Article

Genetic Optimization for Associative Semantic Ranking Models of Satellite Images by Land Cover

Adrian Barb * and Nil Kilicay-Ergin

Great Valley School of Professional Studies, Penn State University, Malvern, PA 19355, USA;
E-Mail: nhe2@psu.edu

* Author to whom correspondence should be addressed; E-Mail: adrian@psu.edu;
Tel.: +1-610-725-5349; Fax: +1-610-648-3377.

Abstract: Associative methods for content-based image ranking by semantics are attractive due to the similarity of generated models to human models of understanding. Although they tend to return results that are better understood by image analysts, the induction of these models is difficult to build due to factors that affect training complexity, such as coexistence of visual patterns in same images, over-fitting or under-fitting and semantic representation differences among image analysts. This article proposes a methodology to reduce the complexity of ranking satellite images for associative methods. Our approach employs genetic operations to provide faster and more accurate models for ranking by semantic using low level features. The added accuracy is provided by a reduction in the likelihood to reach local minima or to overfit. The experiments show that, using genetic optimization, associative methods perform better or at similar levels as state-of-the-art ensemble methods for ranking. The mean average precision (MAP) of ranking by semantic was improved by 14% over similar associative methods that use other optimization techniques while maintaining smaller size for each semantic model.

1. Introduction

Evaluation of geospatial imagery is challenging due to high dimensionality of spatial data and to the coexistence of visual patterns related to multiple semantics in images [1]. As the rate of image collection grows exponentially, it is becoming exceedingly difficult for image analysts to manually extract knowledge from geospatial images in order to deliver focused information for decision making. This necessitates the need for automating remote sensing data analysis and evaluation. Traditional data approaches, such as statistical methods, have limitations in terms of distributional assumptions and restrictions on data input which may prevent them from analyzing unknown and unexpected relationships in geospatial images [2]. Other traditional methods of data mining such as Artificial Neural Networks and Genetic Algorithms (GA) have a black-box characteristic which makes it difficult for users to apply extracted rules to other cases [3]. Besides, data values gain meaning only in the context of the geospatial domain and the existence of multiple semantic interpretations for the same image [4,5], which makes it difficult to apply traditional data analysis methods to images. Therefore, new approaches that consider unique characteristics of image data have emerged for mining patterns from images.

In content-based image retrieval, images are indexed by their visual contents such as color and shapes. However, these low-level features cannot properly capture the high-level image semantics in a user's mind. Therefore, recent studies on content-based image retrieval focus on reducing the semantic gap between low-level features and high-level human semantics by constructing semantic models that can be used for prediction. A comprehensive review of various semantic models are provided in [6] where methods for reducing the semantic gap include using object ontology to define concepts, using machine learning methods to associate low-level features to users' semantics, introducing relevance feedback to learn users' intentions, generating a semantic template to map low-level features to high-level concepts, and combining visual and text content for web image retrieval.

Recent research in the geospatial area provided a variety of in-depth solutions [7–18], to represent the complex, often overlapping geospatial knowledge and to assist image analysts in generating necessary domain specific metadata. The research in [7] describes a framework for modeling and image retrieval using directional spatial relationships among objects. Content-based image retrieval (CBIR) methods were applied to ranking satellite images using possibilistic associations between low-level features and semantics of interest [8]. The researchers in [9,10,13] use Latent Dirichlet Allocation (LDA) semi-supervised methods to annotate images with semantic classes. Both supervised and unsupervised methods are combined in the I^{3}KR [11] framework to enhance image searching capabilities using semantic- and content-based information. The researches in [12,15] efficiently retrieve images using indexing structures on the feature space. The application of self-organizing maps to the analysis of man-made structures in multispectral imagery is investigated in [14]. The research in [16] proposes the integration of a multi-modal content-based system with complex methods of querying on shape, multi-object relationships, and semantics, while the research in [17] automatically detects variations in geospatial images and applies clustering techniques to organize visual pattern variations. The approach in [18] uses ontological knowledge and artificial neural networks to build semantic models of visual patterns using both low-level and descriptive image features. These models can be used to measure the

semantic similarity among image objects. For an in-depth review of spatial data mining and knowledge discovery, the reader is directed to [3].

Among the proposed solutions, associations between low-level features and visual patterns are generated using data mining techniques [8,19] and provide more human-readable insight into the structure of the generated models. Each association rule generates a decision rule where a set of low-level features are selected as the antecedent and a unique semantic as the consequent of the decision rule. The association rules are then evaluated and ranked for their relevance to the high-level visual patterns. Different algorithms have been proposed for spatial association rule mining. Among those, Apriori and AprioriTid algorithms [20] have made significant improvements for generating efficient rules and filtering rules that are trivial or common knowledge. One of the challenges in this area is the computational overhead associated with various spatial predicates in order to derive association rules from large data sets. An approach that derives association rules using fuzzy data mining techniques is proposed in [21] to deal with uncertainty found in spatial data. In [22], self-organization maps are used to mine whether satellite images, and then time dependent association rules, are extracted using Apriori algorithm. For an in-depth review of associative classification mining and spatial associative rule mining, the reader is directed to [23].

The method of feature selection from raw original images is an important step in improving the performance of associative rule mining methods. The process reduces the dimensionality and complexity of the raw image data by eliminating irrelevant and redundant features. A similarity/dissimilarity measure between the selected set of low-level features and high-level semantics determines the effectiveness of the associative models. An important problem in geospatial knowledge discovery is the choice of optimization strategies that can be applied to a feature space. Finding a unique solution in a high-dimensional feature space that contains a large quantity of continuous variables is a challenging task. In particular, in spatial associative mining, subspace generation is exponential to the number of possible subspaces which makes brute force associative methods NP-hard [24]. Feature selection algorithms attempt to reduce feature space complexity by removing irrelevant features [25] using either filtering or wrapper approaches. Brute-force feature selection algorithms are also computationally expensive, while recently proposed feature selection algorithms are greedy in nature and may return inferior performance. Other greedy decision algorithms [26,27] attempt to reduce the complexity of the problem but may be trapped in suboptimal, local maximum solutions. To overcome this problem, additive associative models are used where the newly discovered association rule is added to the model only if the rule's relevance to the semantic model is greater than a predefined threshold [9]. For example in [8,28] additive models were combined with algorithms such as the Sequential Forward Floating Selection Algorithm (SFFS), which applies a number of backward steps as long as the objective function returns better results. Feature selection through association rules is also employed in [29] to reduce the dimensionality of feature vectors.

Evolutionary algorithms are self-adaptive optimization methods that perform global search in a solution space. They tend to perform better with attribute interactions when compared to greedy decision algorithms [30]. Genetic Algorithms (GAs) [31,32] model the space of candidate solutions in chromosome structure where the success of each chromosome is assessed with a fitness function. The best solution or most satisfactory solution is based on natural selection methods that combine successful features existent in a set of previously generated models by selection, crossover and mutation. Since

knowledge about the search space is accumulated during the search process, GAs can eliminate local-maxima traps by adaptively moving the solution space to approach a global optimal. GAs are applied in various spatial data mining domains. In [33], evolutionary programming is used to classify multispectral images using a non-linear combination of spectral and texture metrics. The research in [34] uses GAs to optimize the interpolation of air pollution data while the research in [2] applies to GAs to classify land-cover using object shape found in image. In [35], a spatial clustering method based on GAs and k-medoids is proposed to address spatial clustering with obstacle constraints. The research in [36] uses GAs to discover association rules for image data mining. In [37], a multi-objective optimization algorithm is used to search a number of conflicting objective functions to find Pareto-optimal solution for pixel classification.

GAs have also been applied for feature selection in image retrieval tasks. In [38], GA-based feature selection algorithm is used to select a set of discriminative feature set for satellite images. Separability index is used as the fitness function to evaluate feature subsets and the effectiveness of the algorithm is tested on a neural network classifier. In [39], ranking evaluation functions are proposed as fitness functions in GA-based feature selection to search for the best feature set. In [40], the feature selection method includes a filter-based feature selection using genetic algorithm to improve the precipitation estimation from a remotely sensed imagery.

In this paper, we extend the work in [41] to explore steady-state genetic methods [42] for optimization of associative models for ranking geospatial image regions by land cover. Our goal is to provide an associative method for mapping semantics to visual patterns in domain-specific images. These methods are attractive due to the fact that they can be interpreted much more easily by experts, which can be eventually used in expert training procedures. In previous approaches, we have used Apriori association rule mining techniques for the initial determination of the feature subspaces. However, the training proved to be complex and many of the methods used to reduce the complexity proved limiting and directly affected the quality of pranking. Therefore, in the new approach we use only genetic methods for generation, selection and fine-tuning of the mappings between feature spaces and semantics. The main scope of this article is to evaluate if genetic methods for associative rule mining resulted in performance that is better or similar to the performance of other state-of-the-art techniques. We investigated two models of genetic algorithm for offspring generation; generational GA (standard GA) and steady state GA and chose to use the later. In standard GA, the genetic operators replace the entire old generation with the new off-spring population, whereas in steady-state GA, the population is replaced incrementally such that there is one new member inserted into the new population. A replacement strategy determines which members of the population will be replaced by the new offspring [43]. Each association between a feature and the land cover of interest is modeled as a k-bit exon that contains information about both the features and the characteristics of the feature subspace used. The novelty of our approach is the use of genetic operations at both feature and subspace levels. We evaluate the fitness of models in genetic populations using MAP and compare and contrast it with the SFFS optimization algorithm used in [8]. This paper is organized as follows: In Section 2 we introduce the methodology used to implement genetic algorithms, we present the experimental results in Section 3, and then conclude the article in Section 4.

2. Methodology

In this section we present our methodology for ranking satellite image regions using genetic operations. For each image in the database we generate a feature space *F*. The key feature of the algorithm is that we use sets of association rules between feature subspaces and semantics in a semantic space *S* to rank images by semantic. Each set of associations is generated and evolved using genetic operations at two levels: the feature and subspace levels. At the feature level, we vary the set of features used to identify association rules, while at the subspace level we vary the region for the same feature set that will be used in ranking. For example, for a 38-dimensional space there are 2^{38} unique combinations of features. Using genetic operations we randomly choose and evolve combinations of features using methods such as crossover, shrink, constant, or grow mutations. Once a combination of features is selected, we randomly generate and evolve features' subspaces modeled by sigmoid possibilistic functions. Further, sets of feature spaces are used additively to model correlation to a semantic of interest. To evaluate which subspace is the most relevant we also apply genetic operations at this level.

2.1. Fitness Function

The fitness function for each semantic model is used by the optimization algorithm to determine which combinations of association rules will better model the association between feature subspaces and semantics of interest. In our study, we use the MAP to determine the relevance of each feature subspace, a set of associations that will form a semantic model. However, since each semantic model is an ensemble of associations, with multiple non-zero relevance values, the fitness function is applied as follows: Each association rule maps the region of the feature space $\Theta \subset F$ into the semantic of interest $\varsigma \in S$.

$$\Theta = \{g(m_\varphi)\} \tag{1}$$

$$g(m_\varphi) = \frac{w_g}{2}\left[\tanh\left(\frac{m_\varphi - \lambda_1^L}{\lambda_2^L}\right) - \tanh\left(\frac{m_\varphi - \lambda_1^R}{\lambda_2^R}\right)\right] \tag{2}$$

The function *g* is an asymmetric double sigmoid possibilistic distribution (*L*—left and *R*—right) that models the relevance of a measurement m_φ to a semantic ς. Each half sigmoid is controlled by two parameters: (a) center $(\lambda_1^L, \lambda_1^R)$ and (b) width $(\lambda_2^L, \lambda_2^R)$ while w_g is weight of the relevance retrieved by the *g*. Each possibility distribution is shaped using the relevance assessments provided by image analysts, which we considered as ground-truth semantic information for each semantic of interest. For details of this mapping function, the reader is referred to [8]. The relevance of an image ι to a semantic ς is determined by the relevance of the feature values $m_\varphi(\iota)$ of the image over region of the feature space Θ:

$$T(\iota | \Theta, \varsigma) = w_\Theta \cdot \min\left(g\left(m_\varphi(\iota)\right)\right) \Big| g\left(m_\varphi(\iota)\right) \in \Theta \tag{3}$$

where w_Θ is a weight of the feature subspace Θ that determines its relevance in mapping *F* into ς. Further, for each semantic ς we create a semantic model M_ς defined as the set of mappings of subspaces Θ of *F* into a semantic space *S*:

$$M_\varsigma = \{\Theta \to \varsigma\} \tag{4}$$

The overall relevance $T(\iota|M_\varsigma)$ of an image ι with feature measure $m_\varphi(\iota)$, to a semantic ς is computed by sorting the relevance (rank function) values of image feature measures to each feature subspace $\Theta \in M_\varsigma$ in descending order and then computing:

$$T(\iota|M_\varsigma) = \sum_{|M_\varsigma|} T(\iota|\Theta,\varsigma)/2^{\mathrm{rank}(T(\iota|\Theta,\varsigma))} \tag{5}$$

In this equation, $T(\iota|M_\varsigma)$ is computed as a weighted mean of all the sigmoid relevance values of the associations in the semantic model. We have chosen this average because we want to emphasize the most relevant association while deemphasize the less relevant association that have only a marginal effect.

Finally, for each of the experiments we then compute the fitness function as the MAP of ranking, which provides an aggregate measure of precision (how many of the images retrieved in a search by semantic are actually relevant) across all the recall levels for each model M_ς for $\varsigma \in S$ over a feature space F. The MAP measure is shown below:

$$MAP_\varsigma = \frac{1}{|\iota^\varsigma|}\sum_{j=1}^{\iota^\varsigma}\left(\frac{1}{j}\sum_{k=1}^{j}\left(\mathrm{Precision}(R_{jk})\right)\right) \tag{6}$$

In this formula, R_{jk} is the set of ranked j images from the top to the kth image.

2.2. Encoding

Each generated membership function is considered an exon ε and it is encoded as a decimal string for the sequence $(\varphi, \lambda_1^L, \lambda_2^L, \lambda_1^R, \lambda_2^R)$ using a total 20 decimal digits. The feature φ is recorded as the index of the feature in the feature space using four decimal digits, while for each of the sigmoid parameters we store the most *significant* four digits after the decimal point that resulted after the process of normalization. For readability of the article, we will break a genetic sequence in smaller parts as well as highlight each group of four digits by alternating between italicized and bolded text. For example, $g(m_{F1}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.01}{0.05}\right) - \tanh\left(\frac{x-0.624}{0.01}\right)\right]$ over a feature *F*1 will be encoded as ε = *0001* **0100** *0500* **6240** *0100*.

A gene ζ is a set of conjunctive exons and it is encoded by the sequence $(\eta, rel(\zeta), \varepsilon_1, \dots \varepsilon_n)$ in which η is the number of exons in the gene and $\Theta(\zeta) \in [-1,1]$ represents the relevance of the full membership allowed by the gene. For example, consider a gene having $\Theta(\zeta) = 0.721$ and containing two exons on a two-dimensional feature space {*F*1, *F*2}. Each exon is equivalent to the following sigmoid functions: $g_1(m_{F1}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.887}{0.015}\right) - \tanh\left(\frac{x-0.998}{0.001}\right)\right]$ and $g_2(m_{F2}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.01}{0.05}\right) - \tanh\left(\frac{x-0.624}{0.01}\right)\right]$. This gene is encoded ζ = *0002* **7210** *0001* **8870** *0150* **9980** *0010* **0002** *0100* **0500** *6240* **0100**. For this gene, each point in the feature subspace $F1 \in [0.887, 0.998] \wedge F2 \in [0.01, 0.624]$ has a relevance of $\Theta = 0.721$ while feature points outside this area will have smaller relevance as dictated by the sigmoid functions.

A chromosome χ is a set of disjunctive genes that can be aggregated using the $T(\iota|M_\varsigma)$ function and it is encoded as a concatenation of the constituent genes $\chi = (\zeta_1, \dots, \zeta_k)$. For example, consider that we have a chromosome with two genes: ζ_1 = *0002* **7210** *0001* **8870** *0150* **9980** *0010* **0002** *0100* **0500** *6240*

0100, which was described in the previous paragraph and $\zeta_2 =$ *0001* **2018** *0001* **6670** *8810* **0121** *0040*, that contains one exon equivalent to $g_3(m_{F1}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.667}{0.0121}\right) - \tanh\left(\frac{x-0.881}{0.004}\right)\right]$ over feature *F*1 and with $\Theta(\zeta_2) = 0.2018$. This chromosome is encoded $\chi =$ *0002* **7210** *0001* **8870** *0150* **9980** *0010* **0002** *0100* **0500** *6240* **0100** *0001* **2018** *0001* **6670** *8810* **0121** *0040*. Each chromosome represents a customized region of the feature space. The purpose of our methodology is to identify the optimal region that can maximize the quality of ranking for a semantic. This set of associations will constitute a semantic model for that semantic and will be used for ranking new, unlabeled images that are added to the database.

Finally, a population is a set of chromosomes ($\chi_1,\ldots,\chi_n$) that compete to explain the association between a feature space and a semantic, while a genetic material is a set of chromosomes that return the highest performance in modeling all the semantics of interest.

2.3. Genetic Operations

We perform genetic operations at three levels: exon, gene, and chromosome. Below we enumerate the genetic operations that are performed on each population which are exemplified in Figure 1 on a simplified two-dimensional feature space composed of object convex area kurtosis (*F*1) and orientation skewness (*F*2). In this figure, the vertical axis is the relevance feature points to a semantic of interest.

Figure 1. Example of a sequence of genetic operations: (**a**) random; (**b**) exon shift lambda 1; (**c**) exon shift lambda 2; (**d**) gene growth mutation; (**e**) gene relevance mutation; (**f**) gene constant mutation; (**g**) gene cross by replacing the second exon in (e) with the first exon in (a); (**h**) gene shrink mutation; (**i**) chromosome growth mutation; (**j**) chromosome constant mutation; (**k**) chromosome cross over with first gene at (d); (**l**) chromosome shrink mutation.

(**a**) (**b**) (**c**)

Figure 1. *Cont.*

(**d**)   (**e**)   (**f**)

(**g**)   (**h**)   (**i**)

(**j**)   (**k**)   (**l**)
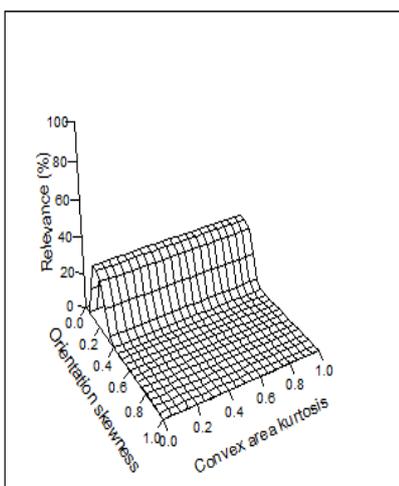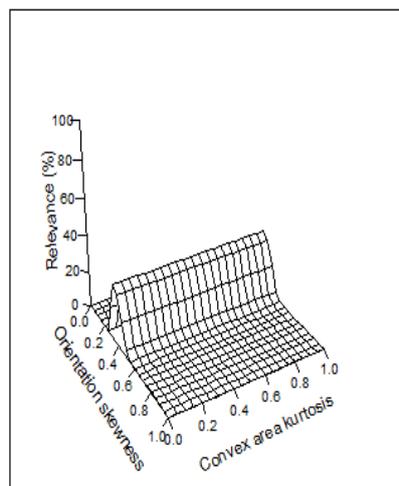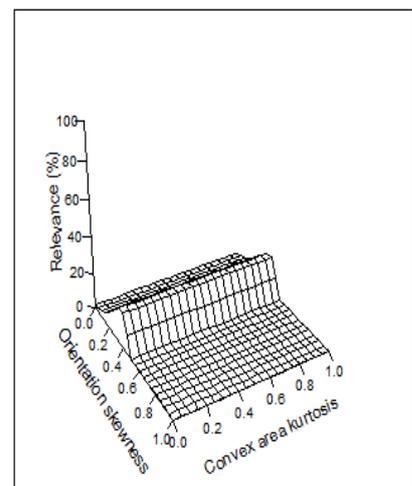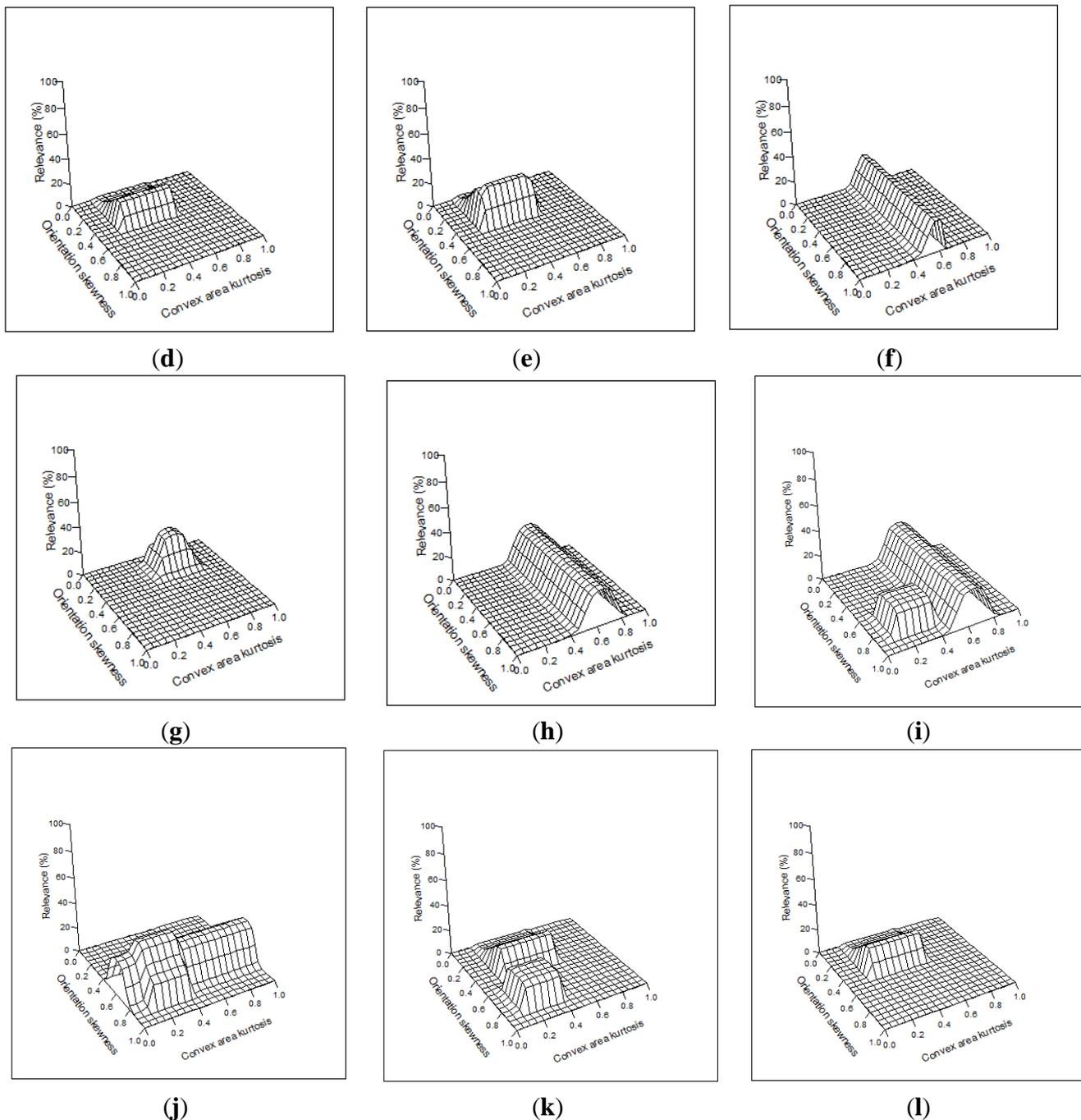
Chromosome Random generation: The first population uses completely random generation of chromosomes. The number of genes in each chromosome is randomly chosen between three and twelve, while each gene has at most five exons. The range of genes in an exon was empirically shown by our experiments to be returned by the associative model while we want to maintain the number of exons in a model to preserve the white-box nature of our semantic models. Figure 1(a) shows relevance of the feature space when using a randomly generated chromosome with one gene, one exon on the *F*2, and with the code *0001* **6510** *0002* **1012** *3410* **0200** *0513*. This is equivalent to a sigmoid function $g(m_{F2}) = \frac{1}{2}\left[\tanh\left(\frac{x - 0.1012}{0.02}\right) - \tanh\left(\frac{x - 0.3410}{0.0513}\right)\right]$.

Exon Shift of λ₁ Parameter: This operation adds variation to the feature interval of maximum relevance by randomly changing λ_1^L and λ_1^R with up to ±5%. Figure 1(b) shows relevance of the feature space when genetically transforming *0001* **6510** *0002* **1012** *3410* **0200** *0513* into *0001* **6510** *000* *23763* *4586* **0200** *0513*. This is equivalent to a new sigmoid function $g(m_{F2})=\frac{1}{2}\left[\tanh\left(\frac{x-0.3763}{0.02}\right)-\tanh\left(\frac{x-0.4586}{0.0513}\right)\right]$ with variation in λ_1^L and λ_1^R over the previous generation.

Exon Shift of λ₂ Parameter: This operation adds variation to the feature interval of maximum relevance by randomly changing λ_2^L and λ_2^R with up to ±5%. Figure 1(c) shows relevance of the feature space when genetically transforming *0001* **6510** *0002* **1012** *3410* **0200** *0513* into *0001* **6510** *0002* **3763** *4586* **2130** *0500*. This is equivalent to a new sigmoid function $g(m_{F2})=\frac{1}{2}\left[\tanh\left(\frac{x-0.3763}{0.213}\right)-\tanh\left(\frac{x-0.4586}{0.05}\right)\right]$ with variation in λ_2^L and λ_2^R over previous generation.

Gene Grow Mutation: This operation adds a new exon to a randomly selected gene in the chromosome. Figure 1(d) shows relevance of the feature space when adding the exon with the code *0001* **2001** *6011* **0100** *0055* on feature *F*1 to the gene in the existing chromosome. The new genetic code of the chromosome is *0002* **6510** *0002* **3763** *4586* **2130** *0500* **0001** *2001* **6011** *0100* **0055**. This is equivalent to a chromosome with relevance $\Theta = 0.651$ and $g_1(m_{F2}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.3763}{0.213}\right)-\tanh\left(\frac{x-0.4586}{0.05}\right)\right]$ and $g_2(m_{F1}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.2001}{0.01}\right)-\tanh\left(\frac{x-0.6011}{0.0055}\right)\right]$.

Gene Relevance Mutation: This operation adds variation to a gene by randomly changing the w_Θ weight of a gene in the chromosome. Figure 1(e) shows relevance of the feature space when increasing the relevance $\Theta(\zeta_2)$ from 0.651 to 0.9999. The new genetic code of the chromosome is *0002* **9999** *0002* **3763** *4586* **2130** *0500* **0001** *2001* **6011** *0100* **0055**.

Gene Constant Mutation: This operation replaces an exon in a randomly selected gene. The selection of the new exon is performed by a random operation. Figure 1(f) shows relevance of the feature space after replacing the exon *0002* **3763** *4586* **2130** *0500* with *0001* **5160** *7613* **0501** *0500*. The new exon is equivalent to $g(m_{F1}) = \frac{1}{2}\left[\tanh\left(\frac{x-0.516}{0.0501}\right)-\tanh\left(\frac{x-0.7613}{0.05}\right)\right]$. The final code of the chromosome is *0002* **9999** *0001* **5160** *7613* **0501** *0500* **0001** *2001* **6011** *0100* **0055**.

Gene Cross Over: This operation switches subsets of exons between two randomly selected genes. Each subset of exons to be switched is also randomly selected. Figure 1(g) shows relevance of the feature space after replacing the second exon in previously described gene *0001* **2001** *6011* **0100** *0055* with the exon from the first random mutation *0002* **1012** *3410* **0200** *0513*. The final code of the chromosome is *0002* **9999** *0001* **5160** *7613* **0501** *0500* **0002** *1012* **3410** *0200* **0513**.

Gene Shrink Mutation: This operation removes an exon in a randomly selected gene. The selection of the exon to be removed is performed by a random operation. Figure 1(h) shows relevance of the feature space after removing the exon *0002* **1012** *3410* **0200** *0513* from the gene described above. The final code of the chromosome is *0001* **9999** *0001* **5160** *7613* **0501** *0500*.

Chromosome Grow Mutation: This operation adds a gene to a randomly selected chromosome with a probability directly proportional with chromosome's relevance. The new gene is generated randomly. Figure 1(i) shows relevance of the feature space after adding a new gene with two exons: *0001* **1210** *4100* **0200** *0500* and *0002* **6200** *8522* **0300** *0050* and weight $w_\Theta = 0.712$. The newly added gene has the code: *0002* **7120** *0001* **1210** *4100* **0200** *0500* **0002** *6200* **8522** *0300* **0050** while the final code of

the chromosome is *0001* **9999** *0001* **5160** *7613* **0501** *0500* **0002** *7120* **0001** *1210* **4100** *0200* **0500** *0002* **6200** *8522* **0300** *0050*.

Chromosome Constant Mutation: This operation randomly selects a chromosome and changes the associated feature for one of its genes. Figure 1(j) shows relevance of the feature space after the feature of the first gene was changed from *F*1 to *F*2 with the resulting code: *0002* **9999** *0002* **5160** *7613* **0501** *0500*. The new chromosome has the code *0001* **9999** *0002* **5160** *7613* **0501** *0500* **0002** *7120* **0001** *1210* **4100** *0200* **0500** *0002* **6200** *8522* **0300** *0050*.

Chromosome Cross Over: This operation switches subsets of genes between two randomly selected chromosomes. Each subset of genes to be switched is also randomly selected. Figure 1(k) shows relevance of the feature space after switching the first gene of the chromosome in Figure 1(d) with the first gene in the previously described chromosome. The final code of the chromosome is *0002* **6510** *0002* **3763** *4586* **2130** *0500* **0001** *2001* **6011** *0100* **0055** *0002* **7120** *0001* **1210** *4100* **0200** *0500* **0002** *6200* **8522** *0300* **0050**.

Chromosome Shrink Mutation: This operation removes a gene from a chromosome with the intent to reduce the complexity of the DNA sequence. The probability of this operation is inversely proportional with the relevance of each chromosome. Figure 1(l) shows relevance of the feature space after removing the second gene from the chromosome. The final code of the chromosome is *0002* **6510** *0002* **3763** *4586* **2130** *05 00* **0001** *2001* **6011** *0100* **0055**

Chromosome Reproduction: This operation makes an exact copy of a chromosome and adds it to the new DNA sequence. The selection of chromosomes used in genetic operations is determined using the roulette wheel selection algorithm [44], which allocates a chance of selection proportional to the fitness of each semantic model in the population.

Figure 2. Flowchart for generating a semantic model using genetic operations.
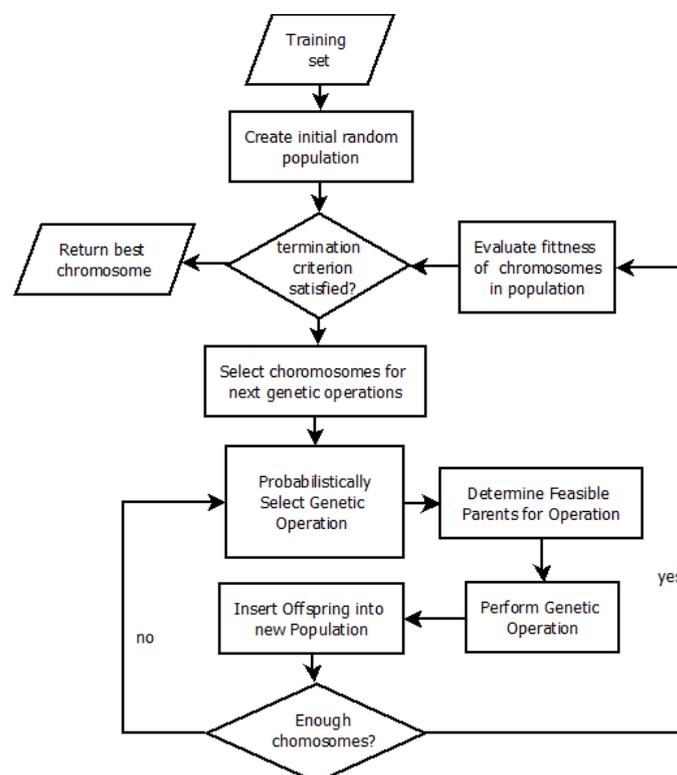
Figure 2 shows the flowchart for generating a semantic model using genetic operations. The input parameters for this process are a training set containing image features that were labeled by image analysts with one or multiple semantics $\varsigma \in S$. This algorithm also takes, as input, the following parameters: the number of chromosomes in each generation of population, the maximum number of generations (iterations) the algorithm will execute, and a threshold on the quality of ranking for which the algorithm would terminate. The algorithm starts with a population in which each chromosome, gene, and exon was randomly generated. The quality of ranking is then evaluated using the MAP measure and it is shown in Equation (6). The top chromosomes are then selected as parents for the chromosomes in the next generation, which is generated using the genetic operation explained in Section 2.3. Finally, when the termination criterion was met—either the quality of ranking of the top chromosome exceeded the preset threshold or the maximum number of iterations was completed—the algorithm returns the most fitted chromosome. This chromosome is converted to a semantic model that is used for ranking of new, unlabeled images.
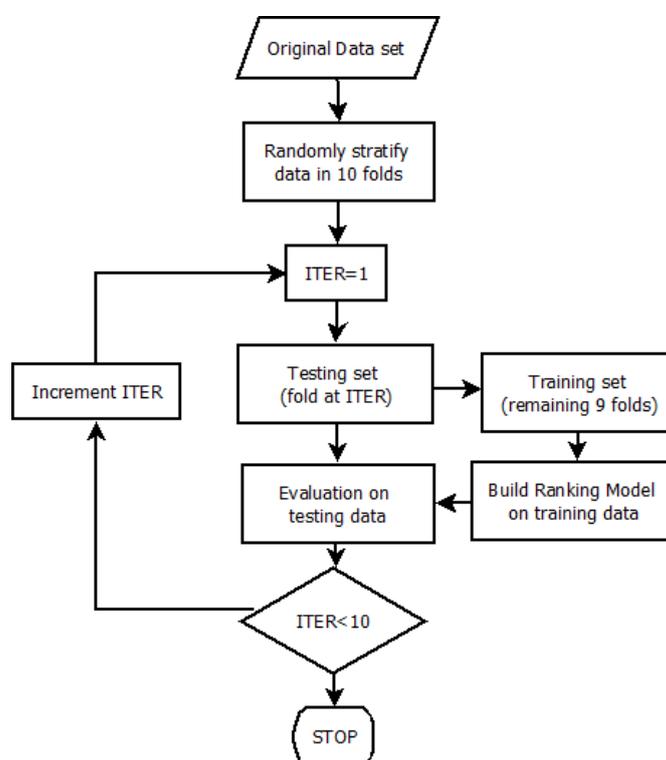
3. Evaluation

We designed three experiments to evaluate the relevance of applying genetic optimization methods to ranking images by semantics: (1) we evaluate the performance of the proposed approach over a large number of genetic operations; (2) we perform an in-depth comparative evaluation of Associative & SFFS and the proposed approach (Associative & Genetic); and (3) we compare the performance of the proposed method with that of six other methodologies. For each experiment we followed the procedure shown in Figure 3: First, the original data was separated into ten subsets using a stratified strategy [45] to ensure that each semantic class in proportionally represented in each fold. Next, using a ten-fold iteration approach, data was separated into testing containing a different subset for each fold and training containing the remaining folds. Then, ranking models were built on the training data and evaluated on testing data. This approach is different from the Associative & SFFS in that the latter uses the following procedure: (1) use Apriori algorithm to generate a large number candidate feature subspaces; (2) sort the generated associations by a harmonic average of confidence and support; (3) generate the parametric sigmoid model using least square method using data distribution over the feature subspace; and (4) generate candidate semantic models by repeatedly adding and applying SFFS methods to the best candidate model.

For our experiments we used two datasets: 2010 WROC satellite imagery of Wisconsin [46] and UCI Statlog Landsat Multi-Spectral satellite [47]. The 2010 WROC satellite imagery contains 18 3-band GeoTIFF image tiles $15,678 \times 11,105$ pixels collected in spring 2010. Each of these tiles was further partitioned in minimal overlapping $1,000 \times 1,000$ tiles. For each tile, a feature extraction algorithm was applied to include the following: For color we extract features from the gray, R, G, B, H, S, V channel as well as color texture. For texture, we extract autocorrelation, contrast, correlation, energy, entropy, Inverse difference moment, and homogeneity. For objects, we extract gray mean, area, centroid, bounding box, major and minor axis length, eccentricity, orientation, convex area, filled area, Euler number, equivalent diameter, solidity, perimeter, and phase congruency. We perform the feature extraction using the Image Processing Toolbox from MatLab. For each of these features average, quartile, standard deviation, skewness, and kurtosis were calculated resulting in a 292 feature

vector for each tile. Further, we selected a number of 100 tiles that were labeled with one or more labels from the *Urban Area* (L100), *Agriculture* (L110), *Grassland* (L150), *Forest* (L160), *Open Water* (L200), *Wetland* (L210), *Barren* (L240), *Shrubland* (L250). In this subset, a number of 72 tiles were labeled with two semantics: *Barren* (L240) overlaps with *Agriculture* (L110) in 26 tiles, with *Grassland* (L150) in 4 tiles, with *Forest* (L160) in 5 tiles, and with *Wetland* (L210) in 4 tiles. Also, *Shrubland* (L250) overlaps with *Grassland* (L150) in 4 tiles and with *Forest* (L160) in 29 tiles. The second data set is the UCI Statlog Landsat Multi-Spectral satellite dataset that contains 6,435 satellite images that were labeled with one of six different soil types: *red soil* (L1), *cotton crop* (L2), *grey soil* (L3), *damp grey soil* (L4), *soil with vegetation stubble* (L5), or *very damp grey soil* (L7). For each image, a 36-dimensional feature space was extracted with the feature corresponding to the 9 intensity values of a 3×3 pixel region (with overlapping regions) in two visible and two near infra-red spectral bands. Semantic models were trained on a randomly selected training set that contains 90% of data while testing was performed on the remaining 10% of data.

Figure 3. Flowchart for the experimental setting.

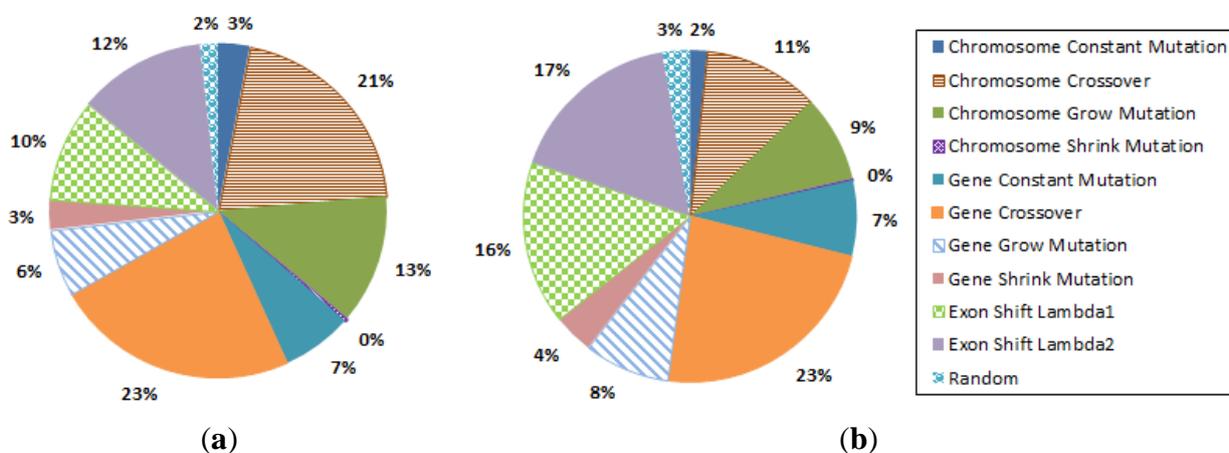*3.1. In-Depth Evaluation of Genetic Operations in the Proposed Method*

For the proposed method, we have recorded each genetic operation that was performed on the genetic population. This resulted in a number of 90,000 genetic operations for the experiments over the UCI Statlog Landsat data set and 120,000 genetic operations for the experiments over the WROC data set. The percentage for each individual operation performed is shown in Figure 4. For example, the crossover operations accounted for 57% of all the operations equally distributed over chromosome and gene mutations. Due to the randomness of the genetic operations, we observed minimal percentile variations for the experiments on the two data sets.

Figure 4. Genetic operations performed as percentage when ranking images by semantics.
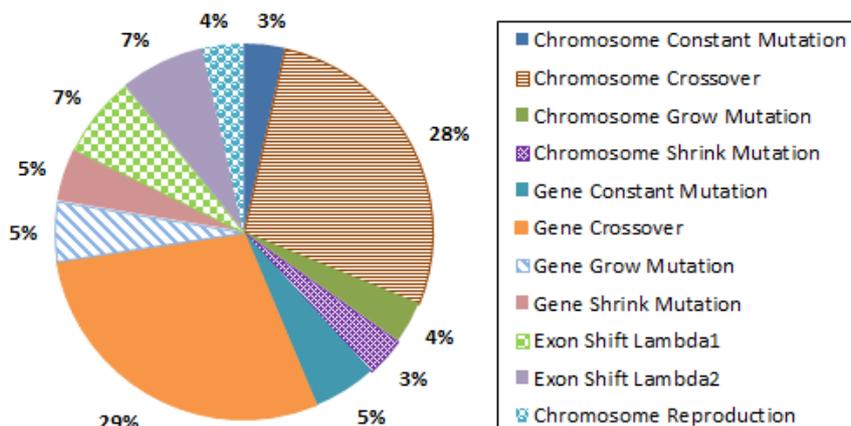
We have also recorded the genetic operation that resulted in the best performing chromosome for each mutated population for each data set-semantic-fold combination. Out of the 21,000 mutated populations, only 6,491 returned better fitted models with 3,517 and 2,974 populations for the UCI Statlog Landsat and WROC data set respectively or a 30.9% genetic mutation success rate. Figure 5 shows the percentage of operations that returned improved semantic models. For example, this figure shows that overall crossover mutations tended to contribute less than average for improvements in semantic models. They returned the best models in 44% and 34% for the UCI Statlog Landsat data set and WROC data sets respectively. On the other end, exon shifts were the most successful in improving semantic models with percentages of 22% and 33% respectively, although they accounted for only 14% of the total genetic operations. It is also noted that the least likely to improve are the models with percentages of less than 0.5%.

Figure 5. Relevant genetic operations as percentage when ranking images by semantics on the (**a**) UCI Statlog Landsat data set and (**b**) WROC data sets.

(**a**)   (**b**)

3.2. In-Depth Evaluation of Associative Methods for Ranking

To evaluate the difference between the two associative methods (Associative & SFFS and Associative & Genetic) we have recorded the MAP measure at each iteration for both the training and

the testing dataset. In this experiment, each generated model is considered one iteration. For example, the Associative and Genetic method with a population of 10 chromosomes and 150 generations will generate 1,500 iterations. At each iteration a new chromosomes/semantic models is evaluated. Similarly, for the Associative & SFFS method, a new iteration is generated by adding a new association to the model. Figures 6 and 7 show the range of MAP when ranking images from the WROC data set for the Associative & SFFS and Associative & Genetic respectively. The results from the UCI Statlog Landsat data set were omitted due to lack of space, but are similar in behavior. For example, in Figure 6, at iteration 1,250 the average MAP returned by the Associative & SFFS method on the training set was 72.33% and 59.99% on the testing set. This shows that on average the Associative & SFFS method overfits the model to the training data by 12.32%. For the same iteration, the MAP value ranged between 49.94% and 98.54% on the training set and between 30.81% and 87.71% for the testing set. Also, this figure shows that the last 150 iterations that produced a better MAP on the training set overfitted the model because they reduced the MAP on the testing set by 0.4%.

Figure 6. Range of MAP by iteration for the (**a**) training and (**b**) testing data sets when ranking images from the WROC data set using the Associative & SFFS.
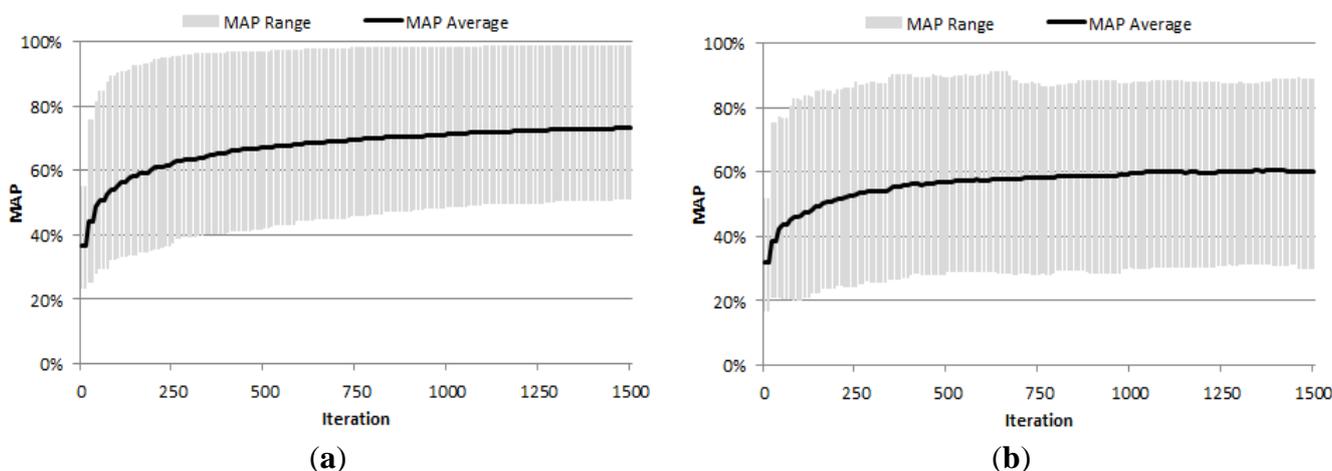

(**a**)

(**b**)

Figure 7. Range of MAP by iteration for the (**a**) training and (**b**) testing data sets when ranking images from the WROC data set using the Associative & Genetic.
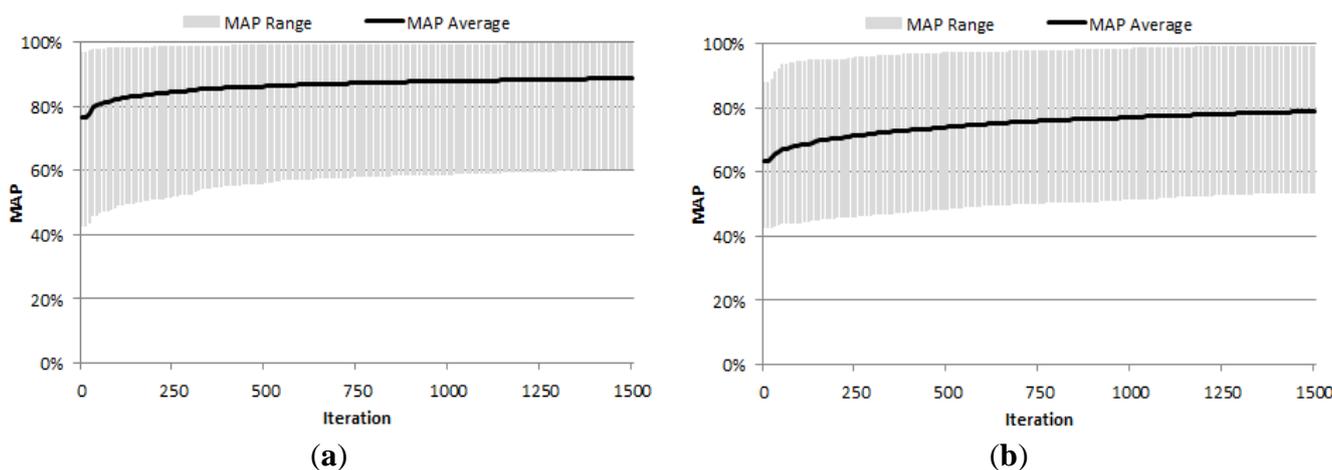

(**a**)

(**b**)

Figure 7 shows similar results for Associative & Genetic method. At iteration 1,250 the average MAP returned by this method were 78.06% and 72.34% on the training and testing set, respectively. This shows that on average the Associative & Genetic method overfits the model to the training data by 5.72% on average, which is less than half of the variation measured for the Associative & SFFS. Similarly, the range of MAP values was smaller than in the case of Associative and SFFS method with values between 53.09% and 98.86% for the training set and 44.45% and 97.27% for the testing set. For the same iteration, the MAP value ranged between 49.94% and 98.54% on the training set and between 30.81% and 87.71% for the testing set.

The results in these figures show that the advantages of the Associative and Genetic method are two-fold: (a) better trained models that achieve higher average MAP on the training data and (b) less overfitting of the models to the training data. To further evaluate the reasons the Associative and SFFS methods overfit, we also recorded the number of rules in the semantic models generated by the two methods. The results of this experiment are shown in Figure 8 for the UCI Statlog Landsat data set and in Figure 9 for the WROC data set. For example, in Figure 8(a), the average number of rules in a semantic model generated by Associative and SFFS at iteration 1,250 on the UCI Statlog Landsat data set is 65.25% with a minimum and maximum of 27 and 1,224 rules, respectively. For the same iteration and data set, the Associative & Genetic method returned on average 12.85 rules with a minimum and maximum of 4 and 18 rules respectively. This shows that the advantage of the proposed method over the Associative & SFFS is given by its parsimonious models [48] which, on average, are five times smaller in size.

Figure 8. Range of rule count by iteration ranking images using Associative & SFFS method on the (**a**) UCI Statlog Landsat and (**b**) WROX data sets.

(**a**)　　　　　　　　　　　　　　　(**b**)

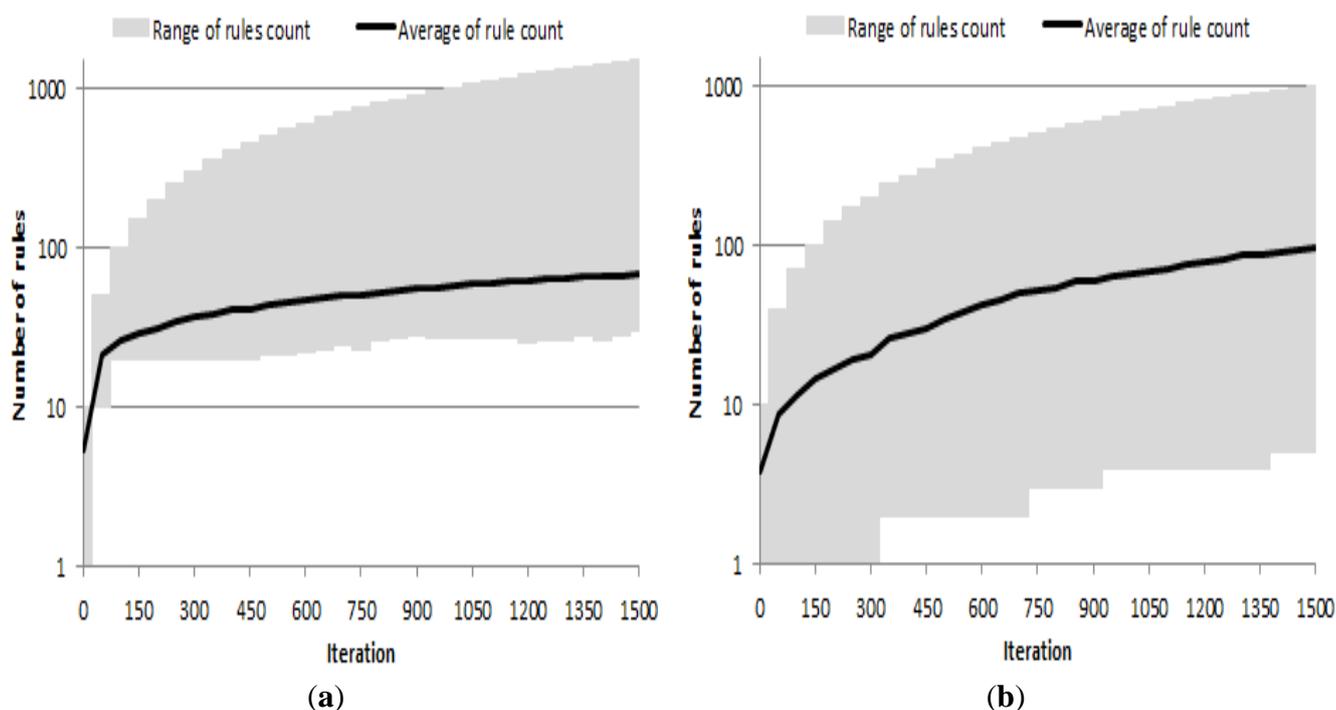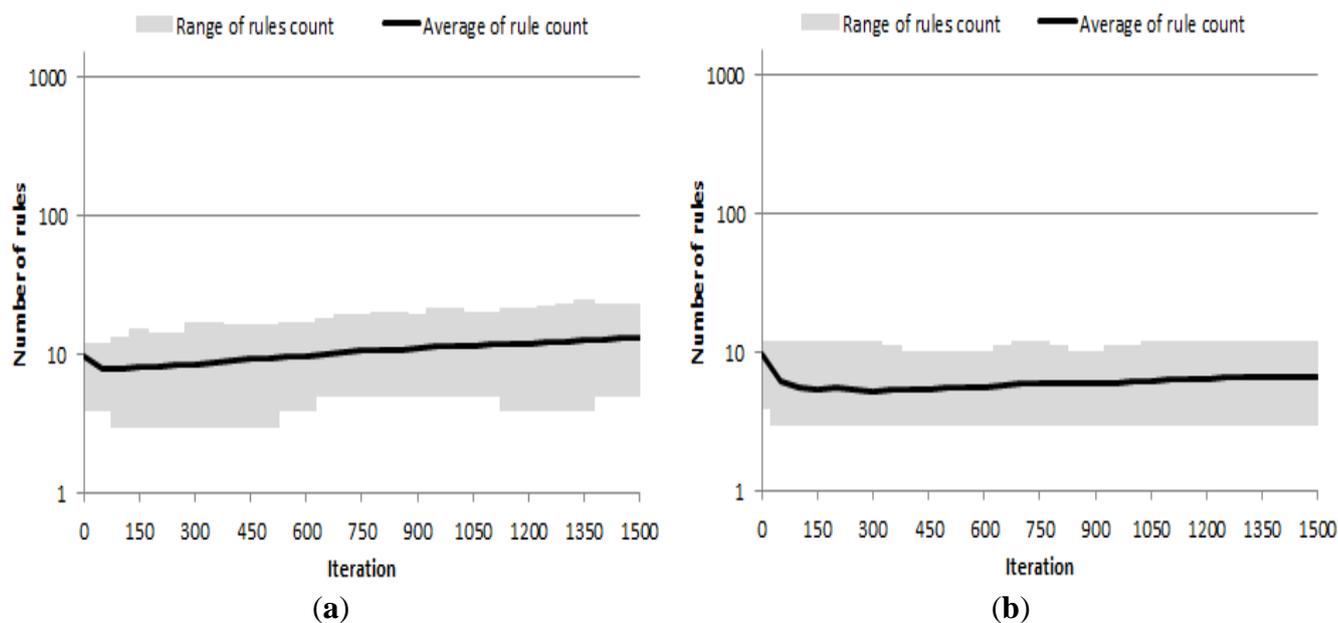**Figure 9.** Range of rule count by iteration ranking images using Associative & Genetic method on the (**a**) UCI Statlog Landsat and (**b**) WROX data sets.

(**a**)   (**b**)

3.3. Comparative Study of Ranking Performance

For this experiment, we designed seven, ten-fold ranking experiments: (1) additive associative combined with SFFS [8], (2) ensemble ranking using artificial neural networks (ANN) [49] with AdaBoost [50], (3) ensemble ranking using C4.5 decision tree (C4.5) [27] with AdaBoost, (4) Logistic Model Trees [51], (5) ensemble ranking using TreeRank with a SVM kernel [52], (6) ensemble ranking using Tree Forest with a SVM kernel [52], and (7) additive associative ranking combined with genetic operations as described in Section 2. All these experiments were implemented in the R statistical environment [53]. For experiments (2) to (6) we have used packages available in R. For experiments (1) and (7) we have used 1,500 optimization steps. The data were preprocessed by applying the Boruta algorithm [54] for variable selection.

Figure 10 shows a comparison of the seven methods for ranking of images in the two data sets described above using mean average precision (MAP) of ranking. When ranking images from the UCI Statlog Landsat dataset, the proposed method retrieved the best results with an average MAP of 87.93%, followed by LMT with a MAP of 86.11%. Both these methods returned a low standard deviation of 2.49% for the Associative & Genetic method and 3.44% for LMT. Low performance was returned by ANN & Adaboost—which is prone to overfitting—and SVM & TreeRank—which is a non-ensemble method—with an average MAP of 66.01% and 71.79%, respectively. These two methods also returned a higher standard deviation of MAP with 6.56% and 6.61%, respectively. When ranking images from the WROC data set, the proposed method retrieved second to best results with an average MAP of 73.30% next to SVM & TreeForest with a MAP of 74.26%. However, the proposed method returned a slightly lower standard deviation at 9.55% as compared to 10.29% for the SVM & TreeForest. LMT ranked fourth for this dataset behind C4.5 & Adaboost. Similarly to the previous results, low performance was returned by ANN & Adaboost, Associative & SFFS, and SVM & TreeRank, with an average MAP of 59.06%, 60.12% and 60.47%, respectively.

Figure 10. MAP results for comparative experiments for ranking images by semantics using different ranking methods on (**a**) the UCI Statlog Landsat data set and (**b**) WROC data set.

(**a**)   (**b**)
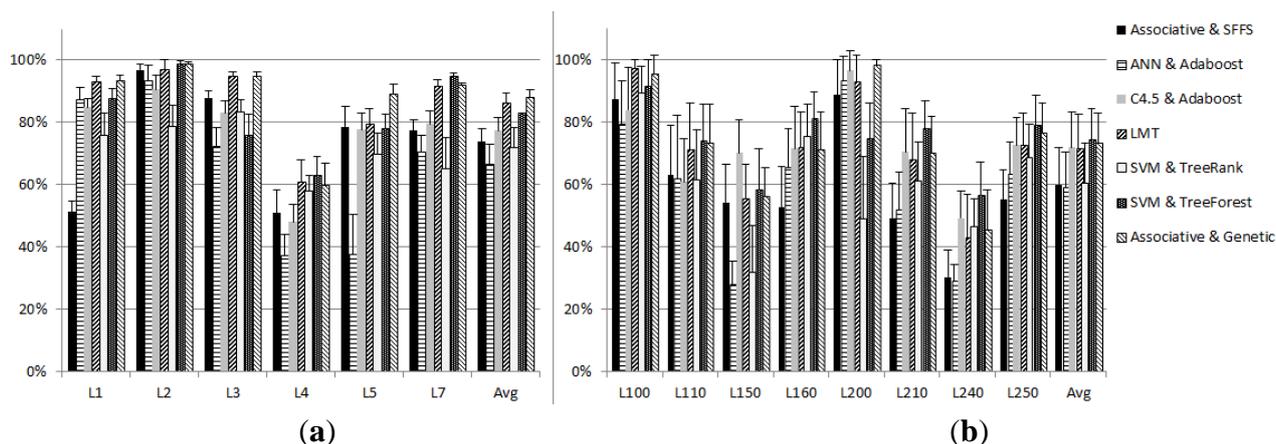
When examining the MAP results for each semantic label, we observe wide variations in performance. For example, the Associative & SFFS method returns a MAP of 51.25% when ranking the semantic *red soil* (L1) on the UCI Statlog Landsat data set. This is 24.65% lower than the next performing method (SVM & TreeRank). On the same data set, the ANN & Adaboost method show very low MAP for the *damp grey soil* (L4) and *soil with vegetation stubble* (L5) with MAP values of 37.40% and 37.80% respectively. The ANN & Adabost also returned low performance for the *Grassland* (L150), and *Barren* (L240) semantics of the WROC data set with MAP values of 27.92% and 29.15% respectively. Variations are also observed in the top performing methods: The proposed method is the best when ranking five semantics across the two datasets, while the SVM & TreeRank is the best when ranking nine semantics across the two datasets. However, on average, the proposed method returned the best results across the two datasets with an average of 80.61%, followed by LMT with 78.85% and SVM & TreeForest with 78.69%. This shows a more consistent behavior of the proposed method with less likelihood of overfitting/underfitting.

For a more in-depth analysis of accuracy of the ranked results we provide precision and recall metrics. Precision measures how many of the images retrieved in a search by semantic are actually relevant, while recall measures how many of the images that are relevant to the target semantic have actually been retrieved. Figure 11 shows in-depth interpolated precision-recall measures for the seven ranking methods. For example, when ranking images from the UCI Statlog Landsat data set, the proposed method returns on average a precision of 95.47% when 20% of the relevant images were recalled. For the same data set and recall level, LMT returned 94.76% while SVM & TreeForest returned 86.61%. The results over the WROC data set show that the proposed method returns the best precision at lower recall rates of less than 30% but performs worse at higher levels of recall. For example, on the WROC data set and a recall of 60%, the Associative & Genetic method ranks fourth with a precision of 68.81% behind SVM & TreeForest, C4.5 and Adaboost, and SVM & TreeRank that returns precisions of 77.09%, 75.85%, and 74.01%, respectively. This trend is noticed also for the Associative & SFFS method which is top three in performance for recalls less than 20% but exhibits performance degradation at higher levels. Associative & SFFS, Associative & Genetic, and LMT show the lowest precision levels at 100% recall which hints to the fact that these methods fail to cover the

whole feature universe, and consequently do not rank some images. This suggests some overfitting issues of models created using these methods which are less evident for methods such as SVM & TreeForest, SVM & TreeRank, or C4.5 and Adaboost.

Figure 11. Average precision-recall results for comparative experiments for ranking images by semantics using seven different ranking methods on (**a**) the UCI Statlog Landsat data set and (**b**) WROC data set.

(**a**)   (**b**)
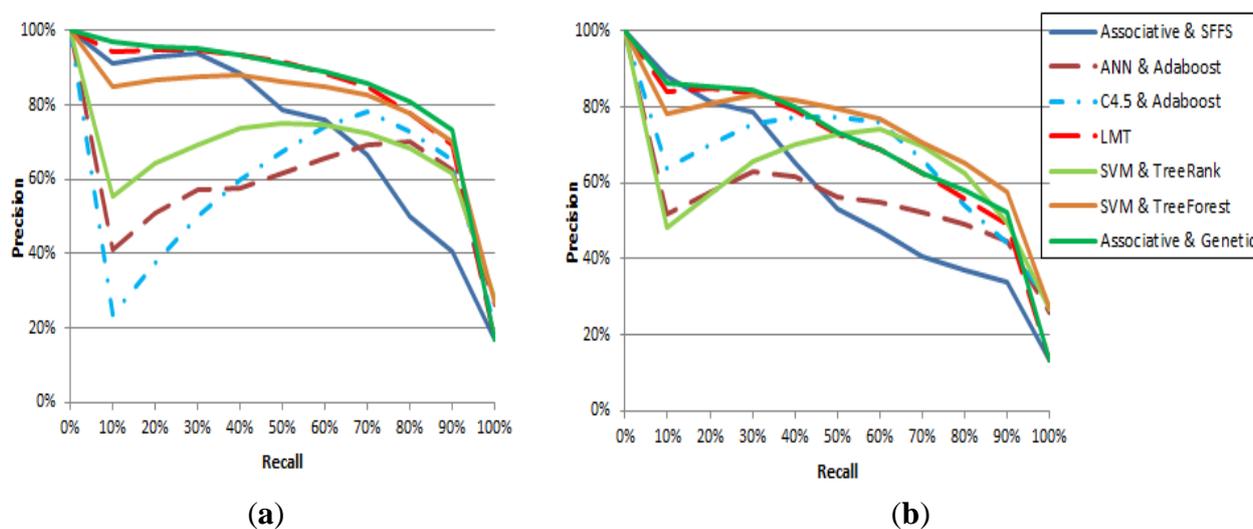
Overall, our conclusion for this experiment is that there are several reasons that cause variations in performance for the methods that we have analyzed. For example, the Associative & SFFS is able to rank only those images for which the Apriori algorithm returned strong associations and athe drop in precision at high recall values signifies that there are some images that are not mapped into any generated feature subspace. The SVM & TreeRank algorithm is the only algorithm that does not use ensemble methods and it is likely to overfit. We observe that ranking quality increases significantly, once ensemble procedures replace TreeRank. Overfitting is likely to be the cause of poor performance returned by ANN & AdaBoost which heavily depends on the characteristic of the neural network while the C4.5 and AdaBoost returns poor result due to its greedy nature.

4. Conclusions and Future Work

We have developed an approach for generating associative models for ranking satellite image regions by land cover. The results of our comparative studies show that the proposed method performs better or has similar performance to that of other ensemble methods. Our method applies genetic methods to return better precision on new untested data while avoiding overfitting by reducing the local minima issues existent in additive models. Overall our results show that the genetic method discovered better association rules faster than the existent additive method. This shows that associative methods offer promising alternatives to visual patterns found in images, although they are prone to overfitting. The key to their success is an adequate learning procedure that is able to avoid local minima. Previous associative approaches use association rule mining algorithms to identify relevant feature spaces but suffer from inadequate measure of association rule relevance, such as support and confidence, which are not optimal for ranking problems. Although our experiments did not provide a clear evidence of the

superiority of the proposed method when compared with other state-of-the-art approaches, the easy to understand nature of the generated models provide a benefit for future research into areas such as expertise and training of image analysts. Genetic models have also the advantage of randomly selecting and testing new feature subspaces which result in better models in shorter time. Although not specifically measured, training time is an important component in any ranking algorithm. As with any other ensemble method, training the proposed method is proportional to the size of the training set, number of rules in a semantic model and number of iterations. This is an improvement over SFFS methods for which reducing the number of rules in a model requires quadratic complexity of number of rules.

Our future work includes a more comprehensive evaluation on different image modalities and semantic sets, especially for data sets that exhibit overlapping visual patterns and which are more difficult to rank. Specifically to genetic operations, we plan to evaluate a better mix of genetic operations that would further improve the performance. We also want to address the training time, which is a well-known drawback of ensemble methods. Cross-region image ranking is also an area of future research since ranking methods are known to return lower precision on data from different regions of the globe.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Datcu, M.; Seidel, K. Human-centered concepts for exploration and understanding of earth observation images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 601–609.
2. Tseng, M.-H.; Chen, S.-J.; Hwang, G.-H.; Shen, M.-Y. A genetic algorithm rule-based approach for land-cover classification. *ISPRS J. Photogramm.* **2008**, *63*, 202–212.
3. Mennis, J.; Guo, D. Spatial data mining and geographic knowledge discovery—An introduction. *Comput. Environ. Urban Syst.* **2009**, *33*, 403–408.
4. Datcu, M.; Seidel, K. Image Information Mining: Exploration of Image Content in Large Archives. In Proceedings of 2000 IEEE Aerospace Conference, Big Sky, MT, USA, 18–25 March 2000; Volume 3, pp. 253–264.
5. Hsu, W.; Lee, M.L.; Zhang, J. Image mining: Trends and developments. *J. Intell. Inf. Syst.* **2002**, *19*, 7–23.
6. Liu, Y.; Zhang, D.; Lu, G.; Ma, W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recogn.* **2007**, *40*, 262–282.
7. Aksoy, S.; Cinbis, R. Image mining using directional spatial constraints. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 33–37.
8. Barb, A.; Shyu, C.-R. Visual-semantic modeling in content-based geospatial information retrieval using associative mining techniques. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 38–42.
9. Blanchart, P.; Datcu, M. A semi-supervised algorithm for auto-annotation and unknown structures discovery in satellite image databases. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **2010**, *3*, 698–717.

10. Bratasanu, D.; Nedelcu, I.; Datcu, M. Bridging the semantic gap for satellite image annotation and automatic mapping applications. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **2011**, *4*, 193–204.

11. Durbha, S.; King, R. Semantics-enabled framework for knowledge discovery from earth observation data archives. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2563–2572.

12. Klaric, M.; Scott, G.; Shyu, C.-R. Multi-index multi-object content-based retrieval. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4036–4049.

13. Lienou, M.; Maitre, H.; Datcu, M. Semantic annotation of satellite images using latent dirichlet allocation. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 28–32.

14. Molinier, M.; Laaksonen, J.; Hame, T. Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 861–874.

15. Scott, G.; Klaric, M.; Davis, C.; Shyu, C.-R. Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1603–1616.

16. Shyu, C.-R.; Klaric, M.; Scott, G.J.; Barb, A.S.; Davis, C.H.; Palaniappan, K. GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 839–852.

17. Sjahputera, O.; Scott, G.; Claywell, B.; Klaric, M.; Hudson, N.; Keller, J.; Davis, C. Clustering of detected changes in high-resolution satellite imagery using a stabilized competitive agglomeration algorithm. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4687–4703.

18. Li, W.; Raskin, R.; Goodchild, M. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1415–1435.

19. Agrawal, R.; Imielinski, T.; Swami, A. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (SIGMOD '93), Washington, DC, USA, 26–28 May 1993; pp. 207–216.

20. Agrawal, R.; Srikant, R. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94), San Francisco, CA, USA, 12–15 September 1994; pp. 487–499.

21. Ladner, R.; Petry, F.E.; Cobb, M.A. Fuzzy set approaches to spatial data mining of association rules. *Trans. GIS* **2003**, *7*, 123–138.

22. Huang, Y.; Chang, T.; Kao, L. Using fuzzy SOM strategy for satellite image retrieval and information mining. *J. Syst. Cybern. Inf.* **2008**, *6*, 56–61.

23. Thabtah, F. A review of associative classification mining. *Knowl. Eng. Rev.* **2007**, *22*, 37–65.

24. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, *Ser*; Springer: New York, NY, USA, 2009.

25. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271.

26. Li, W.; Han, J.; Pei, J. CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. In Proceeding of the IEEE International Conference on Data Mining (ICDM 2001), San Jose, CA, USA, 29 November–2 December 2001; pp. 369–376.

27. Quinlan, J. Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.* **1996**, *4*, 77–90.

28. Pudil, P.; Novovicová, J.; Kittler, J. Floating search methods in feature selection. *Pattern Recogn. Lett.* **1994**, *15*, 1119–1125.

29. Ribeiro, M.X.; Bugatti, P.H.; Traina, C., Jr.; Marques, P.M.A.; Rosa, N.A.; Traina, A.J.M. Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques. *Data Knowl. Eng.* **2009**, *68*, 1370–1382.

30. Freitas, A. A Review of Evolutionary Algorithms for Data Mining. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: New York, NY, USA, 2010; pp. 371–400.

31. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1989.

32. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; The MIT Press: Cambridge, MA, USA, 1992.

33. Momm, H.; Easson, G.; Kuszmaul, J. Evaluation of the use of spectral and textural information by an evolutionary algorithm for multi-spectral imagery classification. *Comput. Environ. Urban Syst.* **2009**, *33*, 463–471.

34. Shad, R.; Mesgari, M.S.; Abkar, A.; Shad, A. Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *Comput. Environ. Urban Syst.* **2009**, *33*, 472–481.

35. Zhang, X.; Wang, J.; Wu, F.; Fan, Z.; Li, X. A Novel Spatial Clustering with Obstacles Constraints Based on Genetic Algorithms and k-Medoids. In Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA '06), Washington, DC, USA, 16–18 October 2006; Volume 1, pp. 605–610.

36. Gao, L.; Dai, S.; Zheng, S.; Yan, G. Using Genetic Algorithm for Data Mining Optimization in an Image Database. In Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), Haikou, Hainan, China, 24–27 August 2007; Volume 3, pp. 721–723.

37. Bandyopadhyay, S.; Maulik, U.; Mukhopadhyay, A. Multiobjective genetic clustering for pixel classification in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1506–1511.

38. De Stefano, C.; Fontanella, F.; Marrocco, C. A GA-Based Feature Selection Algorithm for Remote Sensing Images. In Proceedings of the 2008 Conference on Applications of Evolutionary Computing (Evo'08), Naples, Italy, 26 March 2008; Springer-Verlag: Berlin/Heidelberg, Germany, 2008; pp. 285–294.

39. Da Silva, S.F.; Ribeiro, M.X.; Batista Neto, J.A.D.E.S.; Traina, C., Jr.; Traina, A.J.M. Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decis. Support Syst.* **2011**, *51*, 810–820.

40. Mahrooghy, M.; Younan, N.; Anantharaj, V.; Aanstoos, J.; Yarahmadian, S. On the use of the genetic algorithm filter-based feature selection technique for satellite precipitation estimation. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 963–967.

41. Barb, A.S.; Barb, C.S. Genetic Methods for Associative Semantic Ranking of Landsat Image Regions by Land Cover. In Proceedings of Image Information Mining Workshop, Oberpfaffenhofen, Germany, 24–26 October 2012; European Space Agency and Joint Research Commissions: Oberpfaffenhofen, Germany; pp. 102–105.

42. Syswerda, G. A Study of Reproduction in Generational and Steady-State Genetic Algorithms. In Proceeding of the First Workshop on Foundations of Genetic Algorithms, Bloomington Campus, IN, USA, 15–18 July 1990; pp. 94–101.

43. Vavak, F.; Fogarty, T. Comparison of Steady State and Generational Genetic Algorithms for Use in Nonstationary Environments. In Proceedings of IEEE International Conference on Evolutionary Computation, Nagoya, Japan, 20–22 May 1996 ; pp. 192–195.

44. Davis, L. *Handbook of Genetic Algorithms*; Van Nostrand Reinhold: New York, NY, USA, 1991.

45. Chauvet, G.; Tillé, Y. A fast algorithm for balanced sampling. *Comput. Stat.* **2006**, *21*, 53–62.

46. *The Wisconsin Regional Orthophotography Consortium (WROC)*, 2012. Available Online: http://www.ncwrpc.org/WROC/ (accessed on 1 August 2012).

47. Asuncion, A.; Newman, D.J. *UCI Machine Learning Repository*, 2007. Available Online: http://www.ics.uci.edu/~mlearn/MLRepository.html (accessed on 1 March 2012).

48. Sober, E. *What is the Problem of Simplicity?*; Zellner, A., Keuzenkamp, H., McAleer, M., Eds.; Cambridge University Press: Cambridge, UK, 2002.

49. Bishop, C. *Neural Networks for Pattern Recognition*, 1st ed.; Oxford University Press: New York, NY, USA, 1996.

50. Freund, Y.; Schapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.

51. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Mach. Learn.* **2005**, *59*, 161–205.

52. Clemencon, S.; Vayatis, N. Tree-based ranking methods. *IEEE Trans. Inf. Theory* **2009**, *55*, 4316–4336.

53. R Development Core Team. *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*, Vienna, Austria, 2011. Available Online: http://www.R-project.org (accessed on 1 November 2012)

54. Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13.