

Probabilistic Inference over Image Networks

Claudio Taranto, Nicola Di Mauro, and Floriana Esposito

Department of Computer Science, University of Bari "Aldo Moro"
via E. Orabona, 4 - 70125, Bari, Italy
{claudio.taranto,ndm,esposito}@di.uniba.it

Abstract. Digital Libraries contain collections of multimedia objects providing services for the management, sharing and retrieval. Involved objects have two levels of complexity: the former refers to the inner object complexity while the latter takes into account the implicit/explicit relationships among objects. Traditional machine learning classifiers do not consider the relationships among objects assuming them independent and identically distributed. Recently, *link-based* classification methods have been proposed, that try to classify objects exploiting their relationships (links). In this paper, we deal with objects corresponding to digital images, even if the proposed approach can be naturally applied to different kind of multimedia objects. Relationships can be expressed among the features of the same image or among features belonging to different images. The aim of this work is to verify whether a link-based classifier based on a Statistical Relational Learning (SRL) language can improve the accuracy of a classical k-nearest neighbour approach. Experiments will show that the modelling of the relationships in a real-word dataset using a SRL model reduces the classification error.

1 Introduction

Digital Libraries organized digital collections of multimedia objects available online in computer processable form [5]. These libraries also comprise services and infrastructures to manage, store, retrieve and share objects. Digital images represent a component of multimedia objects involved in the digital library universe. The Machine Learning scientific community has developed methods and models to address the emerging issues arising from the management of these data types. One of main issues concerns the image classification, which addresses the need to automate the process of assigning a class label to a given image. This problem has been mainly tackled by adopting a learning-based or a non-parametric classifier [3]. The former approach requires a training phase in which a learning system induces a model able to identify new unlabelled images, while the latter does not require a learning phase but adopts methods directly applicable to the images to be classified. Both learning approaches require a feature based representation of the images. Furthermore, non-parametric algorithms adopt a similarity measure defined on the feature space aiming to calculate the similarity degree between two images.

Given \mathcal{D} a set of images, the majority of the existing methods adopted for image classification try to estimate the class probability of a query image $q \in \mathcal{D}$, $P(C_q|\mathcal{D}\setminus\{q\})$, assuming that all the images in \mathcal{D} are independent and identically distributed (i.i.d.). In particular, the probability $P(C_q|\mathcal{D}\setminus\{q\})$ is assumed to be factored as in the following way $\prod_i \alpha P(C_q|x_i)$, with $x_i \in \mathcal{D}$.

This paper proposes a Statistical Relational Learning (SRL) [11] method to exploit the relationships among images in order to improve the image classification accuracy by means of a link-based classifier [13, 12]. The main idea is to assume that the images $x_i \in \mathcal{D}$ are not mutually independent and to try to elicit the hidden information representing the probabilistic connections between two images taking into account the possible relationships. To reach this goal, images are represented by means of a complex probabilistic network, where each image corresponds to a node and the connection degree among images is represented by a probabilistic edge. The relationship degree among images may be computed adopting a similarity measure based on their feature based representation. Correlograms [15], a statistics expressing how the colors of an image are spatially correlated, are used in this paper as features representing the images.

The main goal of this paper is to verify whether modelling image classification problem using a SRL language can improve the accuracy of a classical k-nearest neighbour (k-NN) approach. The focus is not on proposing a new image classification algorithm, but on showing that a SRL model can improve a classical classification method. We adopted the probabilistic logic ProbLog [24] as SRL model to describe the structure of the probabilistic network arising from the abstraction process we adopted to represent an image collection. The experimental results obtained from a real world dataset confirmed the validity of the proposed approach.

2 Related Work

The increasing interest in the last years about the problem of image classification provided a lot of models. Here we briefly review some of them that may be related to our proposed statistical relational approach. The Bag of Words model [22] (BOW) has been inherited from the area of document analysis. It describes each image as a collection of features, without analysing neither its geometric structure nor features' position in the image. Given a shared vocabulary for the allowed features, named codebook, each image is then encoded as a distribution of codebook names. In order to classify new images, Probabilistic Latent Semantic Analysis (pLSA) [14], Latent Dirichlet Allocation (DLA) [2] or Vector Space Model (VSM) [25] may be used. The limitation of this approach is that all the spatial relationships among the features are ignored. Part-based methods [4] try to solve the problems of the BOW model. The main idea is to consider an object consisting of a set of N parts connected to each other and describing each component considering specific geometric characteristics (e.g., the constellation model [10]). The adopted feature model, such as color histogram, correlogram and sift, will be enriched by considering location and dimension properties. In

the training phase the system induces a model that will be able to classify objects considering their parts. In order to model the relationships among the images, or among the objects within an image, a graph based representation is usually adopted. This structure provides a powerful model [7], where nodes store local information while the edges encode the spatial relationships, with a companion set of tools, inherited from the graph theory, useful to manage and inspect it.

Given a graph based representation, *collective classification* methods [27] can be used. In collective classification relationships among objects must be taken into account in order to enhance the predictive accuracy of the model: the labelling of an object should depend on the labels of its neighbours. Weighted-Vote Relational Neighbor (WVRN) [20], estimates the probability of a class given a node by summing the probabilities of nodes in the neighbourhood of the same class. Class-Distribution Relational Neighbour (CDRN) [21] uses a vectorial representation of the neighbours by assigning to each node v of known label a vector whose elements represent the probability that a neighbour node have a given label. The Network-Only Bayes (NOB) classifier [6] adopts a naive Bayes approach to compute the label probabilities of a node, assumed to be independent, conditioned on the labels of its neighbours. Finally, the Network-Only Link-Based (NOLB) classifier [19] learns a multiclass logistic regression model using the label distribution in the neighbourhood of nodes with known labels. These methods are based on a link-based classification, but they consider the data as already represented as a network, and then relegating their use for dataset containing explicit relations. While, one of our goals is to elicit the relationships between objects, thereby transforming a set of i.i.d. data in a network and describe the connections found using a relational logic formalism.

3 Image correlograms

There is a large number of features that may be used to describe an image. In [15] has been presented an approach, named *color correlogram*, that combines both global and local image information. This statistics describes how pixels with a given color are spatially distributed in an image, and it is generally more accurate and effective than histogram-based methods. However the time required to compute this statistics may be very high. In [29] has been proposed a new sampling method to approximate the distribution of correlograms proving that the computational time to compute the statistics may be reduce by taking high the similarity based accuracy. In this paper we used image correlograms as a feature representation, even if the generality of the proposed relational approach allows the adoption of any kind of feature.

A correlogram is defined as a table indexed by color pairs, where the k -th entry for the component (i, j) specifies the probability of finding a pixel whose color is j away k pixel from a pixel whose color is i , where k is a distance chosen from a set D .

Let \mathcal{I} be an $n_1 \times n_2$ image, whose colors are quantised into m classes $\mathcal{C} = \{c_1, \dots, c_m\}$ (bins). Given a pixel $p_{xy} \in \mathcal{I}$, $I_{(p)}$ denotes its color and $I_c \triangleq \{p_{ij} \in \mathcal{I} | I_{(p)} = c\}$.

In the following, we assume to use the L_∞ -norm to measure the distance between two pixels. In particular, given $p_{x_1 y_1}$ and $p_{x_2 y_2}$ two pixels, their distance is computed as follows [15]:

$$|p_{x_1 y_1} - p_{x_2 y_2}| \triangleq \max\{|x_1 - x_2|, |y_1 - y_2|\}.$$

The *histogram* h of an image \mathcal{I} is defined as $h_{c_i}(\mathcal{I}) = n_1 n_2 \Pr_{p \in \mathcal{I}}[p \in I_{c_i}]$, for each color class $c_i \in \mathcal{C}$.

Let $d \in D$ be a distance, the correlogram of an image \mathcal{I} is defined as follows:

$$\gamma_{c_i, c_j}^{(d)}(\mathcal{I}) = \Pr_{p_1 \in I_{c_i}, p_2 \in I_{c_j}}[|p_1 - p_2| = d], \quad (1)$$

where $c_i, c_j \in \mathcal{C}$ are two color classes. The size of the color correlogram matrix is $|\mathcal{C}|^2$. A specialisation is the *autocorrelogram* defined as $\alpha_c^{(d)}(\mathcal{I}) = \gamma_{c,c}^{(d)}(\mathcal{I})$.

4 The probabilistic logic ProbLog

Among the many languages considered in SRL there are some based on Logic Programming such as ICL [23], PRISM [16], CLP(BN) [8] and ProbLog [24]. The representation and use of probability theory makes SRL algorithms suitable for combining domain knowledge and data, expressing relationships, avoiding overfitting a model to training data, and learning from incomplete datasets. The probabilistic formalism provides a natural treatment for the stochastic nature of some complex domains. As for classical probabilistic graphical models [18], such as Bayesian networks and Markov networks, statistical relational languages exploit the structure underlying many distributions we want to encode. The same structure often allows the distribution to be effectively used for inference, answering queries using the distribution as a model of the world. Finally, the SRL framework facilitates the effective learning from data of models providing a good approximation to a past experience.

ProbLog is a probabilistic framework that extends Prolog with probabilistic facts and answers several kinds of probabilistic queries, which has been used for learning in the context of large networks where edges are labelled with probabilities. The ProbLog's semantics is an instance of the distributional semantics defined in [26]. A ProbLog program defines a distribution over all its possible non-probabilistic subprograms. Facts are labelled with probabilities and treated as mutually independent random variables indicating whether or not the corresponding fact belongs to a randomly sampled program. The success probability of a query is defined as the probability that it succeeds in such a random subprogram.

A ProbLog program $T = p_1 :: c_1, \dots, p_n :: c_n \cup BK$ defines a probability distribution over subprograms $L \subseteq LT = \{c_1, \dots, c_n\}$:

$$P(L|T) = \prod_{c_i \in L} p_i \prod_{c_i \in LT \setminus L} (1 - p_i).$$

While in Prolog the result of query is binary, i.e. success or failure, in ProbLog the result of a query is the probability of success, i.e. the probability that the query succeeds in a random sample. In particular, the *success probability* $P_s(q|T)$ of a query q in a ProbLog program T is defined as

$$P_s(q|T) = \sum_{L \subseteq LT} P(q|L) \cdot P(L|T),$$

where $P(q|L) = 1$ if there exists a θ substitution such that $L \cup BK \models q\theta$, and $P(q|L) = 0$ otherwise.

The *explanation probability* $P_x(q|T)$ is defined as the probability of the most likely explanation of the proof of the query q :

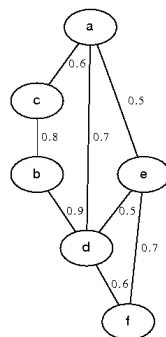
$$P_x(q|T) = \max_{e \in E(q)} P(e|T) = \max_{e \in E(q)} \prod p_i,$$

where $E(q)$ is the set of all explanations for the query q .

Example 1. The following probabilistic ProbLog facts represent the correlations (or similarities) among some images reported in the graph. Each fact $\mathbf{p} :: \mathbf{edge}(\mathbf{a}, \mathbf{b})$ denotes the probability \mathbf{p} of seeing the **edge** between **a** and **b** in a randomly sampled world. The last clause asserts that the two images **A** and **B** are similar whether there is an edge from **A** to **B** or viceversa.

```
0.7 :: edge(a,d).
0.9 :: edge(d,b).
0.8 :: edge(b,c).
0.6 :: edge(c,a).
0.5 :: edge(a,e).
0.5 :: edge(e,d).
0.7 :: edge(e,f).
0.6 :: edge(d,f).
```

```
sim(A,B) :-
    edge(A,B) ; edge(B,A).
```



The ProbLog framework includes different inference methods and its implementation is based on the use of tries and reduced ordered binary decision diagrams (ROBDDs). The execution of ProbLog programs uses SLD-resolution to collect all the proofs for a query. In the case of exact inference, for each successful proof of the query, the probabilistic facts used in the proof are gathered and BDDs are used to solve the disjoint sum problem and to obtain the correct probability of the query. In particular, since probability computation and learning in ProbLog are based on propositional logic, compressing Boolean (propositional) functions by BDDs accelerates them. Another solution adopted to overcome the combinatorial explosion of the BDD in ProbLog is to use approximation methods for inference as those proposed in [17, 28].

5 Image Network representation

The methods usually used for the image classification task assume the images as independent and identically distributed. Here, we want to show that adopting a SRL language modelling a complex network may be used to solve the image classification problem providing better results than those obtained with classical methods.

The first step corresponds to choose a representation language for the images. This is generally done by extracting for each image $x \in \mathcal{D}$ some features. In this paper, we describe each image \mathcal{I} with its corresponding correlogram statistics $\gamma_{c_i, c_j}^{(d)}(\mathcal{I})$, computed as reported in Equation 1. After the feature extraction process, a pairwise image comparison approach should be defined. In particular, it is necessary to define a measure able to identify similar images.

Given a set of K classes, denoted as C_k , a discriminant [1] is a function that takes as input an observation $x \in \mathcal{D}$ and classifies it as belonging to one of the K classes. In the general case of $K > 2$ classes, we consider a single K -class discriminant comprising K linear functions ϕ_k and then assigning an observation x to a class C_k if $\phi_k(x) > \phi_j(x)$ for all $j \neq k$.

The well known k-NN technique represents one of the most adopted non-parametric approaches to estimate the class of a given unseen observation. In particular, to classify a new observation, the k-NN method identifies k nearest observations from the training data and then assigns the new observation to the class having the largest number of representatives among this set [1]. The particular case of $k = 1$ corresponds to the *nearest neighbour rule*, where a test observation is simply assigned to the same class as the nearest observation from the training set. In order to compute the nearest neighbours, a distance function between two observations must be used. For instance, the distance between two observations $x, y \in \mathcal{D}$ may be given by $dist(x, y) = (\sum_i |x_i - y_i|^p)^{1/p}$, where x_i and y_i are the components of the representation of the observations x and y . The number of edges in the network built in this way grows as the degree of each node increases. The degree of a node x corresponds to the number of edges connecting similar images to x .

Since we have to build a probabilistic network (edges labelled with probabilities), given a distance function d , returning a value between 0 and 1, it is possible to build the network, where each image is represented as a node, and each edge connecting two similar images x and y is labelled with the value $1 - d(x, y)$.

In a classical approach, where there is not an underlying network structure, computing the similarity between two images x and y corresponds to apply the distance function only to x and y . In this way we consider x and y as independent, ignoring, for instance, the fact that there may be an image z that is similar to both x and y . The idea is to have a network that allows us to exploit all the available connections among the images belonging to the data set.

After having explained how to build the probabilistic network, it is natural to convert it to a ProbLog program as that reported in the Example 1. In particular, each network's edge corresponds to a probabilistic fact $\mathbf{p} : \mathbf{edge}/2$, where \mathbf{p} is

the similarity degree between the two images that are the arguments of the `edge` predicate. Hence, given the ProbLog program, it is possible to solve a query q asking the similarity degree between two nodes by adopting an inference method. In order to exploit the relations among the images, the ProbLog program has been enriched with the following definition.

```

path(A,B,K) :-
    sim(A,B).
path(A,B,K) :-
    K > 1,
    sim(A,C),
    K1 is K - 1,
    path(C,B,K1).

```

The query `path(x,y,k)` is true for all the possible paths, with a length at most equal to k , starting from the node x and ending to the node y . Since the probability of this query is calculated collecting all its proofs (all the possible paths) then it corresponds to the similarity between two nodes computed considering the subnetwork containing the nodes related to those involved in the query. The value of the k parameter expresses the dimension of the considered subnetwork: $k=1$ corresponds to a classical approach.

Example 2. Given T the ProbLog program reported in the Example 1 enriched with the definition of the predicate `path/3`, the probability of the query predicate `path(a,b,1)` is 0, since there are no edges directly connecting a and b . While the success probability of the query `path(a,b,2)` is the sum of the probabilities of all the subprograms containing edges connecting a and b :

$$\begin{aligned}
P_s(\text{path}(a, b, 2)|T) = & \\
& P(\{\text{edge}(a, c), \text{edge}(c, b)\}|T) + \\
& P(\{\text{edge}(a, d), \text{edge}(d, b), \text{edge}(e, f)\}|T) + \\
& P(\{\text{edge}(a, c), \text{edge}(e, d), \text{edge}(c, b)\}|T) + \dots
\end{aligned}$$

6 Experiments

In this section we present some experimental results to assess whether the SRL proposed approach works better than the classical one. In order to do this, the Caltech 101 dataset [9] provided by the California Institute of Technology has been used. We chosen five classes (Airplaines, Faces, Watch, Motorbikes and Leopards) and 30 images have been randomly selected for each class obtaining a dataset of 150 instances. Each image has been represented by the correlograms calculated on given distances corresponding to 1, 3, 5 or 7. Having a correlogram-based representation, the distance between two images, x and y , is computed as follows, as reported in [15]:

$$d_c(x, y) = |x - y| \triangleq \sum_{c_i, c_j \in \mathcal{C}, d \in D} \frac{|\gamma_{c_i, c_j}^{(d)}(x) - \gamma_{c_i, c_j}^{(d)}(y)|}{1 + \gamma_{c_i, c_j}^{(d)}(x) + \gamma_{c_i, c_j}^{(d)}(y)}. \quad (2)$$

Let \mathcal{D} be a set of images labelled with one of the k classes $\mathcal{Q} = \{q_1, \dots, q_k\}$, and $\#q_i = |\{x \in \mathcal{D} | cl(x) = q_i\}|$, where cl is a function returning the class of an image. To rank images with respect to a query image q the similarity is calculated as $1 - d_c(x_i, q)$. As in classical k-NN-based approaches, for each query image q , the error of the ranking result on the first n ranked images is calculated using the following formula

$$err_q^n = \sum_{t=1}^n \mathbf{1}(cl(x_t) \neq cl(q)), \quad (3)$$

where $\mathbf{1}$ is the indicator function that evaluates to 1 if $cl(x_t) \neq cl(q)$ and 0 otherwise, and x_t are the first n elements of the ranked images list with respect to q .

In our experiment $\#q_i = 30$ for each of the five selected classes. Furthermore, the error, computed using the Equation 3, refers to the first 29 (n) ranked images. In particular, when $err_q^n = 0$ means that all the first n images in the ranked images list have the same class of q .

In order to evaluate the SRL based approach, its results are compared to that obtained by the classical method. As already said, the probabilistic network is built using the function reported in Equation 2. We have to decide how much edges, connecting similar images, to insert in the network for each node (image). Given an image x , the α parameter represents the number of the best similar images to x (its neighbours). For each of these neighbours the similarity probability is computed and added to the probabilistic fact `edge/2` in the ProbLog program. In the following experiments a value of α set to 2, 3, 4 or 5 has been used.

Figure 1 shows (on the left) the probabilistic network constructed by setting to 4 the number of neighbours for each image. In the graph all the nodes filled with the same color, from a grayscale, correspond to the same class. A portion of the network (on the right) with the nodes represented by the corresponding image is depicted in the same figure. We can note that although only the four most similar images are considered as neighbourhood, the network complexity is evident. The application of a classical approach neglects all the relationships shown by the network resulting in a less accurate classification result.

The validation of the proposed SRL approach has been done by comparing each image with all the others and ranking the results adopting the ProbLog exact inference. Using the α parameter involves the insertion of only a subset of all the possible connections. Indeed, in the case of few connections, given two image x and y , the ProbLog success probability for `path(x,y,1)` may be 0, because there may be no link between x and y in the network. The absence of the edge between x and y means that y is not present among the first best α similar images (and viceversa). A similar problem concerns the choice of the parameter `k` in the query `path/3`.

In order to solve the problem of the 0 estimated probabilities, due to connections' absence, the error has been computed considering the first images in the ranked images list whose probability is not equal to 0. In particular, let $z < n$

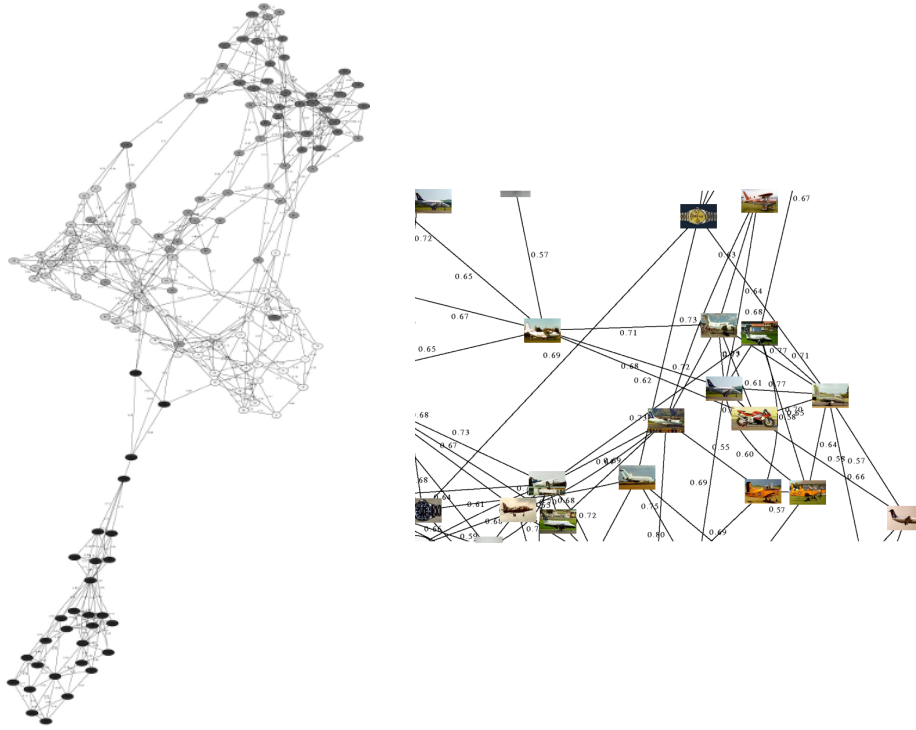


Fig.1. An image network where each image is connected to the four most similar neighbours.

the position in the ranked images list such that all the images x_i , $i \leq z$, have a probability not equal to 0, and $z + 1$ the position in the list of the ranked images corresponding to the first image for which ProbLog provides a success probability equal to 0.

The errors obtained in the first z images has been used to estimate the total error probability over all the $n > z$ images as follows:

$$\widehat{srterr}_q^n = \frac{\frac{k-1}{k}(n-z) + (z \cdot srterr_q^z)}{n}, \quad (4)$$

where

$$srterr_q^z = \sum_{t=1}^z \mathbf{1}(cl(x_t) \neq cl(q)). \quad (5)$$

In particular in Equation 4, $\frac{k-1}{k}(n-z)$ represents the expected number of errors in the remaining $(n-z)$ images, where we assumed that classifications are uniformly distributed among the k classes ($\frac{k-1}{k}$ is the probability to incorrectly classify an image).

Table 1 displays the result obtained on the Airplanes class. In the column headings, d is the distance used for computing the correlogram matrix, α is

the number of neighbors selected for each image to build network, and err_q^n is the number of errors obtained in the first 29 best ranked results with a classical approach. For each value of the k parameter of the `path` query solved by ProbLog, z is the position in the ranked list corresponding to the last image with rank different to 0, $srlerr_q^z$ is the number of error in the first z positions, and \widehat{srlerr}_q^z is the expected number of errors in the first 29 positions. As we can see, the obtained errors with the SRL approach is inversely proportional to both the α and k parameters, and they are fewer than those obtained with the classical method. This result proves the validity of our approach and the advantage of using an SRL language.

Table 1. Results on the Airplanes class

| d | α | err_q^n | k=2 | | | k=3 | | | k=4 | | |
|-----|----------|-----------|-------|--------------|------------------------|-------|--------------|------------------------|-------|--------------|------------------------|
| | | | z | $srlerr_q^z$ | \widehat{srlerr}_q^z | z | $srlerr_q^z$ | \widehat{srlerr}_q^z | z | $srlerr_q^z$ | \widehat{srlerr}_q^z |
| 1 | 0.41 | 2 | 8.40 | 0.10 | 0.60 | 16.67 | 0.19 | 0.45 | 25.43 | 0.27 | 0.34 |
| | | 3 | 12.73 | 0.13 | 0.51 | 25.20 | 0.27 | 0.34 | 29.00 | 0.27 | 0.27 |
| | | 4 | 20.73 | 0.17 | 0.35 | 28.93 | 0.19 | 0.19 | 29.00 | 0.12 | 0.12 |
| | | 5 | 25.10 | 0.22 | 0.30 | 29.00 | 0.19 | 0.19 | 29.00 | 0.16 | 0.16 |
| 3 | 0.41 | 2 | 7.40 | 0.06 | 0.61 | 14.17 | 0.10 | 0.46 | 20.63 | 0.15 | 0.34 |
| | | 3 | 13.90 | 0.14 | 0.48 | 25.90 | 0.22 | 0.28 | 28.90 | 0.21 | 0.21 |
| | | 4 | 18.63 | 0.22 | 0.42 | 28.77 | 0.28 | 0.28 | 29.00 | 0.26 | 0.26 |
| | | 5 | 23.83 | 0.25 | 0.35 | 29.00 | 0.25 | 0.25 | 29.00 | 0.22 | 0.22 |
| 5 | 0.42 | 2 | 7.73 | 0.06 | 0.60 | 14.27 | 0.12 | 0.47 | 22.17 | 0.21 | 0.35 |
| | | 3 | 12.43 | 0.13 | 0.51 | 24.30 | 0.24 | 0.33 | 28.97 | 0.27 | 0.27 |
| | | 4 | 19.37 | 0.18 | 0.38 | 28.90 | 0.21 | 0.21 | 29.00 | 0.18 | 0.18 |
| | | 5 | 24.70 | 0.24 | 0.32 | 29.00 | 0.22 | 0.22 | 29.00 | 0.20 | 0.20 |
| 7 | 0.41 | 2 | 7.27 | 0.07 | 0.62 | 12.97 | 0.13 | 0.50 | 19.17 | 0.24 | 0.43 |
| | | 3 | 12.87 | 0.13 | 0.50 | 24.73 | 0.27 | 0.34 | 28.87 | 0.29 | 0.29 |
| | | 4 | 20.97 | 0.21 | 0.38 | 29.00 | 0.22 | 0.22 | 29.00 | 0.16 | 0.16 |
| | | 5 | 25.27 | 0.28 | 0.35 | 29.00 | 0.24 | 0.24 | 29.00 | 0.19 | 0.19 |

Table 2 shows the results obtained on the other classes: Faces, Watch, Motorbikes and Leopard, where we take fixed the parameter k to 4. The errors obtained with other classes confirm the observations made on the class Airplanes. The computational time of the proposed approach is very high when compared to a classical k -NN. This is due to the ProbLog inference procedure that represents its bottleneck.

7 Conclusion

In this paper we considered the image classification problem. In particular our goal was to see if it is possible to increase the classification accuracy obtained using a classical k -NN approach, where the relationships among the images are ignored. In order to capture the connections among images we have used a SRL

Table 2. Results on the Faces, Watch, Motorbikes and Leopard classes with $k=4$.

| d | α | Faces | | | | Watch | | | | Motorbikes | | | | Leopard | | | | |
|-----|----------|-----------|------|------------|----------------------|-------------|------|-------------|----------------------|------------|-------------|-------------|----------------------|-----------|-------------|-------------|----------------------|--|
| | | err_q^n | z | $sert_q^n$ | \widehat{sert}_q^n | err_q^n | z | $sert_q^n$ | \widehat{sert}_q^n | err_q^n | z | $sert_q^n$ | \widehat{sert}_q^n | err_q^n | z | $sert_q^n$ | \widehat{sert}_q^n | |
| 1 | 2 | | 26.5 | 0.28 | 0.33 | 25.7 | 0.32 | 0.64 | 26.1 | 0.27 | 0.32 | 23.7 | 0.09 | 0.22 | | | | |
| | 3 | 0.33 | 29.0 | 0.25 | 0.25 | 29.0 | 0.67 | 0.67 | 29.0 | 0.24 | 0.24 | 0.13 | 27.6 | 0.07 | 0.10 | | | |
| | 4 | | 29.0 | 0.20 | 0.20 | 29.0 | 0.71 | 0.71 | 29.0 | 0.21 | 0.21 | 0.26 | 28.2 | 0.09 | 0.11 | | | |
| | 5 | | 29.0 | 0.17 | 0.17 | 29.0 | 0.73 | 0.73 | 29.0 | 0.21 | 0.21 | 0.26 | 28.9 | 0.09 | 0.09 | | | |
| | | 2 | 27.4 | 0.25 | 0.28 | 23.6 | 0.57 | 0.61 | 24.8 | 0.22 | 0.30 | 21.8 | 0.11 | 0.28 | | | | |
| 3 | 3 | 0.27 | 29.0 | 0.21 | 0.21 | 28.8 | 0.64 | 0.65 | 29.0 | 0.19 | 0.19 | 0.14 | 27.3 | 0.14 | 0.15 | | | |
| | 4 | | 29.0 | 0.18 | 0.18 | 29.0 | 0.70 | 0.70 | 29.0 | 0.17 | 0.17 | 0.25 | 28.7 | 0.25 | 0.25 | | | |
| | 5 | | 29.0 | 0.17 | 0.17 | 29.0 | 0.77 | 0.77 | 29.0 | 0.18 | 0.18 | 0.25 | 28.9 | 0.15 | 0.15 | | | |
| | | 2 | 26.4 | 0.29 | 0.33 | 24.5 | 0.51 | 0.55 | 25.3 | 0.23 | 0.30 | 28.8 | 0.10 | 0.30 | | | | |
| | | 3 | 0.25 | 29.0 | 0.21 | 0.21 | 29.0 | 0.64 | 0.64 | 29.0 | 0.22 | 0.22 | 0.15 | 27.2 | 0.12 | 0.16 | | |
| 5 | 4 | | 29.0 | 0.18 | 0.18 | 29.0 | 0.69 | 0.69 | 29.0 | 0.19 | 0.19 | 0.23 | 27.9 | 0.12 | 0.14 | | | |
| | 5 | | 29.0 | 0.13 | 0.13 | 29.0 | 0.76 | 0.76 | 29.0 | 0.16 | 0.16 | 0.23 | 28.9 | 0.15 | 0.15 | | | |
| | | 2 | 27.4 | 0.28 | 0.31 | 24.9 | 0.56 | 0.60 | 27.6 | 0.26 | 0.29 | 19.0 | 0.11 | 0.35 | | | | |
| | | 3 | | 29.0 | 0.18 | 0.18 | 29.0 | 0.69 | 0.69 | 29.0 | 0.19 | 0.19 | 0.22 | 27.5 | 0.13 | 0.16 | | |
| | | 4 | 0.25 | 29.0 | 0.14 | 0.14 | 29.0 | 0.74 | 0.74 | 29.0 | 0.18 | 0.18 | 0.22 | 28.1 | 0.13 | 0.15 | | |
| 7 | 5 | | 29.0 | 0.11 | 0.11 | 29.0 | 0.77 | 0.77 | 29.0 | 0.16 | 0.16 | 0.22 | 28.2 | 0.13 | 0.15 | | | |

language. Connections between two images have been weighted with a probability. Experiments have been conducted on a real-word data set comparing the classical method with respect to our presented approach. The results show the validity of the proposed method.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer-Verlag (2006)
2. Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J.: Latent dirichlet allocation. *Journal of Machine Learning Research* 3 (2003)
3. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. *Computer Vision and Pattern Recognition* 0, 1–8 (2008)
4. Burl, M.C., Weber, M., Perona, P.: A probabilistic approach to object recognition using local photometry and global geometry. In: *Proceedings of the 5th European Conference on Computer Vision-Volume II - Volume II*. pp. 628–641. ECCV '98, Springer-Verlag (1998)
5. Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Soergel, D., Agosti, M., Dobрева, M., Katifori, V., Schuldt, H.: The DELOS Digital Library Reference Model. *Foundations for Digital Libraries* (2007)
6. Chakrabarti, S., Dom, B., Indyk, P.: Enhanced hypertext categorization using hyperlinks. In: *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*. pp. 307–318. ACM (1998)
7. Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. *IJPRAI* pp. 265–298 (2004)
8. Costa, V.S., Page, D., Qazi, M., Cussens, J.: CLP(BN): Constraint logic programming for probabilistic knowledge. In: *Proc. UAI-2003*. pp. 517–524. Morgan Kaufmann (2003)
9. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories (2004)
10. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 264–271 (2003)
11. Getoor, L., Taskar, B.: *Introduction to Statistical Relational Learning*. The MIT Press (2007)
12. Getoor, L.: Link-based classification. In: Maulik, U., Holder, L., Cook, D. (eds.) *Advanced Methods for Knowledge Discovery from Complex Data*. Springer-Verlag (2005)
13. Getoor, L., Diehl, C.P.: Link mining: a survey. *SIGKDD Explorations Newsletter* 7, 3–12 (2005)
14. Hofmann, T.: Probabilistic latent semantic analysis. In: *Proceeding of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (1995)
15. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 762–768. IEEE (1997)
16. Kameya, Y., Ueda, N., Sato, T.: A graphical method for parameter learning of symbolic-statistical models. In: *Proceedings of the Second International Conference on Discovery Science*. pp. 264–276. Springer-Verlag (1999)

17. Kimmig, A., Santos Costa, V., Rocha, R., Demoen, B., Raedt, L.: On the efficient execution of problog programs. In: Proceedings of the 24th International Conference on Logic Programming. pp. 175–189. Springer-Verlag (2008)
18. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press (2009)
19. Lu, Q., Getoor, L.: Link-based classification. In: Proceedings of the International Conference on Machine Learning. pp. 496–503 (2003)
20. Macskassy, S.A., Provost, F.: A simple relational classifier. In: Proceedings of the Second Workshop on Multi-Relational Data Mining. pp. 64–76 (2003)
21. Macskassy, S.A., Provost, F.: Classification in networked data: A toolkit and a univariate case study. *Journal of Machine Learning Research* 8, 935–983 (2007)
22. Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. In: In Proc. ECCV. pp. 490–503. Springer (2006)
23. Poole, D.: Abducing through negation as failure: stable models within the independent choice logic. *J. Log. Program.* 44(1-3), 5–35 (2000)
24. Raedt, L.D., Kimmig, A., Toivonen, H.: Problog: a probabilistic prolog and its application in link discovery. In: Proceedings of 20th International Joint Conference on Artificial Intelligence. pp. 2468–2473. AAAI Press (2007)
25. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. *Commun. ACM* 18, 613–620 (November 1975)
26. Sato, T.: A statistical learning method for logic programs with distribution semantics. In: Proceedings of the 12th International Conference on Logic Programming. pp. 715–729. MIT Press (1995)
27. Sen, P., Namata, G.M., Bilgic, M., Getoor, L., Gallagher, B., Eliassi-Rad, T.: Collective classification in network data. *AI Magazine* 29(3), 93–106 (2008)
28. Shterionov, D., Kimmig, A., Mantadelis, T., Janssens, G.: DNF sampling for problog inference. In: International Colloquium on Implementation of Constraint and Logic Programming Systems 2010 (2010)
29. Taranto, C., Di Mauro, N., Ferilli, S., Esposito, F.: Approximate image color correlations. In: Proceedings of the international conference on Multimedia. pp. 1127–1130. MM '10, ACM (2010)