

Copyright 2007 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

A model-based conceptual clustering of moving objects in video surveillance

Jeongkyu Lee ; Pragma Rajauria; Subodh K. Shah

Jeongkyu Lee ; Pragma Rajauria and Subodh K. Shah "A model-based conceptual clustering of moving objects in video surveillance", Proc. SPIE 6506, Multimedia Content Access: Algorithms and Systems, 650602 (January 29, 2007);

doi: 10.1117/12.708229

<http://dx.doi.org/10.1117/12.708229>

A model-based conceptual clustering of moving objects in video surveillance

Jeongkyu Lee*, Pragya Rajauria*, Subodh Kumar Shah*

Department of Computer Sciences & Engineering
University of Bridgeport
Bridgeport, CT 06604-5692 U.S.A.

ABSTRACT

Data mining techniques have been applied in video databases to identify various patterns or groups. Clustering analysis is used to find the patterns and groups of moving objects in video surveillance systems. Most existing methods for the clustering focus on finding the optimum of overall partitioning. However, these approaches cannot provide meaningful descriptions of the clusters. Also, they are not very suitable for moving object databases since video data have spatial and temporal characteristics, and high-dimensional attributes. In this paper, we propose a model-based conceptual clustering (MCC) of moving objects in video surveillance based on a formal concept analysis. Our proposed MCC consists of three steps: ‘*model formation*’, ‘*model-based concept analysis*’, and ‘*concept graph generation*’. The generated concept graph provides conceptual descriptions of moving objects. In order to assess the proposed approach, we conduct comprehensive experiments with artificial and real video surveillance data sets. The experimental results indicate that our MCC dominates two other methods, i.e., generality-based and error-based conceptual clustering algorithms, in terms of quality of concepts.

Keywords: Conceptual clustering, formal concept analysis, EM, moving objects

1. INTRODUCTION

The use of data mining techniques in many applications has increased enormously with the recent advances in computer power. Much of data collected in diverse areas of the industrial and scientific world contains valuable knowledge that is not easily retrievable. The knowledge identified by data mining may take form of predictive rules, clusters, patterns, and conceptual descriptions. To take this advantage into video content analysis, several studies have proposed the video data mining.¹⁻⁴

Clustering analysis, or simply clustering, is one of popular data mining techniques used to find the patterns and groups in video data.^{3,4} Most existing methods for the clustering focus on finding the optimum of overall partitioning. However, these approaches cannot provide any descriptions of the clusters. Conceptual clustering not only partitions the data, but generates resulting clusters that can be summarized by a conceptual description. For example, COBWEB⁵ is an incremental clustering algorithm based on probabilistic categorization trees. However, pure COBWEB only supports nominal attributes. In other words, it cannot be used for abstracting numeric data. To incorporate numeric values, some extensions of the COBWEB system are proposed, such as COBWEB/3,⁶ ECOBWEB,⁷ AUTOCLASS,⁸ Generality-based conceptual clustering (GCC),⁹ and Error-based conceptual clustering (ECC).¹⁰

However, all of the methods mentioned above do not represent the relations among the clusters or the appropriate meaning of each cluster. Also, they are not very suitable for moving object databases since video data have spatial and temporal characteristics, and high-dimensional attributes. To address these, we propose a model-based conceptual clustering algorithm (MCC) of moving objects in video surveillance systems. The proposed MCC differs from existing methods as follows: (1) it utilizes a formal concept analysis to generate concepts, (2) it uses models instead of individual data objects to handle complicated moving objects, and (3) it provides conceptual descriptions of moving object databases such as significant features and relationships.

The rest of the paper is organized as follows. In Section 2, we present a model-based conceptual clustering algorithm of moving objects in video surveillance systems. The performance study is reported in Section 3. Finally, Section 4 presents some concluding remarks.

*E-mail: {jelee, prajauri, subodhs}@bridgeport.edu

2. MODEL-BASED CONCEPTUAL CLUSTERING

In this section, we propose a model-based conceptual clustering (MCC) of moving objects, which is based on a formal concept analysis.^{11–14} Figure 1 shows the process of MCC of moving objects. As seen in the figure, MCC consists of three steps. Step 1 is a ‘*model formation*’. We form a set of models from moving object database by using a clustering analysis. Among existing clustering techniques, we choose Expectation Maximization (EM) algorithm since we can use the byproducts of EM, i.e., the extracted model parameters and the expectation values in the following steps. Each model, i.e., cluster, is characterized by a set of parameters used for computing significance values of features. Step 2 is a ‘*model-based concept analysis*’. In order to extract formal concepts from the models formed in Step 1, we propose a model-based formal concept analysis (MFCA) by applying the models to a formal concept analysis. Step 3 is a ‘*concept graph generation*’. This step refines the formal concepts and relations obtained in Step 2. MFCA typically produces a large number of concepts. There exist some number of similar concepts that need to be merged into a single concept. The final result of MCC is a concept graph that includes a set of conceptual nodes and relational edges. From the concept graph we can generate conceptual descriptions of moving object data.

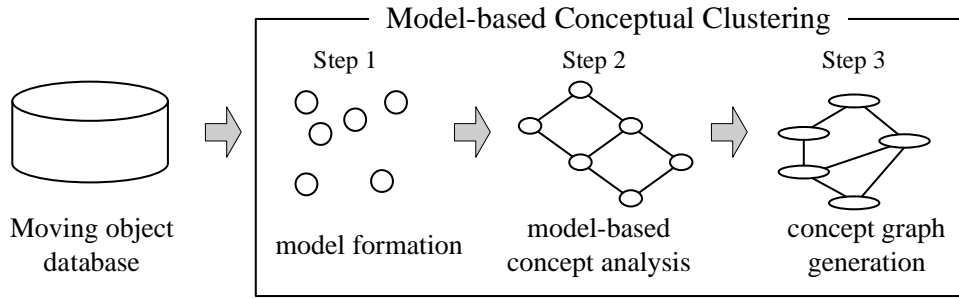


Figure 1. The process of model-based conceptual clustering of moving objects

2.1. Model Formation

In this subsection, we first introduce a distance function *STED* (Spatio-Temporal Edit Distance) that measures the dissimilarity between two moving objects. Then, we describe a model formation based on EM clustering with *STED*.

2.1.1. Spatio-Temporal Edit Distance

In order to cluster moving objects, we need an appropriate distance function that can measure the dissimilarity between two moving objects considering their spatial and temporal characteristic. Let MO_m^s and MO_n^t be the s^{th} and the t^{th} moving objects with time m and n , respectively.

$$MO_m^s = \langle \vec{v}_1^s, \dots, \vec{v}_m^s \rangle, \quad MO_n^t = \langle \vec{v}_1^t, \dots, \vec{v}_n^t \rangle$$

where \vec{v}_i is a feature vector of dimension d , which has the observed values of a moving object at time slot i . The distance function *STED* between MO_m^s and MO_n^t is defined as follows:

DEFINITION 1. The spatio-temporal edit distance (*STED*) between two moving objects MO_m^s and MO_n^t is:

$$STED(MO_m^s, MO_n^t) = \min[STED(MO_{m-1}^s, MO_n^t) + d_{st}(\vec{v}_m^s, \vec{g}), \\ STED(MO_m^s, MO_{n-1}^t) + d_{st}(\vec{g}, \vec{v}_n^t), \\ STED(MO_{m-1}^s, MO_{n-1}^t) + d_{st}(\vec{v}_m^s, \vec{v}_n^t)]$$

where \vec{g} is an edited vector with $\vec{g}_i = \frac{\vec{v}_{i-1}^s + \vec{v}_i^s}{2}$, and the cost function $d_{st}(\vec{v}_i^s, \vec{v}_j^t) = \sum_{l=1}^d \omega_l |v_{i,l}^s - v_{j,l}^t|$.

In the cost function d_{st} of Definition 1, ω_l is a weight of the l^{th} feature of MO where $\sum_{l=1}^d \omega_l = 1$, and $v_{i,l}^s$ indicates a value of the l^{th} component of \vec{v}_i^s . The weight of the l^{th} feature, ω_l is selected based on the variance (σ_l) of its feature as follows:

$$\omega_l = \frac{\log(1 - \sigma_l)}{\sum_{k=1}^d \log(1 - \sigma_k)}, \quad (1 \leq l \leq d) \quad (1)$$

If a certain feature has a higher value of variance, it is considered as less important feature than others, i.e., a lower value of weight.

2.1.2. EM Clustering of Moving Objects

The process of model formation is the same as a model-based clustering, since the clustering conducts unsupervised learning to find models of moving objects, which are characterized by a set of parameters. In order to find the models of moving objects, we employ a model-based EM clustering^{4,15} with $STED$ in Definition 1. Given M mutually independent sample MOs , $\mathcal{Y} = \{y_1, \dots, y_M\}$, the d -dimensional Gaussian mixture density with $STED$ is given by

$$p(y_j|\Theta) = \sum_{k=1}^K \frac{w_k}{2\pi^{1/2}|\sigma_k|} e^{-\frac{1}{2\sigma_k^2}STED(y_j, \mu_k)^2} \quad (2)$$

where K is the number of models, and w_k ($\sum_{k=1}^K w_k = 1$) is the membership probability of the k^{th} model. The log-likelihood (\mathcal{L}) of K mixture model is

$$\mathcal{L}(\Theta|\mathcal{Y}) = \log \prod_{j=1}^M p(y_j|\Theta) = \sum_{j=1}^M \log \sum_{k=1}^K w_k p_k(y_j|\theta_k) \quad (3)$$

where each θ_k is the set of parameters of the k^{th} model. In order to find appropriate models, we estimate the optimal values of the parameters (θ_k) and the weights (w_k) in Equation (3) using EM algorithm, which is a common procedure used to find the Maximum Likelihood Estimates (MLE) of the parameters iteratively. The EM algorithm produces the MLE of the unknown parameters using two alternating steps:

E-step: It evaluates the posterior probability of y_j , belonging to each model k . Let h_{jk} be the probability of the j^{th} MO for a model k , then it can be defined as follows:

$$h_{jk} = P(k|y_j, \theta_k) = \frac{w_k}{p_k(y_j|\theta_k)} \quad (4)$$

M-step: It computes the new parameter value that maximizes the probability using h_{jk} in E-step as follows:

$$w_k = \frac{1}{M} \sum_{j=1}^M h_{jk}, \quad \mu_k = \frac{\sum_{j=1}^M h_{jk} y_j}{\sum_{j=1}^M h_{jk}}, \quad \sigma_k = \frac{\sum_{j=1}^M h_{jk} STED(y_j, \mu_k)^2}{\sum_{j=1}^M h_{jk}}$$

The iteration of E and M steps is stopped when w_k is converged for all k . After the maximum likelihood model parameters ($\hat{\Theta}$) in Equation (3) are decided, each MO is assigned to a model.

The benefits of model formation in the proposed conceptual clustering are as follows: (1) it simplifies the computation of a conceptual clustering algorithm since the number of objects is reduced to the number of models that represent their specific patterns, and (2) byproducts of the EM algorithm, i.e., the extracted model parameters (Θ) and the expectation (h_{jk}) play an important role in conceptual clustering since they provide the relationships between models and features.

2.2. Model-based Concept Analysis

The second step of MCC is a model-based concept analysis to find formal concepts of modelled moving object data. Although the model formation in Step 1 provides a set of models where similar moving objects are grouped, it does not consider interpreting the obtained models. In other words, it is hard to understand what the meaning of each model is, and how the models are formed. In order to address these, we employ a formal concept analysis (FCA) that was introduced by Wille.^{11,12} FCA has been used for data analysis and knowledge representation. We propose a model-based concept analysis (MFCA), which incorporates a model-based statistical data analysis with FCA to represent conceptual meaning of moving object data. The MFCA starts with a model-based formal context that is defined as:

DEFINITION 2. A model-based formal context is a triple $K = (G, F, I)$, where

- G is a set of models characterized by Θ ,
- F is a set of features, and
- I is a set of relations between G and F (i.e. $I \subseteq G \times F$). Each relation $(g, f) \in I$ has a significance value $\lambda(g, f)$.

In the model-based formal context, G consists of the models determined in Step 1 instead of all data objects. Each model is characterized by a set of parameters (Θ). A set of features (F) consists of d number of observed features in MO . A set of relations I indicates how much each feature is relevant to each model. The relevance between the k^{th} model g_k and the l^{th} feature f_l is represented as the *significance value* $\lambda(g_k, f_l)$, which is computed by exploiting a feature selection technique in data mining.¹⁶

Let y_F and $y_{F-\{f_l\}}$ be the full feature vector, and the feature vector without the l^{th} feature, respectively. Consider two posterior probabilities ($P_{i,k,F}$, $P_{i,k,F-\{f_l\}}$) of the k^{th} model based on the full feature vector (y_F), and the feature vector without f_l feature ($y_{F-\{f_l\}}$) as follows.

$$P_{i,k,F} = P(y_{i,F}|\theta_{k,F}), \quad P_{i,k,F-\{f_l\}} = P(y_{i,F-\{f_l\}}|\theta_{k,F-\{f_l\}})$$

where $\sum_{i=1}^m P_{i,k,F} = 1$ and $\sum_{i=1}^m P_{i,k,F-\{f_l\}} = 1$ with m number of objects in the k^{th} model. If f_l is a completely insignificant feature in the k^{th} model, then $P_{i,k,F}$ is equal to $P_{i,k,F-\{f_l\}}$. Otherwise, the difference between two probabilities provides a significance value of f_l in the model. In order to measure a difference between two probabilities, we use the Kullback-Leibler divergence (KLD).¹⁷ The significance value of the l^{th} feature (f_l) in the k^{th} cluster is defined as:

$$\lambda(g_k, f_l) = \sum_{i=1}^m \left| P_{i,k,F} \log \frac{P_{i,k,F}}{P_{i,k,F-\{f_l\}}} \right| \quad (5)$$

The higher the value of λ , the more significance of the feature in a model. Then, we determine all relations between a set of models G and a set of features F .

A model-based formal context K can be represented as a cross-table as shown in Table 1 (a). The context in Table 1 (a) has four models (i.e. g_1, g_2, g_3 and $g_4 \in G$), and three features (i.e. f_1, f_2 and $f_3 \in F$). The relation (I) between a model and a feature is measured by λ in Equation (5).

To remove relations that have very low significance values, we set a predetermined threshold ϵ . Table 1 (b) shows the cross-table of the model-based formal context with $\epsilon = 0.01$.

According to a formal concept analysis, we can consider the features of a formal context as the description of the concept.^{11,12} Therefore, we can derive a model-based formal concept from a model-based formal context as follows (Definitions 3 and 4).

DEFINITION 3. For a set of models $A \subseteq G$, we define $A' = \{f \in F | \lambda(g, f) > \epsilon \text{ for all } g \in A\}$, and a set of features $B \subseteq F$, we define $B' = \{g \in G | \lambda(g, f) > \epsilon \text{ for all } f \in B\}$

Table 1. Example of model-based formal context
(a) Without threshold (b) With threshold ($\epsilon = 0.01$)

	f_1	f_2	f_3		f_1	f_2	f_3
g_1	0.583	0.004	0.431	g_1	0.583	-	0.431
g_2	0.840	0.002	0.003	g_2	0.840	-	-
g_3	0.002	0.454	0.623	g_3	-	0.454	0.623
g_4	0.000	0.517	0.833	g_4	-	0.517	0.833

DEFINITION 4. A model-based formal concept of the context $K = (G, F, I)$ with a threshold value ϵ is a pair (A, B) where $A \subseteq G$, $B \subseteq F$, $A' = B$, and $B' = A$. We call A the extent, and B the intent of the formal concept (A, B) .

The extent covers all models belonging to the formal concept, while the intent comprises all features valid for all those models. Table 2 shows the complete list of concepts for the context in Table 1 (b). For example, a concept C_4 has the extent $A = \{g_3, g_4\}$, and the intent $B = \{f_2, f_3\}$ (see highlighted cells in Table 1 (b)). The extent A and the intent B of the formal concept (A, B) are closely related by the relation I . In other words, the concept (A, B) means the models determined by features f_2 and f_3 .

Table 2. Formal concepts for context in Table 1(b)

Concept	Extent (A)	Intent (B)
C_0	$\{g_1, g_2, g_3, g_4\}$	\emptyset
C_1	$\{g_1, g_2\}$	$\{f_1\}$
C_2	$\{g_1, g_3, g_4\}$	$\{f_3\}$
C_3	$\{g_1\}$	$\{f_1, f_3\}$
C_4	$\{g_3, g_4\}$	$\{f_2, f_3\}$
C_5	\emptyset	$\{f_1, f_2, f_3\}$

It is natural to have a hierarchical order between the model-based formal concepts of the context, i.e. *subconcept* or *superconcept* relation. In order to represent the relation, we exploit lattice theory¹² since it offers a natural way to formalize the ordering of objects. A model-based concept lattice is defined in Definition 5 and 6.

DEFINITION 5. If (A_1, B_1) and (A_2, B_2) are formal concepts of a model-based formal context, then (A_1, B_1) is the subconcept of (A_2, B_2) such that $A_1 \subseteq A_2$ (i.e., $B_2 \subseteq B_1$), denoted as $(A_1, B_1) \leq (A_2, B_2)$. In this case, (A_2, B_2) is superconcept of (A_1, B_1) .

DEFINITION 6. A model-based concept lattice of the model-based formal context $K = (G, F, I)$ is a set of all model-based formal concepts of K with the order \leq , denoted by $\mathfrak{B}(K, \leq)$.

For the purpose of visualization of the lattice, we use a line diagram consisting of circles and lines for all models and features, respectively. Each circle represents a formal concept, and each line indicates a relation, i.e. subconcept (downward) or superconcept (upward) in Definition 5. Figure 2 shows the line diagram of the model-based concept lattice $\mathfrak{B}(K, \leq)$ generated from the model-based formal context $K = (G, F, I)$ in Table 1.

2.3. Concept Graph Generation

Although the generated formal concepts using MFCA provide formal descriptions of models, MFCA typically produces a high number of formal concepts. Among them there exist some similar formal concepts that need to be merged into a single concept. In order to compact the formal concepts, we first propose a similarity measure *ConSim* between two formal concepts.

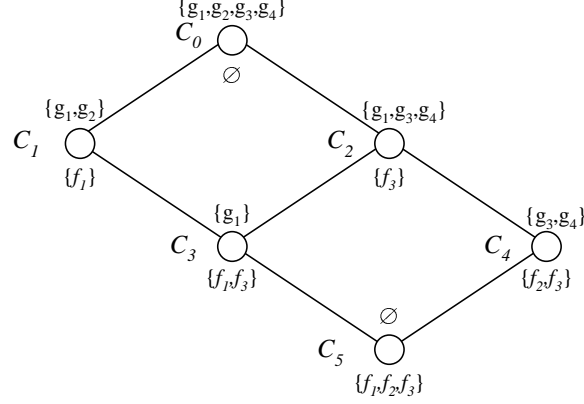


Figure 2. A model-based concept lattice \mathfrak{B} for the context K in Table 1 (b)

To measure a similarity between two concepts generated by MFCA, we exploit the similarity of FOGA framework,¹⁴ then extend it to a model-based formal concept. In FOGA, the similarity between two formal concepts $C_1 = (\varphi(A_1), B_1)$ and $C_2 = (\varphi(A_2), B_2)$, where φ is a fuzzy membership function, is defined as $E(C_1, C_2) = \frac{|\varphi(A_1) \cap \varphi(A_2)|}{|\varphi(A_1) \cup \varphi(A_2)|}$. However, this similarity measure is not enough for our model-based formal concepts since it considers only a fuzzy membership for the similarity, i.e., $\varphi(A_1)$ and $\varphi(A_2)$, but the proposed concept consists of the extent and the intent. We extend it to handle both the extent and the intent of model-based formal concepts for the similarity. Let $C_1 = (A_1, B_1)$ and $C_2 = (A_2, B_2)$ be two model-based formal concepts in $\mathfrak{B}(K, \leq)$. The similarity measure $ConSim$ between C_1 and C_2 is defined as follows.

DEFINITION 7. Given two model-based formal concepts C_1 and C_2 , a concept similarity ($ConSim$) between C_1 and C_2 is defined as:

$$ConSim(C_1, C_2) = \gamma \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} + (1 - \gamma) \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}$$

where $|set|$ is the number of elements in set, and γ is a predefined weight of the extent concept such that $0 \leq \gamma \leq 1$.

The concept similarity $ConSim$ in Definition 7 ranges 0 to 1. The higher the value of $ConSim$ is, the more similar two concepts are. The predefined weight (γ) plays a role to decide a priority between the extent and intent of the concept. For example, if a user has more confidence to the extent of concept (i.e., a set of models) than the intent (i.e. a set of features), γ is greater than 0.5. Otherwise, γ is set to less than 0.5. Throughout this paper, we set $\gamma = 0.5$, which means the extent and intent have the same priority for $ConSim$. Figure 3 shows the computed $ConSim$ with $\gamma = 0.5$ of Figure 2.

Since the MFCA uses a statistical approach for the significance values between features and models, even a small difference of the significance values may cause similar concepts to be separated into different concepts. Moreover, the number of generated formal concepts is usually large since the MFCA allows a model to belong to more than one concept. Such a large number of formal concepts prevent a user from interpreting them easily. In order to address this, we introduce a technique of merging two similar formal concepts into a single formal concept, and a concept graph where the merged concepts are represented.

Let $C_1 = (A_1, B_1)$ and $C_2 = (A_2, B_2)$ be two model-based formal concepts in $\mathfrak{B}(K, \leq)$. The criteria of merging two concept (C_1 and C_2) are defined as follows:

DEFINITION 8. Given two model-based formal concepts C_1 and C_2 , $C_1 \cup_M C_2 = (A_1 \cup A_2, B_1 \cup B_2)$ is a merged concept if C_1 and C_2 satisfy

- $(A_2, B_2) \leq (A_1, B_1)$ C_1 is superconcept of C_2 , and
- $ConSim(C_1, C_2) > T_{sim}$.

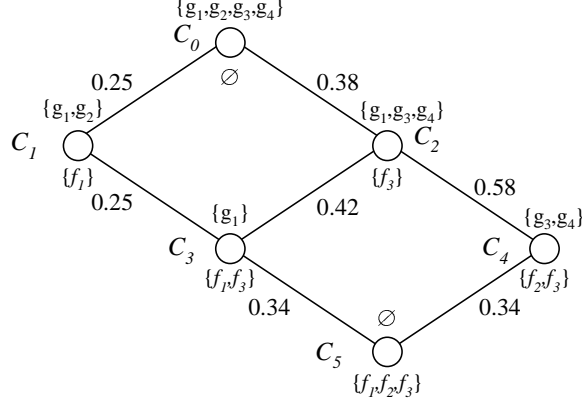


Figure 3. Concept similarities between two concepts of $\mathfrak{B}(K, \leq)$ in Figure 2

where T_{sim} is a predefined threshold value for the concept similarity. For an example of Figure 3, C_2 and C_4 can be merged into a single concept $C_2 \cup_M C_4$ with $T_{sim} = 0.5$. Since $C_4 \leq C_2$ and $ConSim(C_2, C_4) = 0.58 (> T_{sim})$, they satisfy the merging criteria in Definition 8.

However, if two formal concepts are merged into a single concept in $\mathfrak{B}(K, \leq)$, the set of merged concepts is no longer model-based concept lattice. Therefore, we cannot use a model-based concept lattice after merging concepts. In order to address this, we propose a concept graph to maintain the merged concepts and their relations, which is defined as follows:

DEFINITION 9. A concept graph, \mathcal{G} , is a four-tuple $\mathcal{G} = \{V, E, \nu, \xi\}$, where

- V is a finite set of conceptual nodes,
- $E \subseteq V \times V$ is a finite set of relational edges between two concept nodes,
- $\nu : V \rightarrow A_V$ is a function generating the conceptual node attributes, and
- $\xi : E \rightarrow A_E$ is a function generating the relational edge attributes.

A concept graph \mathcal{G} is a directed graph whose edges are ordered pairs of node and attribute. A conceptual node ($v \in V$) corresponds to a (merged) concept where the node attributes (A_V) are sets of models and features. A relational edge ($e(v, v') \in E$) represents a relation between two concepts, i.e., superconcept or subconcept for the edge attributes (A_E). For example, $e(v, v')$ indicates that v is a superconcept of v' (i.e. v' is a subconcept of v). The relational edge is weighted by the similarity between two conceptual nodes using $ConSim$ in Definition 7. The benefit of using the concept graph is that it is more flexible to maintain the merged concepts and their relations without loss of their semantics than a model-based concept lattice in Definition 6.

Figure 4 (a) is the result of concept merging of Figure 3 with $T_{sim} = 0.5$. We can observe that two formal concepts C_2 and C_4 can be merged into a single concept, since $ConSim(C_2, C_4) > 0.5$. Figure 4 (b) shows the concept graph of Figure 4 (a) in which a set of nodes represents concepts of MOs , and a set of edges represents relations between two corresponding concepts.

3. EXPERIMENTAL RESULTS

In order to assess the proposed schemes, we have conducted several experiments with a synthesized data set and a real data set. Using these data sets, we evaluate the performances of our proposed approach, and demonstrate that the quality of generated concepts based on MCC outperforms those of existing conceptual clustering algorithms.

We ran a generality-based conceptual clustering (GCC)⁹ and an error-based conceptual clustering (ECC)¹⁰ for comparison purposes. To be fair, we replace the similarity measure (Sim) used in GCC⁹ with $STED$. ECC

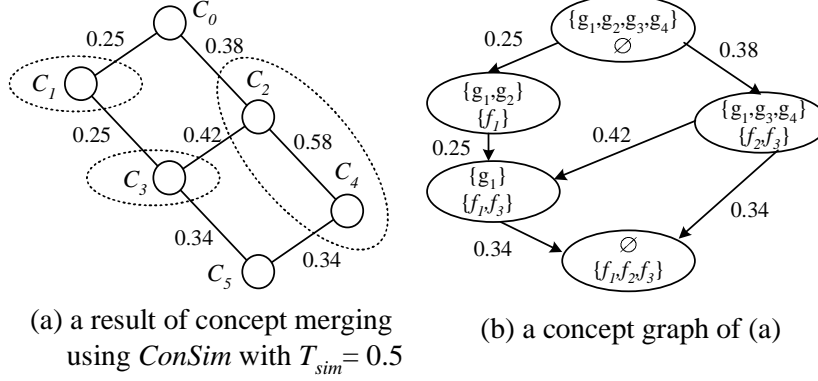


Figure 4. Result of model-based conceptual clustering

uses *BinaryCut*(C) algorithm to find the best cut of the input cluster.¹⁰ However, *BinaryCut*(C) can handle only ordered data, which is not applicable to *MOs*. To address this, we replace *BinaryCut*(C) with k-Means algorithm ($k = 2$) to find the best cut of *MOs* without losing its originality.

3.1. Evaluation Metrics

In order to evaluate the concept graph of MCC, we employ two evaluation metrics: the *relaxation error*,¹⁰ and *F-measure*.¹⁸ The relaxation error (RE) measures the goodness of the generated concepts, which is a popular criterion of evaluating conceptual clustering algorithm. On the other hand, since the concept graph has a hierarchical tree structure, we need a specific metric that can analyze the entire hierarchical tree. On such a metric, we also use F-measure (F_m) that combines the precision and recall ideas from information retrieval.

Given a particular concept of d -dimensional feature *STs*, i.e., $C_k = \{y_1, \dots, y_m\}$, the relaxation error (RE_{C_k}) for the concept C_k is defined as the average pair-wise distance using *STED* in Definition 1 among *STs* in C_k .

$$RE_{C_k} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m P(y_i)P(y_j)STED(y_i, y_j)$$

where $P(y_i)$ and $P(y_j)$ are the probabilities of y_i and y_j occurring in C_k , respectively. RE_{C_k} implies the dissimilarities of data objects (*STs*) in a concept C_k based on the distance function *STED*. Therefore, the relaxation error ($RE(C)$) for entire concepts can be computed as follows:

$$RE(C) = \frac{1}{K} \sum_{k=1}^K RE_{C_k} \quad (6)$$

where K is the total number of concepts. A smaller value of $RE(C)$ corresponds to a higher quality of the concepts.

In addition to the relaxation error, we use F-measure for the evaluation metric. In F-measure, we consider each concept as if it were the result of a query, and each class as if it were the desired set of *STs* for the query. Given a particular concept C_k of size m_k , and a particular class R_j of size m_j , the precision (p) and recall (r) for each C_k and R_j are

$$p_{kj} = \frac{m_{kj}}{m_k}, \quad r_{kj} = \frac{m_{kj}}{m_j}$$

where m_{kj} is the number of data items that are correctly clustered from R_j in C_k . The F-measure of concept C_k and class R_j is computed as follows:

$$F(C_k, R_j) = \frac{2 \cdot r_{kj} \cdot p_{kj}}{r_{kj} + p_{kj}} \quad (7)$$

Table 3. The results of concepts generated by conceptual clusterings on synthesized data sets

Algorithm	D _{0%}	D _{5%}	D _{10%}	D _{15%}	D _{20%}	D _{25%}	D _{30%}
MCC	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60,R ₃ :5,R ₅ :7) C ₂ (R ₂ :60), C ₃ (R ₃ :55) C ₄ (R ₄ :60), C ₅ (R ₅ :53) C ₆ (R ₆ :60)	C ₁ (R ₁ :60,R ₃ :7,R ₅ :9) C ₂ (R ₂ :60,R ₃ :3,R ₆ :8) C ₃ (R ₃ :50), C ₄ (R ₄ :60) C ₅ (R ₅ :51), C ₆ (R ₆ :52)	C ₁ (R ₁ :57,R ₃ :2,R ₅ :2) C ₂ (R ₂ :60,R ₃ :1,R ₆ :2) C ₃ (R ₃ :56), C ₄ (R ₄ :60) C ₅ (R ₅ :3,R ₃ :1,R ₅ :58) C ₆ (R ₆ :58)
ECC	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60,R ₃ :4, R ₅ :5) C ₂ (R ₂ :60), C ₃ (R ₃ :56) C ₄ (R ₄ :60), C ₅ (R ₅ :43) C ₆ (R ₆ :60)	C ₁ (R ₁ :60,R ₃ :12, R ₅ :9) C ₂ (R ₂ :60,R ₃ :5, R ₆ :11) C ₃ (R ₃ :43) C ₄ (R ₄ :60) C ₅ (R ₅ :51) C ₆ (R ₆ :49)	C ₁ (R ₁ :41) C ₂ (R ₂ :35) C ₃ (R ₃ :33) C ₄ (R ₄ :60) C ₅ (R ₅ :44) C ₆ (R ₆ :8,R ₃ :5,R ₆ :47) C ₇ (R ₇ :19,R ₃ :14,R ₅ :16) C ₈ (R ₈ :17,R ₃ :8,R ₆ :13)	C ₁ (R ₁ :38,R ₄ :12) C ₂ (R ₂ :35) C ₃ (R ₃ :27, R ₅ :13) C ₄ (R ₄ :48) C ₅ (R ₅ :37) C ₆ (R ₆ :8,R ₃ :8,R ₆ :38) C ₇ (R ₇ :22,R ₃ :15,R ₅ :10) C ₈ (R ₈ :17,R ₃ :10,R ₆ :22)	C ₁ (R ₁ :33,R ₄ :18) C ₂ (R ₂ :31,R ₄ :16) C ₃ (R ₃ :27,R ₅ :13) C ₄ (R ₄ :26) C ₅ (R ₅ :20) C ₆ (R ₆ :12,R ₃ :9,R ₆ :38) C ₇ (R ₇ :27,R ₃ :14,R ₅ :27) C ₈ (R ₈ :17,R ₃ :10,R ₆ :22)	C ₁ (R ₁ :31,R ₄ :1) C ₂ (R ₂ :29) C ₃ (R ₃ :21,R ₃ :32,R ₆ :28) C ₄ (R ₄ :10,R ₄ :33) C ₅ (R ₅ :14,R ₃ :28,R ₅ :34) C ₆ (R ₆ :32) C ₇ (R ₇ :26) C ₈ (R ₈ :15, R ₄ :26)
GCC	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60) C ₂ (R ₂ :60) C ₃ (R ₃ :60) C ₄ (R ₄ :60) C ₅ (R ₅ :60) C ₆ (R ₆ :60)	C ₁ (R ₁ :60,R ₃ :5,R ₅ :3) C ₂ (R ₂ :60,R ₃ :4,R ₆ :7) C ₃ (R ₃ :51) C ₄ (R ₄ :60) C ₅ (R ₅ :57) C ₆ (R ₆ :53)	C ₁ (R ₁ :51,R ₃ :10,R ₅ :8) C ₂ (R ₂ :49,R ₃ :7,R ₆ :26) C ₃ (R ₃ :8,R ₂ :6,R ₅ :39) C ₄ (R ₄ :1,R ₃ :3,R ₄ :60) C ₅ (R ₅ :4,R ₅ :52) C ₆ (R ₆ :2,R ₆ :34)	C ₁ (R ₁ :42,R ₃ :20,R ₅ :34) C ₂ (R ₂ :39,R ₃ :37,R ₆ :24) C ₃ (R ₃ :15,R ₂ :18,R ₄ :32) C ₄ (R ₄ :3,R ₃ :3,R ₅ :26) C ₅ (R ₅ :3,R ₄ :28,R ₆ :36)	C ₁ (R ₁ :31,R ₃ :22,R ₅ :60) C ₂ (R ₂ :34,R ₃ :28,R ₆ :60) C ₃ (R ₃ :27,R ₂ :22,R ₄ :60) C ₄ (R ₄ :2,R ₂ :4,R ₃ :10)

Consequently, the F-measure for an entire hierarchy of any concepts ($F_m(C)$) is computed by taking the weighted average of all values for $F(C_k, R_j)$ in Equation (7) as follows:

$$F_m(C) = \sum_{j=1}^c \frac{m_j}{m} \max_k F(C_k, R_j) \quad (8)$$

where c and m is the total number of classes and STs , respectively. The values of $F_m(C)$ range from 0 to 1. The higher the F-measure is, the better the quality of concept is. We use both $RE(C)$ in Equation (6) and $F_m(C)$ in Equation (8) to compare the quality of generated concepts C with those of other conceptual clustering algorithms in the following subsection.

3.2. Synthesized data set

We first generate a synthesized data set for STs , then present the results of MCC comparing with GCC and ECC in terms of a quality of concepts.

3.2.1. Data Description

In order to evaluate the proposed MCC algorithm, an synthesized data set for moving objects is generated and used for the experiments. The synthesized data set consists of 6 classes, i.e., $\{R_1, \dots, R_6\}$, and each class contains 60 moving objects (MOs) that are distributed by Gaussian with $\sigma = 5$ from a seed MO . Each MO is described by three features, i.e., spatial location (x, y) and its size (z) , and has various time lengths ranging from 10 to 20. We generate 7 data sets that have different amounts of noise ranging from 0% to 30%, denoted by $D_{0\%}, \dots, D_{30\%}$, respectively.

3.2.2. Results

Using the seven artificial data sets ($D_{0\%}, \dots, D_{30\%}$), we ran three conceptual clustering algorithms, i.e., MCC, ECC, and GCC, to generate concepts. The results of concepts generated by the three methods on the seven data sets are listed in Table 3. In Table 3, C_k is the k^{th} concept that is generated by the corresponding clustering algorithm. R_j and the number in parentheses indicate the j^{th} class that is a true concept, and the number of data in each of class, respectively. The subscript k of C_k is selected by the majority class of data. If two concepts share the same majority class, the concept whose class is larger than the other takes the class as the majority. For the other concept, we select the next majority class in the concept. For example, suppose that two concepts $C_k(R_1 : 35, R_2 : 20)$ and $C_{k'}(R_1 : 25, R_3 : 20)$ are generated. Since the number of data of R_1 in C_k is larger

than that of R_1 in $C_{k'}$, we choose $k = 1$ and $k' = 3$ for C_k and $C_{k'}$, respectively. By comparing the subscripts of the concept and class, we evaluate the number of data that are classified correctly or incorrectly.

First, we evaluate the quality of concepts in Table 3 using the relaxation error (RE). The RE s computed by Equation (6) for the generated concepts are plotted in Figure 5(a). As shown in the figure, the proposed MCC outperforms ECC and GCC methods in terms of the goodness of clusters, i.e., our MCC is around 30% better than ECC. This demonstrates the advantage of using the significance value λ in the model-based concept analysis in Section 2. Figure 5(a) also shows the robustness of MCC to noise. However, RE metric is not enough to measure the hierarchical structures since it considers only the overall goodness measure of concepts. In order to address this, we evaluate the quality of concepts using F-measure that analyzes the entire hierarchical tree.¹⁸ Figure 5(b) gives the quality of each method using F_m in Equation (8). From the figure, it is observed that the quality of ECC using F_m is decreasing significantly as the variance of noise gets larger unlike using RE . However, MCC is still the most robust conceptual clustering to the noise. Overall, the quality of the MCC proposed in this paper is superior to ECC and GCC in terms of the quality measured by a relaxation error and F-measure.

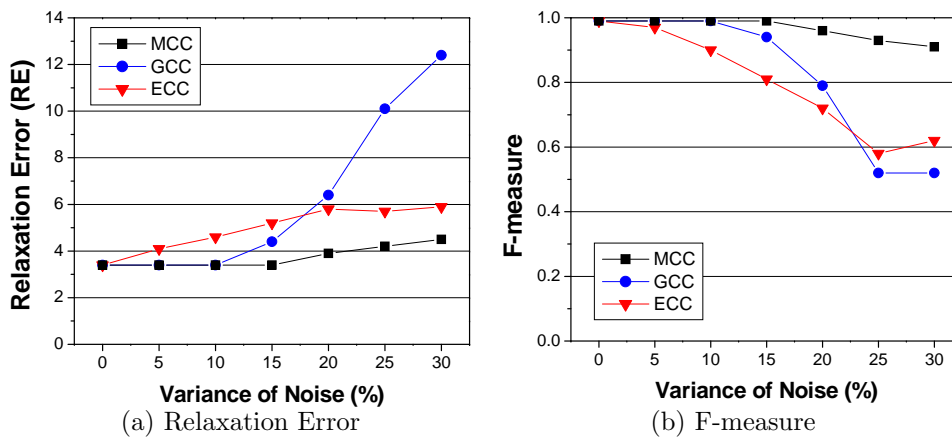


Figure 5. Quality of generated concepts using RE and F-measure

3.3. Real data sets

We apply the proposed method to a real video data set.

3.3.1. Data Description

The real video data set are obtained from video streams captured by a video camera inside of building. The main moving objects in the captured video are persons who are entering door and walking in the room. We extract total 297 moving objects by using the techniques that is based on spatio-temporal region graph.⁴ We take 4 spatio-temporal information of moving object that are x -coordinate (x , horizontal), y -coordinate (y , vertical), $size$ (s , number of pixels), and $color$ (c , average color). In this experiment, we will find their patterns, and the concepts that describe the meanings of the patterns.

3.3.2. Results

In order to generate the concepts, we apply the proposed MCC method to the real data set. For the distance measure $STED$, we compute the weight of each feature by using Equation (1). The computed weights used for $STED$ are $\omega_x = 0.34$, $\omega_y = 0.26$, $\omega_s = 0.24$, and $\omega_c = 0.16$. The model formation that is the first step of MCC generates 6 models, which are seen in Figure 6(a). Each model is visualized by using (x, y) in the figure, and the number in a parenthesis is the number of data belonging to the corresponding model. However, it is still hard to find the meaning of each model from Figure 6(a). To address it, we perform a model-based concept analysis, and generate the concept graph of the real video data set. Figure 6(b) gives the generated concept graph of

real data set. During the process of MCC, we use $T_{sim} = 0.5$ for the concept merging. Each circle and each arrow represent a conceptual node and a relational edge of the concept graph, respectively. The conceptual node includes a set of models for the extent (upper set), and a set of feature for the intent (lower set). The number on a relational edge indicates the concept similarity between two corresponding conceptual nodes computed by $ConSim$ in Definition 7.

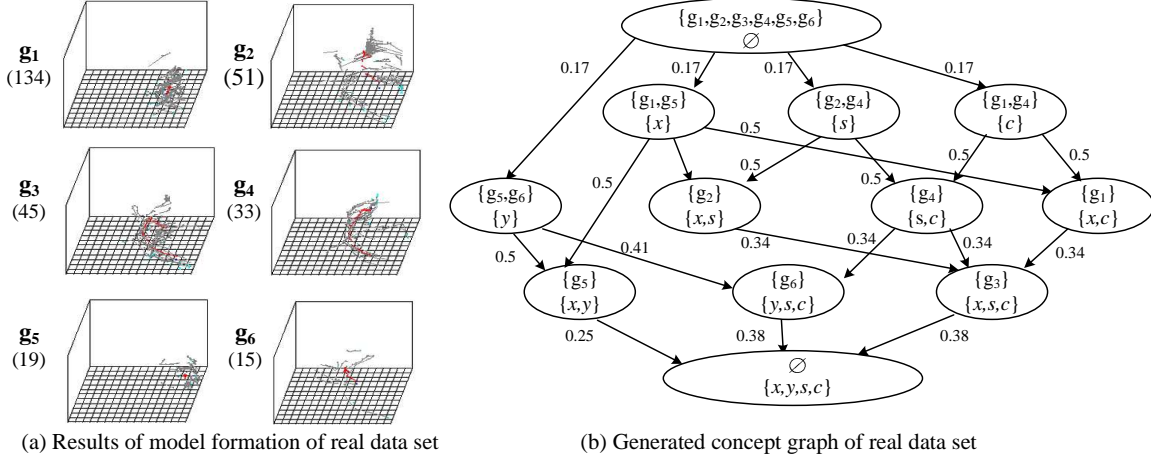


Figure 6. Results of model-based conceptual clustering of real video data set

From the concept graph, we extract concept for the video data set. Table 7 summarizes the generated concept (10) and relations (13). Using the extent and intent, we can describe meanings of the concepts. For example, C_4 is a concept that is formed by two significant features, i.e., x-coordinate (x) and color (c). C_4 includes a model g_2 whose member data are the instances of C_4 . Unlike existing conceptual clustering algorithms,^{9,10} the proposed MCC provides the relationships between two concepts. The superconcepts and subconcepts of C_4 are $\{C_1, C_2\}$, and $\{C_9\}$, respectively.

Concept	Extent	Intent	# of Instances	Super-concept	Sub-concept
C_0	$\{g_5, g_6\}$	$\{y\}$	34	-	C_7, C_8
C_1	$\{g_1, g_5\}$	$\{x\}$	153	-	C_7, C_4, C_6
C_2	$\{g_2, g_4\}$	$\{c\}$	84	-	C_4, C_5
C_3	$\{g_1, g_4\}$	$\{s\}$	167	-	C_5, C_6
C_4	$\{g_2\}$	$\{x, c\}$	51	C_1, C_2	C_9
C_5	$\{g_4\}$	$\{c, s\}$	33	C_2, C_3	C_8, C_9
C_6	$\{g_1\}$	$\{x, s\}$	134	C_1, C_3	C_9
C_7	$\{g_5\}$	$\{x, y\}$	19	C_0, C_1	-
C_8	$\{g_6\}$	$\{y, c, s\}$	15	C_0, C_5	-
C_9	$\{g_3\}$	$\{x, c, s\}$	45	C_4, C_5, C_6	-

Figure 7. Summary of generated concepts and relations

4. CONCLUDING REMARKS

In this paper, we study the problem of conceptual clustering of moving objects in video surveillance systems. A model-based conceptual clustering (MCC) using a formal concept analysis is developed. MCC consists of the following steps: ‘model formation’, ‘model-based concept analysis’, and ‘concept graph generation’. We then

generate concepts and relations for moving objects from the concept graph. Unlike existing conceptual clustering algorithms that provide only the descriptions of clusters, MCC provides not only the meanings of the generated concepts, but also the relations between two concepts. Obviously, our proposed MCC algorithm may be very attractive for other data domains, such as time-series data and spatio-temporal data. Our future work is directed towards applying the proposed schemes to real-life systems, and developing conceptual query processing.

REFERENCES

1. J. Oh and B. Bandi, "Multimedia data mining framework for raw video sequences," in *Proc. of ACM Third International Workshop on Multimedia Data Mining (MDM/KDD2002)*, (Edmonton, Alberta, Canada), July 2002.
2. J. Y. Pan, H. J. Yang, C. Faloutsos, and P. Duygulu, "Automatic multimedia cross-modal correlation discovery," in *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 653–658, ACM Press, 2004.
3. J. Lee, J. Oh, and S. Hwang, "STRG-Index: Spatio-Temporal Region Graph Indexing for Large Video Databases," in *Proceedings of the 2005 ACM SIGMOD*, pp. 718–729, (Baltimore, MD), June 2005.
4. J. Lee, J. Oh, and S. Hwang, "Clustering of Video Objects by Graph Matching," in *Proceeding of the 2005 IEEE ICME*, pp. 394–397, (Amsterdam, Netherlands), July 2005.
5. D. H. Fisher, "Knowledge Acquisition Via Incremental Conceptual Clustering," *Machine Learning* **2**(2), pp. 139–172, 1987.
6. K. McKusick and K. Thompson, "COBWEB/3: A Portable Implementation," *Technical Report No. FIA-90-6-18-2*, pp. 139–172, 1990.
7. Y. Reich and S. J. Fenves, "The Formation and Use of Abstract Concepts in Design," *Concept formation knowledge and experience in unsupervised learning*, pp. 323–353, 1991.
8. P. Cheeseman and J. Stutz, "Bayesian classification (AutoClass): theory and results," *Advances in knowledge discovery and data mining*, pp. 153–180, 1996.
9. L. Talavera and J. J. Béjar, "Generality-Based Conceptual Clustering with Probabilistic Concepts," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, pp. 196–206, February 2001.
10. W. W. Chu, K. Chiang, C.-C. Hsu, and H. Yau, "An Error-based Conceptual Clustering Method for Providing Approximate Query Answers," *Communications of the ACM* **39**(12es), p. 216, 1996.
11. R. Wille, "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," *In: Ival Rival (ed.), Ordered Sets*, pp. 445–470, 1982.
12. B. Ganter and R. Wille, *Formal Concept Analysis - Mathematical Foundations*, Springer-Verlag, Berlin - Heidelberg, 1999.
13. A. Hotho and G. Stumme, "Conceptual Clustering of Text Clusters," in *Proceedings of FGML Workshop*, 2002.
14. T. T. Quan, S. C. Hui, and T. H. Cao, "FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web," in *Proceedings of the 2004 Knowledge Discovery and Ontologies Workshop*, pp. 37–48, September 2004.
15. R. Chandramouli and V. K. Srikantam, "On mixture density and maximum likelihood power estimation via expectation-maximization," in *Proceedings of the 2000 conference on Asia South Pacific design automation*, pp. 423–428, 2000.
16. M. H. C. Law, A. K. Jain, and M. A. T. Figueiredo, "Feature Selection in Mixture-Based Clustering," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 625–632, December 2002.
17. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, 1991.
18. B. Larsen and C. Aone, "Fast and Effective Text Mining using Linear-Time Document Clustering," in *Proceedings of the fifth ACM SIGKDD*, pp. 16–22, 1999.