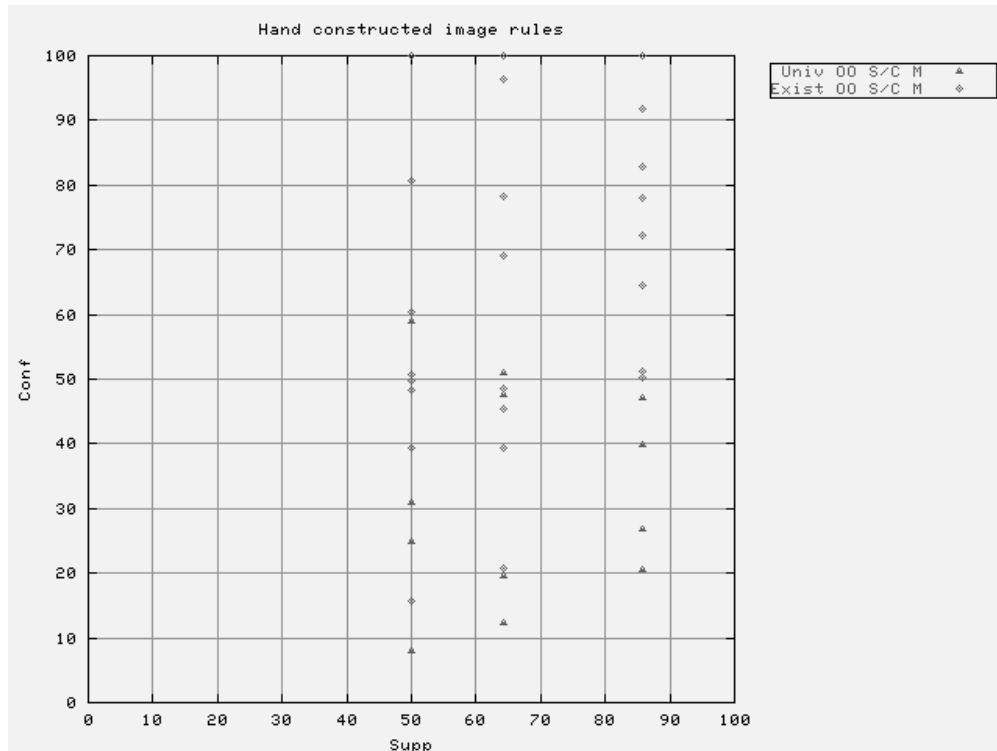


Figure 4. Scatter Plot of Rule Metrics for Figure 3

In Figure 3 we present a simple image with spatial relationships that are apparent to the human eye. We expect to derive strong rules in the left and right directions. Figure 4 shows a scatter plot for the rules generated where the x-axis is support and the y-axis is confidence. High support and confidence rules are in the upper right. We plot only  $C_M$  values. As expected the universal rules exhibit lower support and confidence in general than the existential rules. The highest scoring rule states that for every object in class  $G$ , there is an object in class  $F$  to the left with support 85.7 and mean confidence 100.0.

We have also developed a number of biasing techniques to generate synthetic images with relationships among classes of objects that were more probable in certain directions. The task of generating synthetic images turned out to be very challenging. The mining algorithm typically discovered the associations that were “planted” in the images, but these were often not the strongest resulting associations. Visual examination of the high scoring rules and the images from which they were derived revealed that the associations that were mined were meaningful. See [11] for more detail.

For a real world test we ran the system on some classified seafloor sonar data from the OKEANOS

project [ ], the spatial arrangement of which was unknown. One of these images is shown in Figure 5.

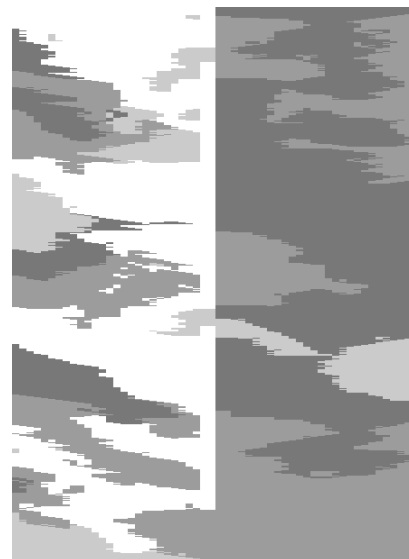
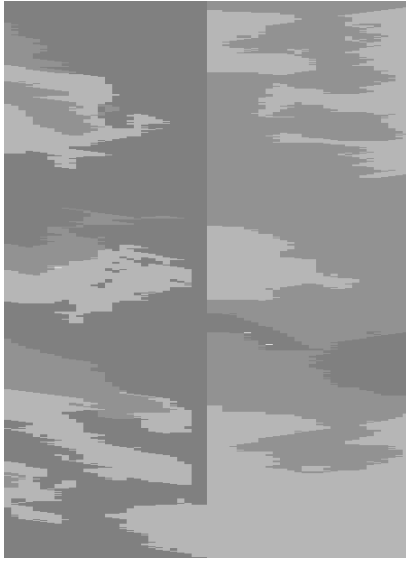


Figure 5. Segmented seafloor image



**Figure 6** Objects from Figure 5 with High Scoring Rules

In Figure 6 the objects in the two light shades of gray are those involved in the highest scoring rule derived. This rule states that for every object of class K (lightest gray), there is an object of class L (medium gray) in the direction  $3\pi/2$  with support 63.4 and confidence in the interval [94.5, 100]. This rule was judged to be both valid and interesting.

### 5.0 Summary and Conclusion

We have developed a method for mining fuzzy spatial association rules that express typical relative position relationships for classes of objects. We have developed several different rule forms involving both existential and universal quantifiers to capture these rules. The method of Bloch for describing fuzzy relative position has been combined with methods for fuzzy association rule mining. Experimental results demonstrate that we generate interesting, often non-obvious rules using these methods.

Although we have restricted our attention to crisp objects, this approach is also easily extended to fuzzy objects. Likewise, the method could easily be modified to use other families of relations for relative position other than that defined by Bloch and additional spatial predicates.

### Acknowledgement

This work was partially funded by a grant from the Naval Oceanographic Office NAS1398033 DO92.

### 6.0 References

- [1] Agrawal, R., T. Imielinski, and A. Swami. 1993. Mining associations between sets of items in massive databases. *Proc. 1993 SIGMOD*

- International Conf. on Management of Data, Washington DC, May 26-28, 1993, NY, pp. 207-216*
- [2] Bloch, Isabelle. 1999. Fuzzy relative position between objects in image processing: A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 657-664.
- [3] Bridges, S. M., J. Hodges, B. Wooley, D. Karpovich, and G. B. Smith, "Knowledge discovery in an oceanographic database," *Applied Intelligence*, No. 11, pp. 135-48, 1998.
- [4] Keller, J. M. and X. Wang, "Comparison of Spatial Relation Definitions in Computer Vision," *Proc. ISUMA-NAFIPS '95*, pp. 679-684.
- [5] Koperski, K. and J. Han. 1995. "Discovery of spatial association rules in geographic information databases." In *Adv. in Spatial Databases: Proc. of the 4<sup>th</sup> International Symp. on Large Spatial Databases*, Portland, Maine, August 6-9, 1995. pp. 47-66, Springer.
- [6] Kuok, C., A. Fu, and M. Wong. 1998. Mining fuzzy association rules in databases. *SIGMOD Record* 27(1): 41-6.
- [7] Matsakis, P., J. M. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera. "Linguistic Description of Relative Positions in Images," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 31(4), pp. 573-588, 2001.
- [8] Miyajima, K. and A. Ralescu, "Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations," *Fuzzy Sets and Systems*, vol. 65, pp. 225-236, 1994.
- [9] Ordonez, C. and E. Omiecinski, "Discovery association rules based on image content," in *Proceedings of the 1999 IEEE Forum on Research and Technology Advances in Digital Libraries*, Baltimore, MD, May 19-21, 1999, pp. 38-49.
- [10] Shekhar, S., Y. Huang, W. Wu, C. T. Lu, and S. Chawla. "What's Spatial About Spatial Data Mining: Three Case Studies," in *Data Mining for Scientific Engineering Applications*, V. Kumar, R. Grossman, C. Kamath, K. Nambaru, Eds. Kluwer Academic, 2001.
- [11] Smith, George Brannon, "Mining Fuzzy Spatial Association Rules from Data," Masters Thesis, Mississippi State University, August 2001.
- [12] Winston, P. "Learning Structural Descriptions from Examples," in *The Psychology of Computer Vision*, P. Winston, Ed. New York: McGraw-Hill, 1975.

