Using Gaussian Processes to Improve Zero-Shot Learning with Relative Attributes

Yeshi Dolma, Vinay P. Namboodiri

Department of Computer Science and Engineering,
Indian Institute of Technology, Kanpur

Abstract. Relative attributes can serve as a very useful method for zero-shot learning of images. This was shown by the work of Parikh and Grauman [1] where an image is expressed in terms of attributes that are relatively specified between different class pairs. However, for zero-shot learning the authors had assumed a simple Gaussian Mixture Model (GMM) that used the GMM based clustering to obtain the label for an unknown target test example. In this paper, we contribute a principled approach that uses Gaussian Process based classification to obtain the posterior probability for each sample of an unknown target class, in terms of Gaussian process classification and regression for nearest sample images. We analyse different variants of this approach and show that such a principled approach yields improved performance and a better understanding in terms of probabilistic estimates. The method is evaluated on standard Pubfig and Shoes with Attributes benchmarks.

1 Introduction

Consider the task of recognizing a person at test time when we are not provided with any images of the person at training. This setting for classification is termed zero-shot learning, i.e. the classifier is provided with no training image for obtaining the classification. A technique used to recognize unseen classes is through the use of attributes [5]. These attributes describe a person in terms as the gender of a person, or type of hair that person has. However, as shown by [1], a more natural description is obtained by describing the attributes of a person in relation to those that are known. For instance, we can say that 'Tracy Morgan's face is chubbier as compared to 'Anderson Cooper' but less as compared to 'Karl Rove'.

In this paper, we consider this problem of zero-shot recognition of different objects like faces or shoes using relative attributes. The initial work by [1] used relative attributes in zero-shot recognition by using a Gaussian mixture model of the relative attributes. However, a simple Gaussian mixture model does not transfer the knowledge effectively in the model. In this paper, we propose a more principled approach where we use a Gaussian Process prior over the relative attributes in order to obtain zero-shot recognition. This approach while being principled also enables us to model the variance in the samples. We further analyze different variants of using Gaussian process prior for obtaining zero-shot recognition of samples.

In our approach we use two stages of Gaussian processes. In the first stage, we use a Gaussian process based classifier to classify the set of classes that are known in training. In the second stage, we use Gaussian process based regression to obtain the zero shot recognition for samples in test that have no training examples. The two stages allow for effective knowledge transfer from known training samples of a fixed set of categories to unknown test samples of a set of categories for which no training samples are present.

The main contribution of this work is to demonstrate a two-stage framework using Gaussian process that allows us to obtain principled probabilistic estimates of the relative attributes for zero shot learning. We obtain in this framework not only the probablistic estimates of $p(y|x)$ where y is the class label and x is the feature set, but also the uncertainty in estimating $p(y|x)$ that is extremely relevant in the zero-shot setting. We demonstrate the efficacy of our method with detailed comparison to the previous work [1] on standard benchmark datasets.

The rest of the paper is organized as follows: In the next section we give a brief overview of the related work. In section 3 we provide the background that briefly provides an overview of the relative attribute zero shot learning based setting. In section 4 we provide detailed description of the proposed method and its variants. Section 5 discusses the experiments performed and the results obtained from the experiments and we finally conclude in section 6 with directions for future work.

2 Related Works

The use of attributes for zero shot learning was initially proposed by [5]. In their work they had shown that animals could be described in terms of binary attribute vectors that captured the properties of each class. This was then used to recognize an unseen class in terms of its attributes. [7] extend the work by considering the attribute representation problem as one of label embedding and learn the embedding instead of using a direct attribute presentation [6]. Further work has been undertaken where they consider that the attributes may be unreliable [8]. Another interesting line of work has been analysed by [10] where the authors analysed the use of pure textual descriptions instead of well defined attribute representations. A recent work explores the structure of the semantic manifold in terms of semantic class label graph for representing the distance [14]. Another explores the co-occurrence of visual concepts for zero shot classification [15].

These methods have addressed the attribute representation. However, in our work we address the method used for zero-shot recognition. The basic premise is that just using a clustering would not exploit the structure of the data for zero-shot recognition. Recently there has been interesting work by [9] where the authors show that using Bayesian local learning they are able to analyse when two images are indistinguishable for a specific attribute. In our work we jointly rely on multiple attributes and treat the problem of identifying the sample through Gaussian process regression.

The present work relies on relative attributes which were proposed by [1]. In their work the authors introduced relative attributes and showed that they were applicable for a number of use-cases including zero-shot learning of unseen classes. Further, [4] have shown that relative attributes could be coupled with relative feedback and this would be useful for image search cases such as searching for a shoe. These use-cases that extend relative attributes could also be applicable using the proposed method.

Gaussian process is extensively used in our work. This framework has been excellently presented by [2] in their book. This approach while primarily suited for regression has also been used for other related tasks such as multi-relational learning [11] and for one-shot recognition [12]. In our approach we use it in a two stage approach for classification and regression based on attribute data for zero-shot learning.

3 Background

Our method builds on the work of Devi Parikh and Grauman [1] where the classes are modelled as Gaussian Distributions using *relative attributes*, which depict the strength of an attribute as opposed to binary attributes which shows its presence or the absence in the image.

During training, given a set of training images X represented by N-dimensional feature vector, $x_i \in \mathbb{R}^N$, and a set of M attributes, A_m, the relation between the attribute strength of the seen classes are given as sets of ordered pair $O_m = \{(i,j)\}$ and similarity-pair $S_m = \{(i,j)\}$. These pairs are such that if $(i,j) \in O_m$, then image i has stronger attribute a_m than image j. Similarly, if the pair $(i,j) \in S_m$, image i and image j have similar strength of attribute a_m. Using these pairs as supervision, M ranking functions are learned for each attribute that maps an image to its attribute strength score. These functions transform the images $x_i \in \mathbb{R}^N \implies \mathbb{R}^M$. The images are now M-dimensional vector where mth dimension represents the attribute a_m's rank score. For the unseen classes, the supervision is given with respect to one or two seen classes. An unseen class c^u_k can be described relative to seen classes c^s_p and c^s_q, using all or a subset of M attributes, as $c^s_{pm} \prec c^u_{km} \prec c^s_{qm}$ or $c^s_{pm} \prec c^u_{km}$, or $c^u_{km} \prec c^s_{qm}$, where the unseen class c^u_k has mth attribute stronger than class c^s_p but weaker than class c^u_q.

Now given a novel image j to be classified into one of the seen or unseen classes, a generative model of all the seen classes in $\mathbb{R}^M$ is built. A seen class c^s_p is represented by a Guassian distribution $c^s_p \sim \mathcal{N}(\mu^s_p, \Sigma^s_p)$ where mean is $\mu^s_p \in \mathbb{R}^M$ and Σ^s_p is $M \times M$ covariance matrix. The parameters of the generative model of the unseen classes U are described relative to the parameters of the seen classes, built according to the supervision given. For attribute a_m, if an unseen class c^u_k is described as $c^s_p \prec c^u_k \prec c^s_q$, the mth component of the mean of unseen class μ^u_{km} is set as $\frac{1}{2}\big(\mu^s_{pm} + \mu^s_{qm}\big)$. Similarly for the unseen classes described relative to just one seen class as $c^s_p \prec c^u_k$ or $c^u_k \prec c^s_q$, μ^u_{km} is described as $\mu^s_{pm} + d_m$ or $\mu^s_{qm} - d_m$ respectively, where d_m is the average of the distances between the sorted mean rank scores of seen classes for the mth attribute and the covariance Σ^u_k is $\frac{1}{S}\sum_{i=1}^{S} \sigma^s_i$.

Finally, maximum likelihood is computed and the test image j is assigned the label with the highest likelihood of a seen or an unseen class.

$$c^* = \underset{j \in \{1,..,N\}}{\arg\max}\, \mathcal{P}\big(\tilde{x}_i | \mu_j, \Sigma_j\big) \tag{1}$$

The description of the unseen classes as simply the mean of the related seen classes may not best represent the unseen class and hence a more accurate approach is proposed to represent the unseen class for recognition.

4 Approach

In this section, we first explain our approach to improve zero-shot recognition using Gaussian Processes by providing more accurate and systematic framework to describe the images of the unseen class. Second, we describe in Section 4.2, Gaussian-process based classifier for the seen classes and then, in Section 4.3, Gaussian Process (GP) based method that improves the accuracy of recognition for the unseen class using k-nearest training images. In Section 4.4, we explain a variant of our method that relies on multiple versions of distributions. This method is however subsumed in terms of performance by the GP-kNN algorithm.

4.1 Gaussian Processes for Zero-Shot Recognition

Gaussian Process is a distribution of random variables such that any finite number of distribution of these variables is jointly Gaussian. The observations in the process occur in a continuous domain. Any Gaussian process $f(x)$ can be specified as

$$f(x) \sim \mathcal{GP}\big(m(x), k(x^T, x)\big) \tag{2}$$

where the process's mean function and the covariance funtion are respectively:

$$m(x) = \mathbb{E}[f(x)], \quad k(x^T, x) = \mathbb{E}\big[\big(f(x)m(x)\big)\big(f(x)m(x)\big)\big]. \tag{3}$$

Let a regression model with Gaussian noise be given as

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{w}, \quad y = f(\mathbf{x}) + \mathcal{N}\big(0, \sigma_n^2\big) \tag{4}$$

where $\mathbf{x}$ is the input vector, $\mathbf{w}$ is the vector of weights (parameters) of the model and f is the function value. The outcome observed is represented by y, assuming that the additional noise term is an independent zero-mean Gaussian distribution. We assume a zero-mean Gaussian prior $w \sim \mathcal{N}\big(0, \Sigma_p\big)$. Given the model and the noise assumption, the *likelihood* and the *posterior*, given by combining the prior with the likelihood using the Bayes' rule, are respectively as follows.

$$\mathcal{P}(\mathbf{y}|X,\mathbf{w}) = \mathcal{N}(X^{\mathbf{w}}, \sigma_n^2 I) \tag{5}$$

$$\mathcal{P}(\mathbf{y}|X) = \int \mathcal{P}(\mathbf{y}|X,\mathbf{w})\mathcal{P}(\mathbf{w})d\mathbf{w} \tag{6}$$

Finally the predictive outcome f_* at x_* is given by

$$\mathcal{P}(f_*|\mathbf{x}_*, X, \mathbf{y}) = \mathcal{N}\big(\frac{1}{\sigma_n^2}\mathbf{x}_*^T A^{-1} X\mathbf{y}, \mathbf{x}_*^T A^{-1}\mathbf{x}_*\big) \tag{7}$$

Further details of the full Bayesian treatment for Gaussian process is presented by Rasmussen and Williams [2].

Our two-tier method uses Gaussian process (GP) based classifier in the first step and Gaussian process regression for a more accurate description of unseen class in the second step. In the first step, for each test image j, if the GP-based classifier outputs a prediction greater than a certain set threshold τ, the classifier corresponding to a seen-class c_p^s labels image j as 'class-p'. This takes care of those test images which are very similar to a seen class's training images, thus suggesting that the target unknown-class has higher probability to be one of the seen classes. The GP-based classifier for the seen classes is explained in subsection 4.2.

In the second step, for a test image j which is not labeled by any of the GP-classifiers of the first step, new Gaussian models representing the unseen classes are created by modeling more accurate description of the attribute value of the unseen class based on k sample images chosen from the training set which are nearest to the test image j. These new distributions are also taken into account, along with their initial Gaussian distribution, to represent the unseen classes. Based on the maximum likelihood computed for all the distributions the final label is assigned. The method is explained further in the following subsections.

4.2 Gaussian Process based classifier

During training, we are given a set of training images X belonging to S number of seen classes and a set of attributes, A_m. These training images are represented by $\mathbb{R}^N$ feature vector. Using the supervision given for the relative attributes between these seen classes, a ranking function is learnt which transforms the $\mathbb{R}^N$ image feature vector to $\mathbb{R}^M$ vector in attribute-space.

For all the training images j, Mahalanobis distance of the image from every seen class c_p^s is computed. This distance shows how many standard deviations away an image j is from a seen class. The distance comes out smaller for images similar to the seen class, according to the attributes, and larger for images that are dissimilar. By taking the average of these distances, Mahalanobis distance is calculated for each pair of seen classes.

For every seen class c_p^s, a Gaussian Process classifier is created, in the attribute space, with the training images from c_p^s and c_q^s as positive and negative samples, where class c^s_p and c^s_q are most distant from each other. The GPML tool box [13] is used for the computation.

These Gaussian process classifiers, each corresponding to a seen class, are used to find the posterior mean given the test image as the input. If the posterior mean of the prediction is greater than the set threshold τ, (experimentally set to 0.9), the test image j is labelled positive by the classifier. In case more than one classifier labels an image positive, the label by the classifier with a more positive mean is assigned.

4.3 Zero-shot Recognition using Gaussian Process - *k*NN Approach

In the previous approach, given a generative model for all the classes, each class is represented by a Gaussian distribution. The unseen classes are modeled using supervision given for all or a subset of M attributes (see Section 3). Every class-p, seen and unseen, has a set of parameters corresponding to the mean μ_p and the covariance Σ_p of the class. The label is assigned to the test image based on the highest likelihood value computed for each of the classes.

In our proposed approach to improve zero-shot recognition, for all the test images which are not labelled by any of the seen-classes' GP-based classifier, Gaussian process is used to improve the recognition in the following way as is shown in figure 1.

1. From the set of training images, k-nearest samples are chosen whose Euclidean distance is shortest from the test image j. These k images resemble the test image most closely, in the attribute space. (See example in the Figure 2 for two test samples- Michelle Wie and Ben Stiller).
2. For every unseen class c^u_o, for an attribute a_m, if the supervision is given with respect to two seen classes c^s_p and c^s_q as $c^s_{pm} \prec c^u_{om} \prec c^s_{qm}$, then the mth component of the mean of the unseen class, μ^u_{om} is computed using Guassian process (GP) and the k nearest neighbours.
   The unseen class is represented by a set of k means and covariances, $(\mu^{iu}_o, \Sigma^{iu}_o)$, $i \in \{1, ..k\}$. A GP is created with the rank scores of the mth attribute of the training images from seen classes c^s_p and c^s_q as positive and negative training samples respectively. Now, the mth component in each of the μ^{iu}_o is the posterior prediction mean output, with the mth attribute rank score of the ith-nearest training samples (chosen in Step 1) $i \in \{1, ..k\}$ as input to the above constructed GP.
3. For the attribute whose supervision is given with respect to just one seen class, as $c^s_{pm} \prec c^u_{om}$ or $c^u_{om} \prec c^s_{qm}$, the mth component of the mean of the unseen class is taken as $\mu^s_{pm} + d_m$ or $\mu^s_{qm} - d_m$ respectively. Here d_m is the average of the distances between the sorted mean rank scores of seen classes for the mth attribute.
4. To assign label to the test image, the likelihood score is computed by $\mathcal{P}\big(\tilde{x}_j|\mu_i, \Sigma_i\big)$, where μ_i and Σ_i is the mean and covariance of all the classes, including the k new sets of $(\mu^{iu}_o, \Sigma^{iu}_o)$ constructed for the unseen classes in the previous step. The label is finally assigned based on the maximum likelihood value.

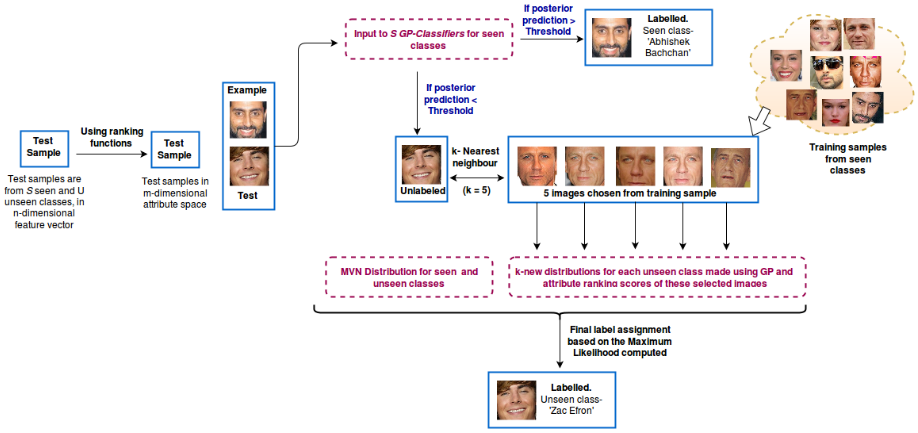
Test Sample
Test samples are from *S* seen and U unseen classes, in n-dimensional feature vector
Using ranking functions
Test Sample
Test samples in m-dimensional attribute space
Example
Test
Input to *S GP-Classifiers* for seen classes
If posterior prediction > Threshold
Labelled. Seen class- 'Abhishek Bachchan'
If posterior prediction < Threshold
Unlabeled
k- Nearest neighbour
(k = 5)
5 images chosen from training sample
Training samples from seen classes
MVN Distribution for seen and unseen classes
k-new distributions for each unseen class made using GP and attribute rankina scores of these selected imaaes
Final label assignment based on the Maximum Likelihood computed
Labelled. Unseen class- 'Zac Efron'

Fig. 1. *Basic outline of the proposed GP based method for Zero-shot recognition.* Test image in n-dimensional feature space is first transformed to m-dimensional attribute space using the ranking function learned for each attribute. These images are then given as an input to be labelled by GP-based classifiers for the seen classes, determined by a threshold for the predicted posterior. k-training samples from seen classes are then chosen according to their euclidean distance from the unlabeled test samples. Using Gaussian process, explained in Section 4.3, and the attribute rank scores of these chosen images to the GP, multivariate normal distributions (MVN) are modelled to represent the unseen class more accurately. The label corresponding to the distribution which gives the maximum likelihood, is assigned to the image.

4.4 Tray of Multivariate Normal Distributions - a variant of our proposed method

We also experimented with a variant of our proposed GP-kNN method, and studied its performance in a subset of PubFig dataset. In this step, for all those test images which are not labelled by any of the GP-classifiers from the first step (Section 4.2), likelihood of the image belonging to each class is computed. If the likelihood of the test image to belong to a seen-class is highest, the label is assigned to it accordingly. However, if the likelihood of the test image to belong to one of the unseen classes is highest, instead of one set of mean μ_i and covariance Σ_i, multiple sets or a 'tray' of mean and covariances representing that class is dynamically created as we come across test samples. The image is labeled accordingly and a new distribution (μ'_i,Σ_i), where the mth component of μ'_i is the posterior mean predicted with the test image's mth attribute score as input, is added to the tray. For subsequent test images, the likelihood for labeling, will be computed using all the earlier distributions representing the classes and those which are added to the tray.

In this approach rather than using GP regression, we had considered dynamic updation of the multi-variate normal distribution for the unseen classes. Keeping a dynamically increasing tray of multivariate normal distributions to compute the likelihood and assign label to the test image, accomodates the idea that a labeled test sample may improve the description of the unseen class, for the following test images, than the original gaussian mixture model. However, improvement by this method is dependent on the order of test images which led to the development of more systematic algorithm (GP-kNN) for the unseen classes' description. Moreover, as shown in section 5.4, this method does not perform as well as the GP kNN regression method.

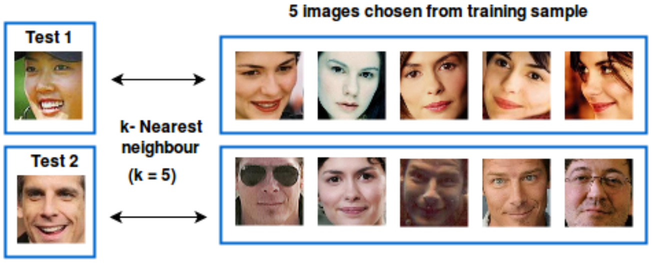

Fig. 2. k-nearest neighbours computed for two unlabelled test samples: Michelle Wie and Ben Stiller. From the training set of 5 seen classes, k-nearest neighbour (k=5) based on the Euclidean distance from the test image is seen. The neighbors selected depends on attributes. The shape of face and age is similar for the nearest neighbors in this example.

5 Experiments

We evaluate our method for zero-shot recognition using GP-based classifier and k-nearest neighbors and compare our accuracy rate with the results obtained by GMM based clustering, as in the work of Parikh and Grauman [1]. We report results to demonstrate a more systematic and accurate description of the unseen class and validate the improvement achieved in recognition.

5.1 Setup

Our experiments used two datasets: a subset of **Public Figure Face Database** (PubFig) [3] and **Shoes with Attributes Dataset** [4]. The PubFig dataset consists of images of 60 different personalities, each image being represented by a 73-dimensional feature vector. Four sets of experiments were done on this dataset to validate our method where in each set, 8-10 classes of people are randomly chosen. The effect of changing the number of seen classes, the number of attributes to describe the classes and varying the supervision in terms of 10 different relative attributes is also demonstrated.

The experiment on Shoes with Attributes Dataset is done by taking 8 classes of shoes which are visibly distinct from each other, in terms of 10 relative attributes. The effect of varying supervision in terms of the number of classes seen is also presented. The images are represented as concatenation of the 960-dimensional gist descriptor with 30-dimensional color histogram image features. The feature vector was chosen to be same as the relative attributes work [1] to which it is being compared.

5.2 Zero-shot Learning Results

Results of PubFig Dataset: Four sets of experiment are done on this dataset consisting of randomly chosen classes and 10 relative-attributes. Table 1 shows in detail the classes that were randomly chosen, the attributes taken into consideration and the partial ordering of the subset of relative attributes given as supervision for the unseen classes, in one of the experiments. (For example supervision '(8) : $J \prec S \prec H$' means that Scarlett Johansson has narrower eyes than Hugh Laurie and Jared Lato has narrower eyes than Scarlett Johansson). In figure 3 we show some examples where our proposed method does better than the GMM based. The green labels are correct labels assigned by our GP-based method and labels in red are the incorrect labels. In an example, for a test sample of class 'Miley Cyrus', both of the methods fail as the relative attribute supervsion given is not sufficient to distinguish it from the class 'Alyssa Milano'.

By varying supervision in terms of attributes to relate classes, our method follows a general trend of increasing accuracy rate with increase in the number of seen classes. This is not only because with greater number of seen class the supervision is more elaborate but also because as the number of seen classes increases, the number of test images that are labeled correctly in our first step by the GP-based classifiers also increases.

Secondly, our GP-based method, using the k-sample images nearest to the test image, provides a more accurate description of the unseen class as opposed to Gaussian mixture model of the classes where the unseen class is described as means of the seen classes. This can be clearly seen as our method outperforms the GMM based recognition. 120-150 test images uniformly belonging to each of the seen and unseen classes, are randomly taken for evaluation. The graphs below shows the accuracy curve obtained by GP-based method vs. GMM-based method.

Graph 1 (top-left) and Graph 2 (top-right) presents the performance curve of our proposed method vs the GMM based method. For 10 classes (seen and unseen), 10 attributes are used to relatively describe the classes for learning the ranking function and a subset of these attributes for unseen classes' supervision. The classes and the set of attributes vary for both the experiments. The classes are randomly selected and the attributes are such chosen that they are capable of representing these classes and vary well among the classes to make them distinct. To study the effect of supervision in terms of the proportion of seen classes, the number of seen classes were varied from 4 to 10, keeping the total number of classes same. It is seen that as we see more number of classes, the overall accuracy percentage increases for a test set of 150 images as the unseen classes can be related to more number of seen classes to make itself more distinguishable. The testset consists of randomly selected images, uniformly belonging to each of the classes.

Graph 3 (bottom-left) and Graph 4 (bottom-right) validates the performance of our method in the same way. Here, the 8 classes were randomly chosen and were represented by 10 relative attributes for both of the experiments. The proportion of seen classes were varied from 4 to 8 (all seen) and an increasing graph for accuracy in the recognition is observed. The test set consists of 120 randomly selected images, uniformly belonging to each of the seen and unseen classes.

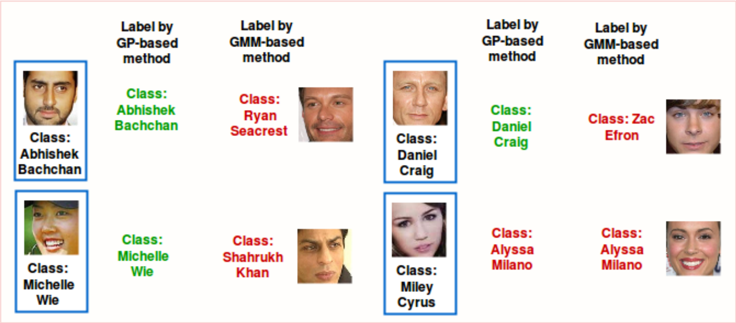

Fig. 3. The figure shows some examples of Prediction using our GP-based method and GMM-based method. The color green shows the correct prediction and label in color red shows the incorrect prediction.

Results of Shoes with Attributes Dataset: In the experiment to evaluate our method in shoes with attribute dataset, 8 distinct classes of the dataset with 6 attributes relating them were chosen. The relative attribute supervision is similar to that provided in the previous experiment. In figure 5 we show examples where our proposed method does better labeling than the GMM based method. The labels in green are correct labels for the test samples, assigned by our GP-based method and labels in red are the incorrect labels. For test sample of Rainboots, using the relative atributes chosen, it was difficult to distinguish ‘rainboots’ from ‘boots’.

The performance result obtained in this dataset is very similar to the one obtained with the PubFig dataset. The classes in this dataset are chosen such that they can be humanly perceived as distinct from each other without confusion (*e.g.* keeping only ‘Athletic shoes’ and not -both Sneakers and Athletic shoes and keeping ‘pumps’ instead of both pumps and high-heels). The accuracy of our method increases as we increase the number of seen classes and outperforms the GMM-based method. In the graph of Figure 5.2, the proportion of seen classes are varied keeping the total classes same.

5.3 Varying the number of attributes

Variation in the performance by varying the number of attributes to describe the seen and the unseen classes is seen. For a PubFig dataset consisting of 8 classes (5 seen and 3 unseen), the number of attributes used to describe these

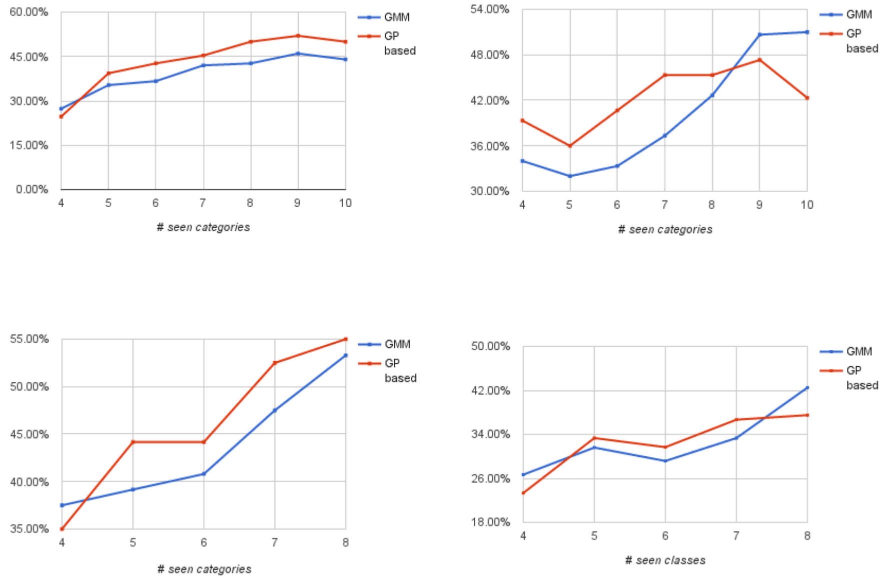

Fig. 4. *Performance curve for experiment with PubFig Dataset.* The accuracy rate is presented for four different sets of experiments done on PubFig and changes in the accuracy for recognition as the proportion of seen classes is varied.

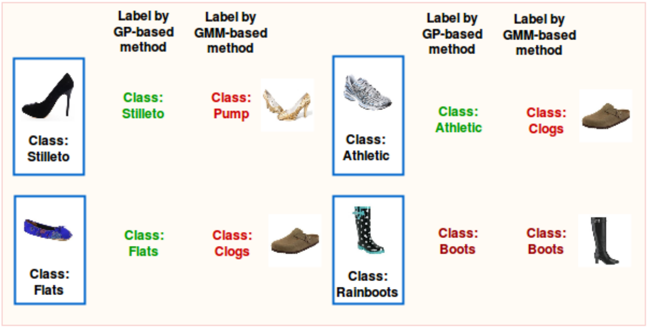

Fig. 5. The figure shows some examples of Prediction using our GP-based method and GMM-based method. The *green* color shows the correct prediction and label in color *red* shows the incorrect prediction.

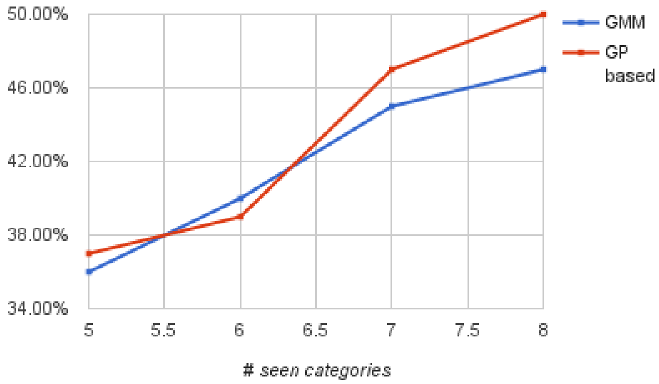

Fig. 6. Performance curve evaluated on Shoes with Attribute Dataset with 8 different categories of shoes represented by 6 relatively defined attributes. The accuracy of recognition increases as the number of seen classes increases from left to right. The accuracy is compared to GMM based method for recognition. The test set consisted of 100 images randomly chosen and belonging to all the classes.

classes relatively, were varied. In the graph of Figure 5.3, number of attributes to describe the classes are varied in the x-axis from 6 to 11. It is seen that greater the number of relative attributes learned to represent a class, the more descriptive it is of the class and hence the recognition rate increases. Our proposed GP-based method outperforms the GMM-based method for the recognition. The test set consisted of 120 images randomly chosen and uniformly belonging to all the classes.

5.4 Comparing performance of various methods for Zero-shot learning

Performance of proposed GP-based method is compared to GMM based method and MVN-tray method (See Section 4.4). The curve in Figure 5.4, shows the accuracy achieved by different methods on 6 classes of PubFig dataset. The classes were chosen at random and 7 relative attributes were used to describe the classes. From left to right, while Gaussian Mixture Model (GMM) achieves an accuracy of 56.60% , a variant of our method of keeping a dynamically increasing tray of the mutivariate normal (MVN) distribution for each unseen class, as more test samples are seen, improves upon it. In this case, when more than one seen classes' classifier gives a positive output in the first step of our algorithm, the test image is not assigned any label.

Slight modification is done to this MVN-Tray method which improves the accuracy further. In case of a tie between two classifiers which outputs a positive

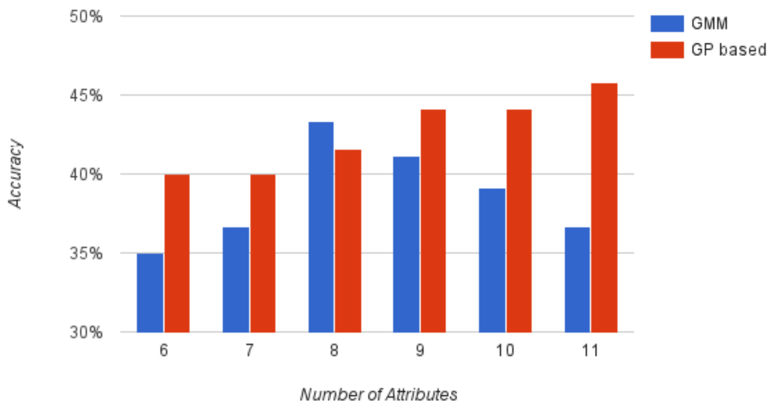

Fig. 7. The graph shows performance of our proposed method vs GMM-based method, as the number of attributes to describe the classes is varied. The setup is 8 randomly chosen classes from PubFig dataset with 5 seen and 3 unseen classes. The x-axis shows the number of attributes used to model a class.

prediction for the test image, label is assigned to the test image by the classifier with more positive prediction posterior as opposed to MVN-Tray where no label is assigned in such a case. This variant of MVN-Tray method is named as 'MVN-Tray-Modified' in the figure. Finally, our proposed algorithm (GP-kNN) presents a more principled method using Gaussian process with k-nearest sample images, to improve the recognition of test images belonging to the seen classes, using GP-based classifiers, as well as the unseen classes by better description of the class using GP. The overall accuracy, using this method, increases to 63.33%. The test set for this experiment consisted of 90 randomly chosen images belonging to all the classes.

6 Conclusion

In this paper we propose a two stage Gaussian process (GP) based zero-shot learning method using relative attributes. The method is extensively evaluated on two standard datasets. The results from the method show consistent improvement over the basic Gaussian mixture model based approach for zero-shot learning that was proposed earlier [1]. The method while being more accurate is also more descriptive. The GP based classifier allows us to estimate the uncertainty in a test sample to belong to one of the seen classes. The GP kNN based regression allows us to obtain reliable estimates of the attributes distribution for the unseen class in terms of the relative attribute representation. These allow us to obtain a better understanding of the mid-level representation obtained through relative attributes.

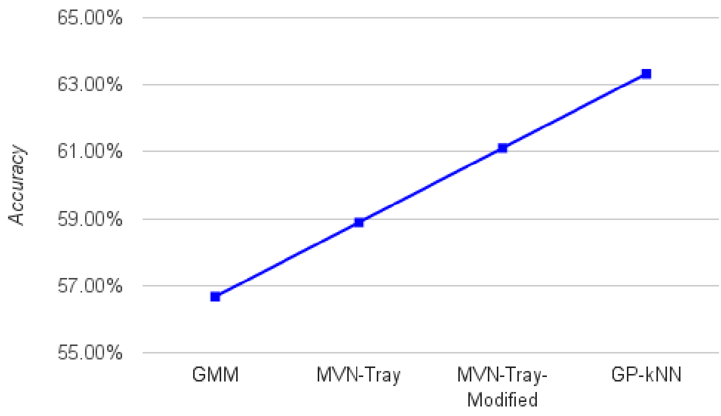

Fig. 8. *Accuracy curve for different approaches.* The curve depicts the accuracy of zero-shot recognition achieved by four different approaches. The accuracy of recognition increases as we go from left to right with GMM based method, MVN-Tray method, MVN-Tray-Modified for 'tie breaks' and our final proposed method using GP kNN.

Attributes	**Classes**	**Supervision**
Male (1)	Alex Rodriguez (A)	*seen-class*
White (2)	Clive Owen (C)	*seen-class*
Young (3)	Hugh Laurie (H)	*seen-class*
Smiling (4)	Jared Leto (J)	*seen-class*
Chubby (5)	Miley Cyrus (M)	(1): M <J (3): A <M (6): J <M <C (10): C <M <A
Visible Forehead (6)	Viggo Mortensen (V)	(3): V <A (4): V <C (5): V <J (10): H <V <C
Bushy Eyebrows (7)	Scarlett Johansson (S)	(1): S <J (3): S <M (8): J <S <H (9): C <S <H (10): A <S
Narrow Eyes (8)	Zac Efron (Z)	(1): Z <A (3): Z <J (5): H <Z <A (6): Z <C
Pointy Nose (9)		
Big Lips (10)		

Table 1. *Classes, relative attributes and supervision in one of the experiments with PubFig Dataset.* Given four seen classes, and the unseen classes are described using relative attributes with respect to the seen classes. Note that Supervision column marks the labels available for training.

In future we would like to undertake research to obtain structured attribute representations that are relative and are also structured with respect to the uncertainty or unreliability of the attribute. Further, it would be interesting to study the effect of the proposed method in the context of relative feedback.

References

1. Parikh, Devi and Grauman, Kristen: In Proceedings of International Conference on Computer Vision (ICCV). pp. 503-510. IEEE. Computer Society. (2011)
2. Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning. The MIT Press. (2006)
3. Biswas, Arijit and Parikh, Devi: Simultaneous Active Learning of Classifiers and Attributes via Relative Feedback. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2013)
4. Tamara L. Berg, Alexander C. Berg, Jonathan Shih: Automatic Attribute Discovery and Characterization from Noisy Web Images. In European Conference on Computer Vision (ECCV). (2010)
5. Christoph H. Lampert and Hannes Nickisch and Stefan Harmeling: Learning to detect unseen object classes by betweenclass attribute transfer. In IEEE International Conference on Computer Vision and Pattern Recognition(CVPR). (2009)
6. Lampert, Christoph H. and Nickisch, Hannes and Harmeling, Stefan: Attribute-Based Classification for Zero-Shot Visual Object Categorization. In IEEE Trans. Pattern Anal. Mach. Intell. **36(3)**, 453–465, (March, 2014)
7. Zeynep Akata and Florent Perronnin and Zaïd Harchaoui and Cordelia Schmid: Label-Embedding for Attribute-Based Classification. In Conference on Computer Vision and Pattern Recognition,Portland, OR, USA, June 23-28, 2013. pp. 819–826 (2013)
8. D. Jayaraman and K. Grauman: Zero-shot learning with unreliable attributes. NIPS. (2014)
9. A. Yu and K. Grauman: Just Noticeable Differences in Visual Attributes. In International Conference on Computer Vision (ICCV). (December,2015)
10. Elhoseiny, Mohamed and Saleh, Babak and Elgammal, Ahmed: Write a Classifier: Zero-Shot Learning Using Purely Textual Descriptions. In IEEE International Conference on Computer Vision (ICCV). (December, 2013)
11. Zhao Xu and Kristian Kersting and Volker Tresp: Multi-Relational Learning with Gaussian Processes. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI) 2009, Pasadena, California, USA, July 11-17. pp. 1309–1314 (2009)
12. Erik Rodner and Joachim Denzler: One-Shot Learning of Object Categories Using Dependent Gaussian Processes. In Pattern Recognition - 32nd DAGM Symposium, Darmstadt, Germany. September 22-24, 2010. Proceedings. pp. 232–241 (2010)
13. Rasmussen, Carl Edward and Nickisch, Hannes: Gaussian Processes for Machine Learning (GPML) Toolbox. In J. Mach. Learn. Res. (3/1/2010) Volume 11,dec,2010. **1532-4435**. pp. 3011-3015
14. Zhen-Yong Fu and Tao A. Xiang and Elyor Kodirov and Shaogang Gong: Zero-shot object recognition by semantic manifold distance. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015, Boston, MA, USA, June 7-12, 2015. pp. 2635-2644 (2015).
15. Mensink, T. E. J. and Gavves, E. and Snoek, C. G. M.: COSTA: Co-Occurrence Statistics for Zero-Shot Classification. In IEEE Conference on Computer Vision and Pattern Recognition. (2014)