ASSOCIATION RULE MINING WITH SUBJECTIVE KNOWLEDGE

E.C. Kulasekere[1], K. Premaratne[1], M.-L. Shyu[1], and P.H. Bauer[2]

[1]Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33124, USA

[2]Department of Electrical Engineering
University of Notre Dame, Notre Dame, Indiana 46556, USA

ABSTRACT

In this paper, an analytical framework for association rule mining based on the Dempster-Shafer (DS) evidential reasoning is proposed. The method we propose associates itemsets in a database with basic probability assignments (bpa s) encountered in DS theory to numerically quantify the complex inter-relationships that exist among the itemsets, thus incorporating the subjective human reasoning that may otherwise be unaccounted for. In order to recast an association within this framework, measures of support and confidence in association rule mining derived via certain conditional notions are used. These measures utilize the associated subjective knowledge of the itemsets in order to discover the interesting patterns as opposed to a simple measure of frequency of occurrence of itemsets. The manner in which the frequency of occurrence is used in the existing methods also fail to capture the associations generated by the multiplicity of an item. However the method we propose uses the subjective assignment of a bpa in order to address this issue. The association rules thus formed capture the qualitative nature of the relationships among itemsets in the database which is not sufficiently well captured in the traditional data mining analysis methods.

Keywords: Association Rules, Data Mining, Subjective Knowledge, Dempster-Shafer Theory.

1. INTRODUCTION

Data management has evolved from primitive manual processing by humans to the development of database management systems with automated query processing and knowledge discovery. This process, widely referred to as knowledge discovery in databases (KDD) or data mining [1, 3, 12, 13, 14], aims at discovering useful information from very large collections of data. The knowledge that is discovered may come in the form of rules describing the properties of data, frequently occurring patterns, clustering of similar items in the database, etc. The need for efficient and effective data mining procedures has increased due to the capability of gathering large amounts of data as a result of recent technological advances. To keep up with this demand, many automated data processing methods have been adopted while sacrificing the complex subjective decision making capability of human involvement. To address this issue, we propose a framework related to the Dempster-Shafer (DS) theory [11] that can be used to simulate the subjective decision making capability of a human.

In order to illustrate our framework, we concentrate on association rule mining which is one of the central KDD tasks. Association rule mining aims at discovering interesting associations or correlation relationships among itemsets in a large data set. In general, the association rules are derived depending on the frequency of occurrence of the itemsets within a database. One major drawback of such a method is its inability to differentiate the inter-relationships among the itemsets. Such inter-relationships may provide valuable information which can influence the final outcome of the association among the itemsets. On the contrary, if the analysis is carried out by a user, the inter-relationships of the items would be taken into consideration as subjective knowledge together with the frequency of occurrence to arrive at the associations among the itemsets. That is, the degree of importance of the itemsets in a database would play a role in identifying the conditional implications among them. Therefore, the assessment of an itemset within a subjective analysis framework can yield the rules that are more realistic in its applicability. For example, consider the following scenario of a market basket analysis. The output of a frequent itemset search might have an acceptable support level to form an implication rule for paper towels from the

K.P. gratefully acknowledges the support received from the National Science Foundation through grant ANI-9726253 at the University of Miami.

P.B. gratefully acknowledges the support received from the National Science Foundation through grant ANI-9726247 at the University of Notre Dame.

linen isle and apples from the produce department.

However, it is unreasonable to assume that one can use this implication for useful rearrangement of the different products in a store since this fails to adhere to the accepted logic in ordering items. On the other hand, if the relationship between the towels and apples was considered in arriving at the association rule, it may not have achieved the necessary support level. Hence, it is evident that considering the frequency of occurrence as an initial step might result in implication rules that are logically unacceptable. Another area in which the frequency of occurrence does not represent the true situation is when the database has a multiple occurrence of an item in the same transaction or instance. This issue cannot be adequately addressed when one only considers the existence or the non-existence of an item at the time of assessing frequent itemsets. On the contrary, in a subjective reasoning environment the multiplicity of an item generates some amount of information that is useful to arrive at associations which accurately represent the situation at hand.

In this paper, our aim is to provide a framework that attempts to account for the additional subjective knowledge required to simulate human involvement in the decision making process when association rules are derived. The method we propose will explicitly account for the inter-relationships among the itemsets and the relationship generated by the multiple occurrences of the items within an instance while implicitly accounting for the frequency of occurrence of itemsets thus eliminating the drawbacks highlighted above. The proposed DS theory based framework was initiated in [6, 7]. The DS theory has been successfully used in many different areas such as automated processing [10], modeling uncertainty in decision making [2, 5], and modeling of indefinite information related to databases [8, 9]. The advantage of DS theory lies in its ability to numerically quantify knowledge or a lack thereof in an effective manner.

The organization of the presentation is as follows. In Section 2, we will provide a brief introduction to the fundamental notions associated with DS theory. In Section 3, we propose a framework for association rule mining focused on the DS theory and also highlight some of the deficiencies of the existing DS theory and their consequences. In [6, 7], conditional notions that provide ways to avoid deficiencies highlighted in Section 3 were provided. In Section 4, we demonstrate how these may be utilized in the context of association rule mining with subjective knowledge. In Section 5, we propose a new measure of information in an evidential framework and suggest new measures for the support and confidence in association rule mining in the presence of subjective knowledge. Concluding remarks are presented in Section 6.

2. PRELIMINARIES

In Dempster-Shafer (DS) theory, the relevant characteristics of the world is represented by a finite set of mutually exclusive propositions or assumptions called the *frame of discernment (FOD)* and commonly denoted by Θ. In the context of association rule mining, the FOD represents the items in a database. Then our subjective knowledge regarding the degree of support for the 2^Θ subsets of items (or the itemsets) in Θ is quantified by a function known as the *basic probability assignment (bpa).* That is, the bpa to each itemset represents the degree of importance of the co-occurrence of the individual items within the itemset. Since a bpa mass is assigned to the itemset rather than the individual items, it is free to move into any individual item within the itemset thus providing a basis for modeling ignorance and updating one's beliefs as new evidence is received. This is a major departure from the Bayesian formalism. Since no individual item in an itemset can claim the floating mass, it signifies the subjective information we have regarding the degree of belief of the items co-occurring within the itemset. In an association rule mining environment, this corresponds to our subjective knowledge of the significance of the items co-occurring within the database.

Definition 1 ***Basic Probability Assignment.*** *The function* $m : 2^\Theta \mapsto [0,1]$ *is a* basic probability assignment (bpa) *for the FOD* Θ *if it satisfies the following conditions:*
(i) $m(\emptyset) = 0$, *and*
(ii) $\sum_{A \subseteq \Theta} m(A) = 1$.

Those itemsets in a FOD Θ that possess nonzero bpas are called the *focal elements* of Θ and are denoted by $\mathcal{F}(\Theta) = \{A \subseteq \Theta | m(A) > 0\}$. The triple $\{\Theta, \mathcal{F}, m\}$ is referred to as the *body of evidence (BOE)* of the FOD Θ.

Definition 2 ***Belief.*** *The* belief *assigned to* $A \subseteq \Theta$ *in the BOE* $\{\Theta, \mathcal{F}, m\}$ *is* $Bel : 2^\Theta \mapsto [0,1]$ *where*

$$Bel(A) = \sum_{B \subseteq A} m(B). \quad (1)$$

Definition 3 ***Doubt.*** *The* doubt *regarding* $A \subseteq \Theta$ *in the BOE* $\{\Theta, \mathcal{F}, m\}$ *is* $Dou : 2^\Theta \mapsto [0,1]$ *where*

$$Dou(A) = Bel(\overline{A}). \quad (2)$$

Definition 4 ***Plausibility.*** *The* plausibility *of* $A \subseteq \Theta$ *in the BOE* $\{\Theta, \mathcal{F}, m\}$ *is* $Pl : 2^\Theta \mapsto [0,1]$ *where*

$$Pl(A) = 1 - Dou(A) = 1 - Bel(\overline{A}) = \sum_{B \cap A \neq \emptyset} m(B). \quad (3)$$

Note that $Pl(A) \geq Bel(A)$ for any $A \subseteq \Theta$. The *uncertainty interval* associated with A is taken to be $\mathrm{Un}(A) = [\mathrm{Bel}(A), \mathrm{Pl}(A)]$. The length of this uncertainty interval is denoted by

$$\ell(\mathrm{Un}(A)) = \mathrm{Pl}(A) - \mathrm{Bel}(A). \quad (4)$$

In an association rule mining environment, $\mathrm{Un}(A)$ represents the lowest and the highest 'supports' we have about the degree of co-occurrence of the items in the itemset corresponding to A.

3. ASSOCIATION RULE MINING FRAMEWORK

In this section, we present the modeling strategy for association rule mining using subjective knowledge within the DS framework. Association rule mining searches for interesting relationships among the items in a given data set. The complete set of items is represented by a FOD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$, where each θ_i $(i = 1, 2, \ldots, m)$ represents an item in a data set. For example, an item can be $\mathcal{A}$ or $\mathcal{C}$ in Figure 1. Therefore any subset of items in Θ forms the relevant itemsets used to find the conditional implication rules.

ITEMS

Jane Austen	$\mathcal{A}$
Agatha Christie	$\mathcal{C}$
Sir Arthur Conan Doyle	$\mathcal{D}$
Mark Twain	$\mathcal{T}$
P.G. Woodhouse	$\mathcal{W}$

DATABASE

Transaction	Items
1	$\mathcal{A}\ \mathcal{C}\ \mathcal{T}\ \mathcal{W}$
2	$\mathcal{C}\ \mathcal{C}\ \mathcal{C}\ \mathcal{D}\ \mathcal{W}$
3	$\mathcal{A}\ \mathcal{C}\ \mathcal{T}\ \mathcal{W}$
4	$\mathcal{A}\ \mathcal{C}\ \mathcal{D}\ \mathcal{W}$
5	$\mathcal{A}\ \mathcal{C}\ \mathcal{D}\ \mathcal{T}\ \mathcal{W}$
6	$\mathcal{C}\ \mathcal{D}\ \mathcal{T}$

Figure 1: Example database.

In Figure 1, each instance corresponds to a single transaction in the transactional database. Therefore, the task relevant data or items for the k-th instance are assumed to be in $\Theta^k \subseteq \Theta$. The inter-relationships of the itemsets associated with the k-th transaction is represented by a bpa $m^k : 2^{\Theta^k} \mapsto [0, 1]$. It will provide us with a subjective view of how different items in Θ^k are interrelated and influence the outcome of the k-th instance. Then, an association rule is an implication of the form $X \implies Y$ where $X, Y \subseteq \Theta^k$ and $X \cap Y = \emptyset$.

Note that in general it is not necessary for each transaction to span the same itemset. For example, Transaction 1 in Figure 1 is associated with four items while Transaction 6 is associated with three items. This corresponds to each k-th instance being associated with its own FOD satisfying $\Theta^k \subseteq \Theta$ such that $\Theta = \bigcup_{\forall k} \Theta^k$. The complete set of items that is under consideration for Figure 1 is given by $\Theta = \{\mathcal{A}, \mathcal{C}, \mathcal{D}, \mathcal{T}, \mathcal{W}\}$. Therefore, each transaction is associated with a FOD that is a subset of Θ. For example, $\Theta^6 \equiv \{\mathcal{C}, \mathcal{D}, \mathcal{T}\}$. Then the inter-relationships for Θ^6 can be defined by assigning numerical values to 2^{Θ^6} item combinations. For example,

$$m^6(\{\mathcal{C}\}) = m^6(\{\mathcal{D}\}) = m^6(\{\mathcal{T}\}) = 0.1,$$
$$m^6(\{\mathcal{C}, \mathcal{D}\}) = m^6(\{\mathcal{D}, \mathcal{T}\}) = m^6(\{\mathcal{T}, \mathcal{C}\}) = 0.2,$$
$$m^6(\{\mathcal{C}, \mathcal{D}, \mathcal{T}\}) = 0.1.$$

The currently available association mining methods derive the conditional implication rules depending on the frequency of occurrence of itemsets within the database. Subsequently, two measures known as support and confidence which represent the 'interestingness' of the implication rules are used to prune and extract a subset of rules that are useful. These measures are related to the joint and conditional Bayesian probabilities of the itemsets occurring in the database [3]. In the traditional method of finding the frequent itemsets, one does not account for the inter-relationships among the itemsets that might provide useful information in determining the possibility or relevance of items co-occurring within an itemset. As a result, one might arrive at the itemsets which contain items that have no reasonable logical associations among each other.

We now look at the proposed DS framework in this context and indicate how subjective reasoning can be used to avoid such a situation. In a DS framework, the frequency information associated with an itemset is implicitly embedded in the bpa. For example, the computation of belief via Definition 2 uses all subsets of the itemset A in order to compute $\mathrm{Bel}(A)$ and hence all items that support A are implicitly accounted for by the computation. The bpas of 2^Θ different itemsets of a FOD Θ satisfying Eq.(1) now numerically represent our subjective knowledge of the items co-occurring in a transaction. Hence any computation resulting from a combination of mass functions depends explicitly on the relationship it represents and implicitly on the frequency of occurrence of the items. Therefore the DS framework will produce nontrivial and conclusive results that closely simulate the subjective decision making capability of a human user. Furthermore, repeated scanning of the database is no longer necessary since both frequency and relevance information are embedded in a single bpa. In Section 5, we also show that the bpa can be used to characterize the multiplicity of an item within a transaction thus providing the complete subjective view of an item via a single measure.

As mentioned previously, the bpa provides the subjective knowledge of the inter-relationship of the items within a particular itemset. Hence to develop measures of support and confidence of association rule mining, one has to isolate the knowledge related to a given subset $A \subseteq \Theta^k$. For example, given Θ^1 we need to find the support for the itemset $\{\mathcal{A}, \mathcal{C}\}$. Such a computation will extract the relationship information that can be associated with the itemset $\{\mathcal{A}, \mathcal{C}\}$ from the complex subjective relationship represented by the entire bpa over Θ^1. However, the DS framework does not allow combination or comparison of bpa s from dissimilar FOD s. For example, the bpa representing the inter-relationships for Transaction 4 cannot be combined or compared with the corresponding bpa associated with Transaction 5 in Figure 1. A detailed description of the basis behind this inconsistency can be found in [4, 6, 7, 16]. This inconsistency is a result of the relationship in Θ^4 missing the relationship information associated with item $\mathcal{T}$ available to Θ^5. Hence, a conditional framework is essential to isolate the relevant information that is common to the FOD s thus allowing a comparison. This is the topic of next section.

4. CONDITIONAL NOTIONS IN DS FRAMEWORK

In order to proceed further we now present the notions of conditional belief and plausibility applicable within the DS framework [4, 6, 7]. The work in [6] uses a conditional bpa to isolate the relevant itemset and the support each BOE allocates for it. This provides the foundation to define conditional relationships that exist among the itemsets. The conditional belief and plausibility are given by Theorem 1.

Theorem 1 ***Conditional Belief and Plausibility.*** *The* conditional belief *and* conditional plausibility *assigned to* $B \subseteq A \subseteq \Theta$ *in the BOE* $\{\Theta, \mathcal{F}, m\}$ *are respectively* $Bel(B|A) : 2^A \mapsto [0, 1]$ *and* $Pl(B|A) : 2^A \mapsto [0, 1]$ *where*

$$Bel(B|A) = \frac{Bel(B)}{Bel(B) + Pl(A - B)};$$
$$Pl(B|A) = \frac{Pl(B)}{Pl(B) + Bel(A - B)}. \qquad (5)$$

The corresponding *conditional uncertainty interval* is defined as $\mathrm{Un}(B|A) = [\mathrm{Bel}(B|A), \mathrm{Pl}(B|A)]$ for $B \subseteq A$. The *joint belief and plausibility* are taken to be the 'fair' portion of belief and plausibility that one can allocate to $B \subseteq A$ in Θ if our original view was restricted to just the evidence associated with A [6]:

$$\mathrm{Bel}_A(B) = \mathrm{Bel}(A)\,\mathrm{Bel}(B|A);$$
$$\mathrm{Pl}_A(B) = \mathrm{Pl}(A)\,\mathrm{Pl}(B|A). \qquad (6)$$

The corresponding *conditional uncertainty interval* is defined as $\mathrm{Un}_A(B) = [\mathrm{Bel}_A(B), \mathrm{Pl}_A(B)]$.

To illustrate where these notions may become useful, consider two FOD s associated with the two transactions $\Theta^4 = \{\mathcal{A}, \mathcal{C}, \mathcal{D}, \mathcal{W}\}$ and $\Theta^5 = \{\mathcal{A}, \mathcal{C}, \mathcal{D}, \mathcal{T}, \mathcal{W}\}$ from Figure 1. Any reasonable comparison of the inter-relationships from FOD s Θ^4 and Θ^5 has to be limited to the items in the common portion $\Theta^4 \cap \Theta^5$. The inter-relationships of Θ^4 do not account for the missing item $\mathcal{T}$ which exists in Θ^5 and hence should not be considered for the comparison. In fact, if the relational information is computed using the entire FOD by ignoring the fact that some of the items are missing from one FOD, the result one obtains assumes that Θ^4 deems the relationships generated by the missing item $\mathcal{T}$ is irrelevant. Such an assumption is clearly incorrect and will yield inconsistent results at the mass combination stage [6]. However, if we compute the relational information within each FOD restricting our view to the common items which accounts for the contents of the original sets, the comparison that is drawn now is more sensible. For example, considering $X = \{\mathcal{A}, \mathcal{C}\}$ and $Y = \{\mathcal{D}, \mathcal{W}\}$ where $X \cap Y = \emptyset$ together with Transactions 4 and 5 in Figure 1, the relevant information that can be compared without inconsistencies comes from

$$X, Y \subseteq \Theta^4 \text{ and } \mathrm{Un}(X), \mathrm{Un}(Y),$$
$$X, Y \subseteq \Theta^5 \text{ and } \mathrm{Un}_{\Theta^4}(X), \mathrm{Un}_{\Theta^4}(Y).$$

We now proceed to use these conditional measures for belief and plausibility to present measures for the confidence and support of an association.

5. CONFIDENCE AND SUPPORT MEASURES

In DS theory literature, causes of uncertainty have been identified to roughly belong to two categories referred to as randomness and non-specificity. The randomness uncertainty arises due to the incapacity of differentiating among different items [15]. This type of uncertainty is associated with the Bayesian portion of a given BOE. In an association mining environment, the degree of support for individual items signifies its relevance when numerical values are computed for the relationships. Hence, for items with multiplicity, higher mass values would be assigned thus the resulting impact on relational information is higher.

The second form of uncertainty is non-specificity that arises due to the incapacity of differentiating among the factors within each itemset. This is the non-Bayesian portion of the BOE. For example, the bpa assigned to $\{\mathcal{C}, \mathcal{D}\}$ numerically represents our subjective view of items $\{\mathcal{C}\}$ and $\{\mathcal{D}\}$ co-occurring. The higher the mass assignment for an itemset, the higher its relational information is.

In general, relational information represented in a BOE is a combination of both randomness and non-specificity. In our framework, we monopolize the qualitative nature (instead of quantitative aspect) of the information to characterize the subjective knowledge associated with any item in the database. Hence at the time of information extraction, the relational information represented via the bpa s simulates human reasoning to a certain degree. We illustrate the salient features of the DS modeling strategy described above via the bpa for Transaction 2 in Figure 1:

$$m^2(\{\mathcal{C}\}) = 0.2,\ m^2(\{\mathcal{D}\}) = 0.1,$$
$$m^2(\{\mathcal{W}\}) = 0.1,\ m^2(\{\mathcal{C},\mathcal{D}\}) = 0.35,$$
$$m^2(\{\mathcal{D},\mathcal{W}\}) = 0.15, m^2(\{\mathcal{C},\mathcal{D},\mathcal{W}\}) = 0.1.$$

Observe the following:

1. The relatively higher mass assigned to $\{\mathcal{C}\}$ in comparison to $\{\mathcal{D}\}$ and $\{\mathcal{W}\}$ represents the significance of the multiplicity of item $\mathcal{C}$. Uncertainty due to randomness is generated by this assignment.

2. The higher values of mass assigned to $\{\mathcal{C},\mathcal{D}\}$ represents a stronger inter-relationship between $\{\mathcal{C}\}$ and $\{\mathcal{D}\}$. Uncertainty due to non-specificity is generated by this assignment.

3. In general, not all relationships are possible. For example $\{\mathcal{C},\mathcal{W}\}$ is not a focal element in the above assignment. This indicates that the relationship is not possible in the given BOE.

Hence it can be seen that one bpa function can be used to represent two different kinds of subjective information associated with the inter-relationships of the items in a transaction. In [6], the following measures are introduced:

Definition 5 ***Joint Uncertainty Measures for Randomness and Non-Specificity.*** *The* joint uncertainty measures *given A associated with* randomness *and* non-specificity *assigned to exactly $B \subseteq A \subseteq \Theta$ in the BOE $\{\Theta, \mathcal{F}, m\}$ are defined as*

$$R_A(B) = \sum_{C \subseteq A} \sum_{D \subseteq B} Bel_C(D) \log \frac{2}{Bel_C(D) + Pl_C(D)};$$
$$N_A(B) = \sum_{C \subseteq A} \sum_{D \subseteq B} Pl_C(D) \log \frac{1}{1 - \ell[Un_C(D)]}. \tag{7}$$

A measure that accounts for the complete relational information that can be associated with a BOE $\{\Theta, \mathcal{F}, m\}$ can be defined via the function $H_A(B) = R_A(B) + N_A(B)$, where $H_A : 2^{\Theta} \mapsto [0, \infty)$. Further details appear in [6].

The information measures derived above can then be used to find suitable expressions for the support and confidence of an association within the DS framework. Considering Transactions 4 and 5 shown in Figure 1, we have $\Theta^5 = \{\mathcal{A}, \mathcal{C}, \mathcal{D}, \mathcal{T}, \mathcal{W}\}$. As is the case in the DS framework, suppose each instance is associated with a bpa $m^k : 2^{\Theta^k} \mapsto [0,1]$ providing us with a subjective view of how different items are interrelated and influence the k-th transaction. To recast an association rule within the framework proposed in this work, for any proper subset $C \subseteq \Theta$ with $C = \bigcap_{\forall k} \Theta^k$, the *support* and *confidence* of the association rule $X \implies Y$ may be quantified via [6]

$$\text{support}(X \implies Y) \equiv \sum_k H_C^k(X \cup Y);$$
$$\text{confidence}(X \implies Y) \equiv \sum_\ell \frac{H_C^\ell(Y)}{H_C^\ell(X)}, \tag{8}$$

where the summation is over all k instances and all the instances ℓ that have an implication on X, respectively.

Contrary to the conventional association rule mining where comparisons are performed in terms of frequency of occurrence, these definitions of *support* and *confidence* take one's subjective assessments into account. In the traditional association rule mining approach, these measures are defined within the confines of a Bayesian probability space. Correspondingly, if the FOD s have bpa s that are Bayesian, the resulting DS framework indicates that there are no inter-relationships among the items under consideration.

The support and confidence functions can now be used to consider rules that satisfy the given minimum support and confidence thresholds as *strong* with respect to $X \implies Y$.

6. CONCLUSION

In this paper, we have provided an alternative to the conventional association rule mining that is based on frequency of occurrence as a basis for identifying the rules. The proposed analytical framework that is based on the DS theory presents a novel way of incorporating subjective knowledge into the association rule mining process. Such a framework is capable of simulating the subjective analysis environment of a human user thus providing a qualitative basis for extracting association rules rather than the more quantitative approaches proposed in the existing literature.

In the conventional association rule mining, since only the frequency of occurrence of the itemsets and the existence or the non-existence of an item at the time of assessing frequent itemsets (the initial and the most important step in association rule mining) are considered, it is entirely possible to derive inconsistent or illogical associations from the data. The primary causes of this are its inability to effectively gauge the inter-relationships among items and the multiplicity of an item when forming the frequent itemsets. On the other hand, the proposed framework avoids such issues by explicitly quantifying the subjective information we have regarding the inter-relationships within the items and the multiplicity of an item in a transaction via bpa s that implicitly account for the frequency of occurrence. Hence, the association rules derived from such a framework is consistent with logical human reasoning. A more general framework of extracting relational information from a subjective environment can be developed with the conditional framework and the measures of support and confidence of association rule mining mentioned above.

7. REFERENCES

[1] D. Cai, M.F. McTear, and S.I. McClean, "Knowledge Discovery in Distributed Databases using Evidence Theory," *Int. J. Intelligent Syst.*, 15(8):745-761, July 2000.

[2] L.M. De Campos and J.F. Huete, "A New Approach for Learning Belief Networks using Independence Criteria," *Int. J. of Approx. Reasoning*, 24:11-37, 2000.

[3] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2000.

[4] F. Janez and A. Appriou, "Theory of Evidence and Non-Exhaustive Frames of Discernment: Plausibilities Correction Methods," *Int. J. Approx. Reasoning*, 18:1-19, 1998.

[5] A.N. Kaplan and L.K. Schubert, "A Computational Model of Belief," *Artificial Intelligence*, 120:119-160, 2000.

[6] E.C. Kulasekere. *Representation of Evidence from Bodies with Access to Partial Knowledge.* Ph.D. dissertation, University of Miami, Coral Gables, Florida, USA, August 2001.

[7] E.C. Kulasekere, K. Premaratne, M.-L. Shyu, and P.H. Bauer, "Conditioning and Updating Evidence," in review, *International Journal of Approximate Reasoning*, 2002.

[8] E.-P. Lim, J. Srivastava, and S. Shekhar, "Resolving Attribute Incompatibilities in Database Integration: An Evidential Reasoning Approach," *Proceedings of the 10th International Conference on Data Engineering*, pp. 154-163, Feb. 14-18, 1994.

[9] E.-P. Lim, J. Srivastava, and S. Shekhar, "An Evidential Reasoning Approach to Attribute Value Conflict," *IEEE Trans. on Knowledge and Data Eng.*, 5(8):707-723, Oct. 1996.

[10] R.R. Murphy and D. Hershberger, "Handling Sensng Failures in Autonomous Mobile Robots," *The Int. J. Robot. Research*, 18(4):382-400, Apr. 1999.

[11] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 41, William Street, Princeton NJ 08540, 1976.

[12] M.-L. Shyu, S.-C. Chen, and R.L. Kashyap, "Discovering Quasi-Equivalence Relationships from Database Systems," *ACM Eighth Int. Conf. on Information and Knowledge management (CIKM'99)*, pp. 102-108, Nov. 1999.

[13] M.-L. Shyu, S.-C. Chen, and R.L. Kashyap, "A Generalized Affinity-based Association Rule Mining for Multimedia Database Queries," *Knowledge and Information systems (KAIS): An International Journal*, 3(1):319-337, Aug. 2001.

[14] M.-L. Shyu, S.-C. Chen, and C.-M. Shu, "Affinity-based Probabilistic Reasoning and Document Clustering on the WWW," *24th IEEE Computer Society International Computer Software and Applications Conference*, pp. 149-154, Oct. 2000.

[15] R.R. Yager, M. Fedrizzi, and J. Kacprzyk. *Advances in Dempster-Shafer Theory of Evidence.* John Wiley and Sons, Inc., 1994.

[16] J. Zhang, E.C. Kulasekere, K. Premaratne, and P.H. Bauer, "Resource Management of Task Oriented Distributed Sensor Networks," *IEEE Symp. Corcuits and Syst., ISCAS-2001*, 3:513-516, May 2001.