UACI: Uncertain Associative Classifier for Object Class Identification in Images

L. Manikonda[1], A. Mangalampalli[1], V. Pudi[1]

[1]Center for Data Engineering, IIIT, Hyderabad 500032, India.
Email: lydia@research.iiit.ac.in

Abstract

Uncertainty is inherently present in many real-world domains like images. Analyses of such uncertain data using traditional certain-data-oriented techniques do not achieve best possible accuracy. UACI introduces the concept of representing images in the form of probabilistic or uncertain model using interest points in images. This model is an uncertain-data-based adaptation of Bag of Words, with each image not only represented by the visual words that it contains, but also their respective probabilities of occurrence in the image. UACI uses an Associative Classification approach to leverage latent frequent patterns in images for the identification of object classes. Unlike most image classifiers, which rely on positive and negative class sets (generally very vague) for training, UACI uses only positive class images for training. We empirically compare UACI with three other state-of-the-art image classifiers, and show that UACI performs much better than the other classifying approaches.

Keywords: Uncertain Mining, visual object identification, Associative Rule Mining, Associative Classification

1 Introduction

Advancements of recent technology has provided an opportunity to store and record large quantities of different types of data continuously. In many applications, data are inherently noisy, such as the huge data collected by sensors or by satellites. In some other cases for preserving the privacy [1] like in case of medical databases, customer transaction analysis, and demographic datasets, noise is added deliberately. Similarly, data being collected during the surveys which involves questionnaires and interviews, may be uncertain in nature. Since in such kind of applications information captured in the transactions have items associated with an existential probability, traditional mining techniques which were used for certain databases are not applicable. This has created a need for mining uncertain data [2]. An uncertain item is an item $x \in I$ present in a transaction $t \in T$ which is associated with an existential probability $P(x \in t) \in (0, 1)$. An uncertain transaction t is a transaction that contains uncertain items. A transaction dataset T containing uncertain transactions is called an uncertain dataset.

It is difficult for humans to discover underlying knowledge and patterns from huge image datasets. So, extracting knowledge from images is increasingly in demand. The focus is to extract the most relevant image features into a form suitable for mining. The mined patterns from such local image features are used for object identification and detection. Association Rule Mining (ARM) enables the extraction of such *latent patterns* in the form of association rules. These rules are based on their respective frequencies, and thus represent the *dominant trends* and statistically significant associations in the given dataset. A new classification approach called associative classification [3],[4],[5] leverages these patterns to build classifiers. Thus, associative classifiers are *highly accurate* and *very robust* because low-frequency patterns (noise) are eliminated during the ARM stage. Associative Classification when applied on uncertain probabilistic representation of data like images, provides even better accuracy and performance as compared to the corresponding traditional certain representation.

In this paper, we present our algorithm UACI which adapts uncertain associative classification to fit the image classification perspective. In UACI, we extract Scale Invariant Feature Transform (SIFT) [6] points from each image. SIFT points are local image features which are image scale and rotation invariant. They are based on the appearance of the object at particular interest points. Each image has different number of SIFT features. Then, clustering is done by combining all the SIFT features of all images present in the given dataset. The

978-1-4244-9631-0/10/$25.00

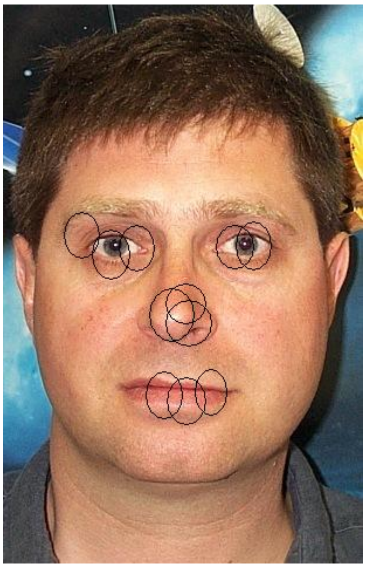

Figure 1: Example of Visual Words obtained through SIFT points in an image

images are expressed in terms of a modified Bag of Words (BOW) model. This modified BOW model consists not only of various words/clusters in the given vocabulary, but also the probabilities associated with these words/clusters in each image. Each image is represented as a transaction/record of words/clusters with associated probabilities to provide an uncertain data representation of the image. Thus, our model not only encapsulates the words/clusters associated with each image, but also their respective probabilities of occurrence in the given image. Fig. 1 illustrates various visual words (eyes, nose, and lips), obtained by using SIFT points, present in a human face. These visual words are combined with their respective probabilities to get a representation of each image as illustrated in Fig. 2. Patterns obtained from the occurrence (along with probability) of visual words in each image, help in the identification of image object classes.

Another salient property of UACI is that it is a one-class classifier, *i.e.* it requires only positive class images for training. In the image domain, especially for object class identification problems, the negative class is not well-defined and could include all images in the universal set that are not present in the positive class. While using a traditional two-class classifier, this could lead to the training of a lopsided classifier, which could be over-fitting too. But, UACI relies on only the positive class training set, and thus is very robust by avoiding the aforementioned problems.

Thus, UACI mainly aims at identifying the presence or absence of object classes in images using uncertain association rules and uncertain associative classification. Our main contributions through this paper are:

- Creation of a new concept of representing images in the form of *uncertain transactions/records* using *visual words/clusters* and their respective *probabilities* in each image.
- Use Uncertain ARM to extract *patterns*, in the form of *uncertain association rules*, from images.
- Come-up with a *one-class Uncertain Associative Classifier* which would identify the presence of object classes in images. For training only uncertain association rules obtained from positive class training images are used, with *no reliance on negative class training images.*

This paper is organised as follows: In Section 2, we introduce the related work of associative classification of Images. In Section 3, we discuss our algorithm. Detailed experimental results are presented in Section 4. In Section 5, we draw conclusions and present future work.

2 Related Work

In this paper, we consider the whole dataset to be uncertain instead of partial missing values in data. Uncertainty has been shown in terms of probabilistic distributions. There are few methods [7], [8] which try to solve specific classification tasks instead of developing a general algorithm for classifying uncertain data. But none of these algorithms have addressed the issue of developing classification algorithms for uncertain datasets.

Identifying frequent itemsets and association rule mining [9], [10], [11] from uncertain databases have also been becoming an active area of research. The support of itemsets and confidence of association rules are integrated with the existential probability of transactions and items. U-Apriori is proposed in [12] which essentially mimics the Apriori algorithm, except that it performs counting by computing the expected support of the different itemsets. The expected support of a set of items in a transaction is obtained by simply multiplying the probabilities of different items in the transaction. But much work has not been done by combining the association rule mining and classification techniques to mine the uncertain datasets. CBA [13] was one of the first associative classifiers and focuses on mining a special subset of association rules, called class association rules (CARs). Thabtah in [3] provides a very detailed survey of current associative classification algorithms. The author has compared most of the well-known algorithms, like CBA [13], CMAR [14], CPAR [5] among others, highlighting their pros and cons, and describing briefly how exactly they work. He lays more emphasis on pruning techniques, rule ranking, and predication methods used in various classifiers, and also provides valuable insights into the future directions which should be undertaken in associative classification. In this paper, we deal with an associative classifier for uncertain datasets.

To mine valuable information from a large image content, classifying image by content (low-level visual features) is an important way and is challenging [15]. Different types of classifiers [16], [17] have been built in order to perform this task. Application of Association Rule Mining with the help of spatial configurations and local neighborhoods has been described in [18] and [19] in detecting features in images and videos respectively. There is also an approach [20] based on the bag-of-words model for generic visual categorization. [21] adapts fuzzy associative classification for object class detection in images using the interest points.

3 Our Algorithm and Image Classification

In this section, we explain how our algorithm is used for classification of object classes in images which are represented by a probabilistic model. Given a image dataset, SIFT vectors are extracted from each image and are clustered using K-Means clustering algorithm creating a modified Bag of words model. This is followed by the identification of association rules which are transformed in to uncertain classification rules during training. While actual classification, the SIFT vectors extracted from the test images are interpolated with the centers of the clusters generated during the training phase. Information gain is calculated for each classification rule and classification process is performed by considering a threshold δ.

3.1 Probabilistic Data Generation

Given the positive class training dataset(I^{tr}), the first step in our algorithm is to extract different SIFT vectors($SIFT_i$) from each image, I_i^{tr}. K-means clustering algorithm is used to cluster all the generated SIFT vectors($SIFT_{Total}$) for the images in the training dataset. It generates a clustering C which has $\underline{K}$ number of different clusters - $C_1, C_2, \ldots, C_k$. After clustering, each image I_i^{tr} is represented in the modified form of Bag-Of-Words model, where each word represents the cluster associated with a probability value. Each image I_i^{tr} in this model is represented as a record r. As described in the Algorithm 1, initially each cluster id j present in each record r is associated with a total count of the number of SIFT features grouped in to that cluster C_j.

We normalize the values associated with each cluster C_j so as to maintain consistency throughout all the records. This leads to the transformation of the model in to a probabilistic or uncertain dataset which takes care of the problem experienced by the general Bag of Words model in which each word belongs to each image/record with the same probability of 1.

$Image_1$ $:<$ $Cluster_1, Prob_1$ $>,<$ $Cluster_2, Prob_2$ $>, \ldots, <$ $Cluster_i, Prob_i$ $>, \ldots <Cluster_k, Prob_k>$

$\vdots$

$Image_n$ $:<$ $Cluster_1, Prob_1$ $>,<$ $Cluster_2, Prob_2$ $>, \ldots, <$ $Cluster_i, Prob_i$ $>, \ldots <Cluster_k, Prob_k>$

Figure 2: Modified Bag of Words model representation for Uncertain Data

Algorithm 1 Probabilistic Dataset Generation

Given the training dataset of images $I^{tr} = \{I_1^{tr}, I_2^{tr}, \ldots, I_n^{tr}\}$
for each image ${I^{tr}}_i \in I$ **do**
  calculate $SIFT_i$ for ${I^{tr}}_i$
end for
$SIFT_{Total} = \bigcup_{i=1}^{n} SIFT_i$
Cluster $SIFT_{Total}$ into $K-$clusters, where the clustering $C = C_1, C_2, \ldots, C_k$
for each image ${I^{tr}}_i \in I$ **do**
  for each feature $f \in SIFT_i$ **do**
    Identify the cluster C_j to which the feature f belongs.
    if $f \in C_j$ **then**
      $Frequency(C_j)$ is incremented by 1
    end if
  end for
end for
for each image ${I^{tr}}_i \in I$ **do**
  for each cluster C_j in C **do**
    normalize the frequency value associated with C_j in ${I^{tr}}_i$ using the $\sum_{i=1}^{n} Frequency(C_j)_i$
  end for
end for

The probability value assigned with each cluster id is the normalized value of the frequency of SIFT vectors of an image clustered in to that particular cluster. Each image in the probabilistic data generated will be in the format as shown in Fig.2. Choosing the number of clusters while clustering is also an important step. Because, lesser the number of clusters, more is the loss of information about the images which is also the same in case of higher number of clusters. Hence, choosing the optimal number of clusters is important. In our algorithm while testing we have used the number of clusters based on the dataset considered.

3.2 Uncertain Associative Classifier Training

Most of the algorithms train their respective classifier with positive class and negative class datasets. But in UACI classifier, only a positive-class dataset is used for training the classifier. The first step in training is to generate association rules for the uncertain model. For generating the uncertain association rules, we have used an uncertain ARM algorithm which relies on the partitioning approach and TIDlists. The main reason for building an uncertain associative classifier instead of a certain associative classifier is to handle the probability associated with the cluster ids in the modified BOW model. After the generation of association rules, entropy and information gain are calculated for each rule generated. Given a rule $X \rightarrow Y$, X is an itemset composed of varying number of attributes and Y is the class label of the rule which is obtained from the dataset.

The probability of Y is considered to be the maximum probability of all the attributes in each rule. The information gain $IG(Y|X)$ (as shown in eq.(3)) of a given attribute X with respect to the class attribute Y is the reduction in uncertainty about the value of Y when we know the value of X. The pseudocode for training the classifier is shown in the Algorithm 2.

$$H(Y) = -\sum_{i=0}^{z} p_i \log p_i \tag{1}$$

$$H(Y|X) = -\sum_{j=1}^{n} P(X = x_j) H(Y|X = x_j) \tag{2}$$

$$IG(Y|X) = H(Y) - H(Y|X) \tag{3}$$

ARM generates a large set of rules(R), many of which are redundant. Pruning methods are used in order to improve the efficiency. For the pruning process, information gain(IG) of each rule r_i and rule length rl_i i.e., number of attributes in each rule. Each rule r_q is compared to all r_{q+1} to $r_{m'}$ rules. A given rule r_q (with information gain IG_q and rule length rl_q) is pruned ($R = R - r_q$) if there exists another rule r_s (with information gain IG_s and rule length rl_s) which is a superset of r_q, and $rl_q < rl_s$ and $IG_q < IG_s$. The size of R reduces from $m^{'}$ to $m^{''}$ after applying the pruning technique.

3.3 Image Classification

Image classification is done by using a set of uncertain classification rules derived during training. Given set of testing images I^{te} for classification, a set of SIFT vectors $SIFT_i$ are extracted from

Algorithm 2 Training the associative classifier

Apply ARM for the modified BOW model of I^{tr}
ARM generates rule set R which has $m^{'}$ number of rules
rule length of any rule $r_q \in R$ is rl_q
Information Gain of $r_q \in R$ is IG_q
for each rule $r_q \in R$ **do**
  for rule $r_s \in [r_{q+1}$ till $r_{m'}]$ **do**
    compare r_q with r_s
    if $rl_q < rl_s$ and $IG_q < IG_s$ **then**
      prune the rule r_q
    end if
  end for
end for
Pruning the redundant rules reduces the number of rules in R by $(m^{'} - m^{''})$
for each rule $r(X \rightarrow Y) \in R$ **do**
  calculate the entropy of X using eq.(1)
  calculate the average conditional entropy $H(Y|X)$ for Y with respect to X using eq.(2)
  calculate the Information Gain $IG(Y|X)$ using eq.(3)
end for

each image I_i^{te}. Let us say that there are f SIFT vectors for each image, where the value of f is based on the image considered. Each SIFT vector of an image should be mapped with the cluster centers $med_1, med_2, \ldots, med_k$ of the K-clusters obtained during the training phase. The cluster center(med_j) which is nearer to a SIFT vector(v) should be identified using eq.(4). The same procedure is applicable for all the remaining SIFT vectors $SIFT_i$ of a given image I_i^{te}. As shown in the Algorithm 3 the frequency with which the SIFT vectors belong to clusters are calculated and the image is interpolated in the same format which was done similarly in the training phase of positive class. Each image in the given set should be modified in to the format shown in Fig.2

$$CosineDistance = \frac{med_j \cdot v}{\|med_j\| \|v\|} \tag{4}$$

Each image ${I^{te}}_i$ from the positive class or negative class test datasets is classified similarly as follows. As explained in the Algorithm 4, each rule r_s in the generated rule set R(with $m^{''}$ rules) is applied to the image record r. We identify each of the t attributes that are in common with the antecedent of each rule and the given image record. Considering each of the probabilities $\{prob_1, prob_2, \ldots, prob_t\}$ associated with the t attributes from r, we calculate the average probability $avgProb_i^{r_s}$ by using eq.(5). We multiply this value with the information gain $IG(r_s)$ associated with the rule r_s as shown in eq.(6) and consider this obtained result as the uncertain information gain $UIG(r_s)$. For an image

Algorithm 3 Interpolation of testing dataset

Centers of the clusters obtained during the training process are $meds = \{med_1, med_2, \ldots, med_k\}$
Given the image dataset for testing I^{te} contains $I_1^{te}, I_2^{te}, \ldots, I_l^{te}$
for each image ${I^{te}}_i \in I^{te}$ **do**
  calculate $SIFT_i$ for ${I^{te}}_i$
  for each SIFT vector $v \in SIFT_i$ **do**
    for each center $med_j \in meds$ **do**
      calculate Cosine Distance of SIFT vector v from med_j
      if $f \in C_j$ **then**
        $Frequency(C_j)$ is incremented by 1
      end if
    end for
  end for
end for
for each image ${I^{te}}_i \in I$ **do**
  for each cluster C_j in C **do**
    normalize the frequency value associated with C_j in ${I^{te}}_i$ using the $\sum_{i=1}^{n} Frequency(C_j)_i$
  end for
end for

${I^{te}}_i$, we calculate the uncertain information gain obtained while applying each rule in the rule set R and add the values as shown in eq.(7). This is the Total Uncertain Information Gain($TUIG$) which is verified with a threshold δ. If the value of $TUIG$ is greater than or equal to δ, then the image belongs to the positive class or else it belongs to the negative class. Hence, given a set of images classification is done.

$$avgProb_i^{r_s} = \frac{\sum_{j=1}^{t} prob_j}{t} \tag{5}$$

$$UIG = \prod (IG)(avgProb_i^{r_s}) \tag{6}$$

$$TUIG = \sum_{s=1}^{m''} UIG_s \tag{7}$$

4 Experimental Results

In this section, we present the experimental results of the proposed uncertain association-based UACI algorithm. We have studied the performance of UACI as compared to I-FAC, BOW and SVM on the basis of false-positive-rate (FPR) versus recall curve. For our algorithm, we have considered a minimum support $\approx$ 0.01 for all the datasets studied. The support for each datasets is different and relies on how dense or sparse the dataset is, the number of items (singletons) involved in the dataset, and the average length of transactions in

Algorithm 4 Classification of the testing dataset

for Each image $I_i^{te} \in I^{te}$ **do**
  consider a record r_i of the image dataset
  for each rule $r_s \in R$ **do**
    Identify the t common attributes
    Calculate $avgProb_j^{r_s}$, using the probabilities associated with each of the t attributes
    calculate UIG
  end for
  calculate $TUIG$
  if $TUIG \geq \delta$ **then**
    Classify the image that it belongs to positive class
  else
    Classify the image that it belongs to negative class
  end if
end for

the dataset [22]. We have used 500 clusters for each dataset for the clustering process.

In BOW, we count how many times each visual word in the code-book occurs in an image. The results for BOW have been taken from the baseline of [18], which uses 3000 clusters to create the code-book. And, the results for SVM (RBF Kernel) and I-FAC have been taken from [21]. CALTECH Cars (Rear) background dataset was used as negative training set for BOW and SVM. UACI and I-FAC do not expect any negative class training set.

The datasets considered are:

1. **GRAZ Bikes:** Positive class training and positive class test sets respectively are randomly picked 25 and 38 images from the GRAZ bikes dataset. The first 200 images from CALTECH-101 background class dataset were used as negative class test set.

2. **ETHZ Giraffes:** Training was done on 93 images of giraffes downloaded from Google Images. The positive class test and negative class test datasets were 87 giraffe images and the rest 168 images respectively from the ETHZ Shape Classes dataset.

3. **CALTECH Faces:** 52 randomly picked images from the CALTECH Human Faces (Front) dataset were used for each of the positive class training and test sets. The first 200 images from CALTECH-101 background class dataset were used as negative class test set.

The performance of UACI, when compared with the other algorithms, on the basis of the three datasets used is shown in figures 3, 4, 5. The FPR-versus-recall curve is calculated by varying the va-

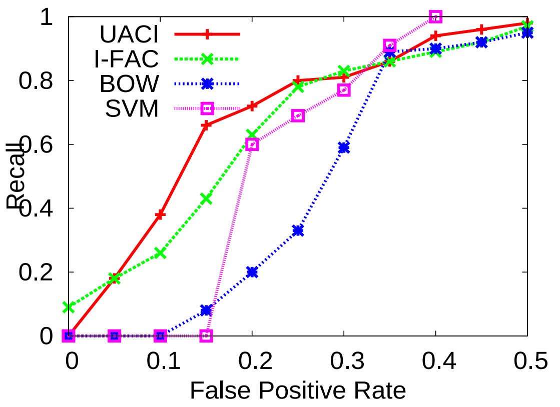

Figure 3: Graz Bikes dataset results

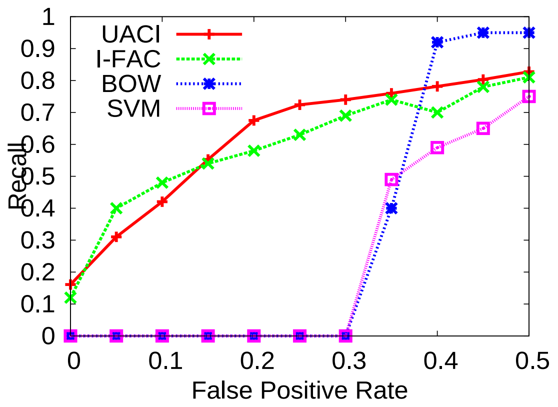

Figure 4: Giraffe dataset results

lue of δ. The value of δ for each dataset can be determined by cross-validation on the respective positive class and negative class test datasets. UACI consistently *out-performs* BOW and SVM on all datasets. Moreover, the performance of UACI is *better than* that of I-FAC for the GRAZ Bikes and Faces datasets, and comparable for the Giraffes dataset. Another highly desirable characteristic of UACI, which is very important for object class identification for images, is that it performs *very well at lower ranges of FPRs* (0.1–0.3), especially for the GRAZ Bikes and Faces datasets. The *good performance* of UACI is because of the following three features of UACI:

- First, is the representation of images in probabilistic or uncertain model instead of the traditional BOW model. This model captures the uncertainty which is inherently present in each feature of an image.

- Second, is the extraction of latent frequent patterns in the form of uncertain association rules. These patterns eliminate much of the noise and represent dominant trends and statistically significant associations in the given dataset, because of which the resultant associative classifier has a very high degree of accuracy.

- Third, the one-class classifier paradigm helps in building a classifier which is not affected by the negative class (generally very vague in case of images). This helps in creating a generic classifier which does not over-fit the negative class training set and can work very well with all kinds of test images.

5 Conclusions and Future Work

Data uncertainty is very common in many real-world applications. One such application is images that we have considered here. In this paper we have developed a new associative classifier which can be used for classification of images represented in the form of uncertain data records. This algorithm directly classifies positive class and negative class test images without any need of training the classifier on negative class dataset. By identifying the inefficiency of the traditional bag of words model, we have developed a modified bag of words model which helped in classifying positive class and negative class test images to a greater extent as shown in the results. Performance of this algorithm is better when compared to the other state-of-the-art classification algorithms for object class identification in images. The avenues of future work include performing further evaluation on different models with varying number of clusters and integrating meta-data (tags) with feature-based data for better mining of patterns.

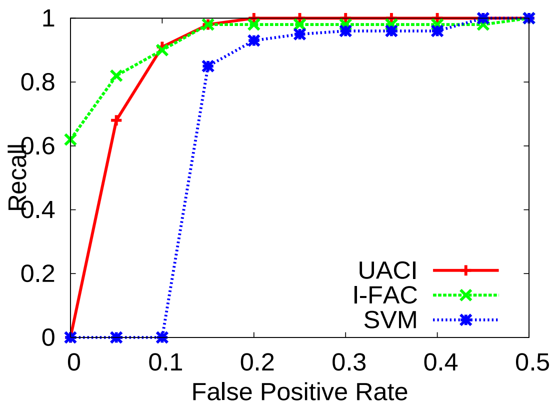

Figure 5: Faces dataset results

References

[1] Y. Xia, Y. Yang, and Y. Chi, “Mining association rules with non-uniform privacy concerns,” 2004, proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery.

[2] C. C. Aggarwal and P. S. Yu, "A survey of uncertain data algorithms and applications," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 5, pp. 609–623, 2009.

[3] F. A. Thabtah, "A review of associative classification mining," Knowledge Engineering Review, vol. 22, no. 1, pp. 37–65, 2007.

[4] A. Veloso, W. M. Jr., and M. J. Zaki, "Lazy associative classification," pp. 645–654, 2006, proceedings of the Sixth International Conference on Data Mining.

[5] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," 2003, proceedings of the third SIAM International Conference on Data Mining.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[7] E. Vilches, I. A. Escobar, E. E. Vallejo, and C. E. Taylor, "Targeting input data for acoustic bird species recognition using data mining and HMMs," 2007, proceedings of the Seventh IEEE International Conference on Data Mining Workshops.

[8] J. Chaker and O. Habib, "Genre categorization of web pages," 2007, proceedings of the Seventh IEEE International Conference on Data Mining Workshops.

[9] T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic frequent itemset mining in uncertain databases," 2009, proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

[10] C. K.-S. Leung, C. L. Carmichael, and B. Hao, "Efficient mining of frequent patterns from uncertain data," 2007, seventh IEEE International Conference on Data Mining Workshops.

[11] Q. Zhang, F. Li, and K. Yi, "Finding frequent items in probabilistic data," 2008, proceedings of the 2008 ACM SIGMOD international conference on Management of data.

[12] C.-K. Chui, B. Kao, and E. Hung, "Mining frequent itemsets from uncertain data," 2007.

[13] B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," 1998, proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining.

[14] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules," 2001, proceedings of IEEE International Conference on Data Mining.

[15] Y. Uehara, S. Endo, S. Shiitani, D. Masumoto, and S. Nagata, "A computer-aided visual exploration system for knowledge discovery from images," 2001, in Second InternationalWorkshop on Multimedia Data Mining (MDM/KDD2001).

[16] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," IEEE Transactions on Image Processing, vol. 10, no. 1, pp. 117–130, 2001.

[17] L. Bruzzone and D. Fernandez-Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 2, pp. 456–460, 2001.

[18] T. Quack, V. Ferrari, B. Leibe, and L. V. Gool, "Efficient mining of frequent and distinctive feature configurations," 2007, international Conference on Computer Vision (ICCV'07).

[19] T. Quack, V. Ferrari, and L. V. Gool, "Video mining with frequent itemset configurations," 2006, in Proceedings of CIVR.

[20] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," 2004, eCCV International Wrokshop on Statistical Learning in Computer Vision.

[21] A. Mangalampalli, V. Chaoji, and S. Sanyal, "I-FAC: Efficient fuzzy associative classifier for object classes in images," 2010, 20th International Conference on Pattern Recognition.

[22] Z. Zheng, R. Kohavi, and L. Mason, "Real world performance of association rule algorithms," in KDD, 2001, pp. 401–406.