Fuzzy Ontology based Approach for Flexible Association Rules Mining

Alsayed M. H. Moawad
Computer Science Department
College of Computer and Information Technology
Arab Academy for Science, Technology & Maritime Transport
Cairo, Egypt

Ahmed M. Gadallah
Computer Science Department
Institute of Statistical Studies and Research
Cairo University
Giza, Egypt

Mohamed H. Kholief
Information Systems Department
College of Computer and Information Technology
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt

***Abstract*—Data mining is used for extracting related data. The association rules approach is one of the used methods for analyzing, discovering and extracting knowledge and mining the relationships among raw data. Commonly, it is important to understand and discover such knowledge directly from huge records of items stored in a relational database. This paper proposes an approach for generating human-like fuzzy association rules based on fuzzy ontology. It focuses on enhancing the process of extracting association rules from a huge database respecting a predefined domain fuzzy ontology. Commonly, association rules mining based on crisp ontology is found to be more flexible than classical ones as it considers the relationships between concepts or items. Yet, crisp ontology suffers from the problem of information losing resulted from the rigid boundaries of crisp relationships, which are approximated to be 0 or 1, between concepts. In contrast, the smooth boundaries of fuzzy sets make it able to represent partial relationships that range from 0 to 1 between concepts in an ontology in a more flexible human-like manner. Consequently, generating fuzzy association rules based on fuzzy ontology makes it more human-like and reliable compared with other previous ones. An illustrative case study, on two different data sets, shows the added value of the proposed approach compared with some other recent approaches.**

***Keywords*—Fuzzy Ontology; Crisp Ontology; Data Mining; Fuzzy Association Rule**

I. Introduction

The increasing use of databases in different scientific and business fields resulted in huge amounts of stored data. Analysing and understanding this data are needed to extract important information by finding unsuspected relationships among observed data sets, and summarise the data to be understandable and useful to the decision makers [1]. Data mining literature has focused on the issue of developing new techniques that successfully extract information from the vast amounts of data accumulated in large databases in order to achieve the data analysis and machine learning [2].

An ontology is "a specification of a conceptualization" [3]. It provides a shared and common understanding among people and systems. It facilitates defining the relationships between terms and concepts in a given domain. Consequently, fuzzy ontologies were introduced to represent the relationships between terms and concepts in a human-like manner.

Commonly, an ontology can be defined as "the conceptualization of a domain into a human understandable, machine readable format consisting of entities, attributes, relationships, and axioms" [4]. In other words, an ontology can be defined as the knowledge representation and common understanding of a domain. On the other hand, fuzzy ontology represents uncertain information which generally exists in several domains in a human understandable format, and translates human brain into a machine understandable form [5]. Generally speaking, data mining is used to extract valuable knowledge from huge amounts of data respecting the natural relationships between the domain terms and concepts [6]. The imprecise nature of fuzzy logic, compared with crisp logic makes it more flexible and subjective. Using fuzzy logic, data mining techniques and ontology as the base core of this work make it more flexible and human-like.

This paper proposes an enhancement approach to extract association rules based on fuzzy ontology. The rest of this paper is organised as: Section II presents a background. Related works is addressed in section III. In consequence, section IV presents the proposed data mining approach based on fuzzy ontology. An illustrative case study is given in section V. Consequently, section VI presents a comparison between the proposed approach and the Extended SSDM approach. Finally, the conclusion is presented in section VII.

II. Background

This section gives a brief overview about some related aspects of this work including association rule extraction, crisp and fuzzy ontologies and fuzzy against crisp sets.

A. Association rule Extraction

Commonly, the main objectives of data mining are of two kinds: (1) predictive and (2) descriptive. The predictive objective is the process of predicting the value of a specific attribute respecting the values of other attributes. On the other hand, the descriptive objective is concerned with extracting patterns (association rules, trends, clusters, classification rules … etc.) in order to summarise the relationships among the underlined data sets [7].

Association rule mining is one of the most important data mining approaches that aim to extract relationships or local dependencies between items in a given dataset in the form of patterns [8].

Assuming that D stands for a Database, T for Transactions, where each transaction contains a set of items T ∈ D, the form of association rule is: X → Y , where X and Y are fuzzy items in the database; where, X, Y ∈ D and X ∩ Y = ∅. The accuracy of a rule X → Y can be measured by a support measure that can be computed as in (1).

$$Support = \frac{frq(X,Y)}{N} \qquad (1)$$

where, N represents the total number of transactions in the database.

On the other hand, the confidence of an association rule is computed as in (2). A rule X → Y is interesting or satisfied in the set of transactions T with a confidence factor (c) if there is at least c% of the transactions in T that satisfy X also satisfy Y. Accordingly, while the support is a measure of statistical significance, the confidence is a measure of the strength of the rule.

$$Confidence = \frac{frq(X,Y)}{frq(X)} \qquad (2)$$

Generally, an association rule is accepted if its support and confidence are greater than or equal to predefined thresholds namely min-support and min-confidence, respectively. Such rules or subsets of associated items are called frequent item sets. The main objective of the mining process is to find all such satisfied or interesting rules that match the threshold [8].

B. Crisp versus fuzzy ontologies

Classical set theory characterises a set in which all elements take a binary or Boolean {0, 1}. Crisp has discrete terms, it takes only one of two values, for example it takes either 0 or 1, true or false, white or black, but fuzzy takes unlimited number of values in interval [0,1]. Practically speaking, a fuzzy set fits transitional rather than Boolean. Fuzzy and classic logic are not competitive, but complementary. Fuzzy system reflects how people think and translates human brain experiences into machine rules, it has the ability to develop uncertain domains [9, 10].

Commonly, the domain or the universe of discourse of a fuzzy set is the range of all possible values for an input to a fuzzy system. A fuzzy set allows its members to have different grades of membership values in the interval [0,1] as presented in (3). A fuzzy set A on a domain U, is defined by a membership function μ from U to a value in [0, 1]. On the other hand, the support of a fuzzy set F is the crisp set of all points in the universe of discourse U with non-zero membership degrees [9].

$$A\colon \mathrm{U} \rightarrow [0,1] \qquad (3)$$

Ontology specifies the concepts, relationships, and other distinctions that are related to modelling a domain to be shared between users [4, 11]. In consequence, fuzzy ontology allows each object to be related to other objects in the ontology with a matching degree based on the fuzzy set theory invented by Zadeh [1]. The fuzzy membership value μ is used for measuring the relationship between the objects or concepts in specific domain, where $0 < \mu < 1$, and μ corresponds to a fuzzy membership relation such as "low", "medium", or "high" for each object.

The strength of fuzzy logic against classical crisp one is its simplicity and flexibility when dealing with uncertainty. Commonly, when it is necessary to represent parameters of a model whose values are incomplete, vague or uncertain, then fuzzy logic represents a reliable solution. In fuzzy logic, unlike standard conditional logic, the truth of any statement is a matter of degree. Consequently, the power or cardinality of a finite fuzzy set A is given by the sum of the membership degrees of the elements belonging to fuzzy set A [9]. That is symbolically defined as in (4). Since an element can partially belong to a fuzzy set, a natural generalization of the classical notion of cardinality is to weigh each element by its membership degree, which resulted in the following formula for cardinality of a fuzzy set:

$$|A| = \sum \quad \mu_A(\chi), \ \forall \ \ \chi \in \Omega \qquad (4)$$

where, |A| is called the sigma-count of A.

III. Related Work

The work described in [4], which uses ontology to improve support in rule mining, is an example of an approach that considers semantic information during the pre-processing step. In that work, data are raised to more generalised concepts according to the ontology, and then the mining process is performed by a conventional association rule mining algorithm, like Apriori [8]. The authors argue that previous data generalization makes it possible to consider subcategories in support calculation, generating rules with higher support. Furthermore, obtained rules can be easier to interpret, since they contain high level concepts that represent richer information than specific terms in the database.

On the other hand, a relevant work has focused on the post-processing step. In [3], for example, domain knowledge is used to generalise low level rules discovered by usual rule mining algorithms, in order to obtain fewer and clearer high level rules. The authors used ontologies to generalise the objects or concepts in rules after applying the algorithm of data mining, and then they applied the data mining algorithm again to discover the high level in the abstract rules.

Another example is described in [12], where ontologies are employed to determine rule interestingness. This is done by verifying whether discovered rules confirm, contradict or reveal new information when compared to the knowledge available in the ontology. Furthermore, the author also proposed feedback mechanisms to update domain knowledge from generated rules, because new and interesting insights can be discovered from the results of the mining process.

Other approaches, like ExCIS [13], use domain knowledge in both pre-processing and post-processing steps. In this work, the pre-processing step uses an ontology to guide the construction of specific data sets for particular mining tasks. The next step is the application of a standard mining algorithm which extracts patterns from these datasets. Finally, in the post-processing step, mined rules may be interpreted and/or filtered, as their terms are generalised according to an ontology. Therefore, semantic information used in ExCIS supports dataset preparation and allows reducing the volume of extracted patterns.

In summary, the refereed work has used ontologies mainly as concept hierarchies or taxonomies, focusing on generalization relationships between concepts. Such background knowledge was used in order to obtain a reduced number of rules that are more interesting and understandable to the end user. Although domain knowledge has an important role to improve mining results, one bottleneck faced by aforementioned approaches is that the conceptual formalism supported by classic ontology may not be sufficient to represent uncertain information found in many applications [5]. This is because general ontologies contain crisp inter-concept relations and cannot quantify the strength of a relation. According to Wallace and Avrithis [14], relations between real life entities are always a matter of degree, and are, therefore, best modelled using fuzzy relations. For this reason, it is suitable to incorporate fuzzy logics into domain knowledge in order to handle data uncertainty. Thus, some association rule mining approaches have been using fuzzy concepts in taxonomies or concept hierarchies so that the membership degree can be considered when computing support and confidence of association rules.

Chen, Wei and Kerre's work presented in [15] focuses on the matter of mining generalised association rules based on fuzzy taxonomic structures. While conventional taxonomies have a child belonging to its ancestor with degree 1, on fuzzy taxonomies a child can belong to its ancestor with degree μ($0 \leq \mu \leq 1$). The authors extended the algorithm proposed by Srikant and Agrawal [16] so that the computation of support and confidence could be applied in a fuzzy context. After that, Chen and Wei have developed another work [17], where linguistic hedges were also combined in mining fuzzy rules to express more meaningful knowledge.

Another work that also considers fuzzy logic, taxonomies and data mining is described by Hong, Lin and Wang [18]. The algorithm proposed by them integrates fuzzy set concepts and generalised data mining to find cross-level interesting rules from quantitative data. In order to accomplish that, item quantities are transformed into fuzzy sets; and fuzzy rules are generated by modifying Srikant and Agrawal's method [16] to manage hierarchical fuzzy items. Association rules are said to be cross-level because quantitative items may belong to any level of the given taxonomy. Since mined rules are expressed in fuzzy linguistic terms belonging to different semantic levels, information can be more natural and easily understandable by users.

In [19] the work focuses on using fuzzy ontology in the terrorism domain to extract the events of terrorism for example, victims, date, places, time, and tactics. Another work [20] focuses on the matter of mining association rule in transaction table in relational database that uses SQL by association rule algorithm (Apriori) in K-way method to compute frequent item sets. This study seeks to remove the self-joining between the item and itself during generating and computing frequent item sets. The Algorithm try to avoid redundant data to decrease retrieval time and storage space.

In Extended Semantically Similar Data Miner (Extended SSDM) [21], the work focuses on using ontology as a background knowledge as well as similarity degrees between items to represent data mining rules and generalise terms during the mining process.

Although, there are a lot of enhancements of such previous works, there is still a need for more flexible human-like approaches for mining data to reach more reliable knowledge. The proposed approach in this work represents an attempt to satisfy such a need.

IV. The Proposed Association Rules Mining based on Fuzzy Ontology Model

A. Overview

This work enables the use of fuzzy ontology which represents the relationships between items and products in the underlined domain in a human-like manner. Consequently, the mining process can generate more understandable and meaningful association rules, based on fuzzy background knowledge. Figure 1 shows an overview of these steps.

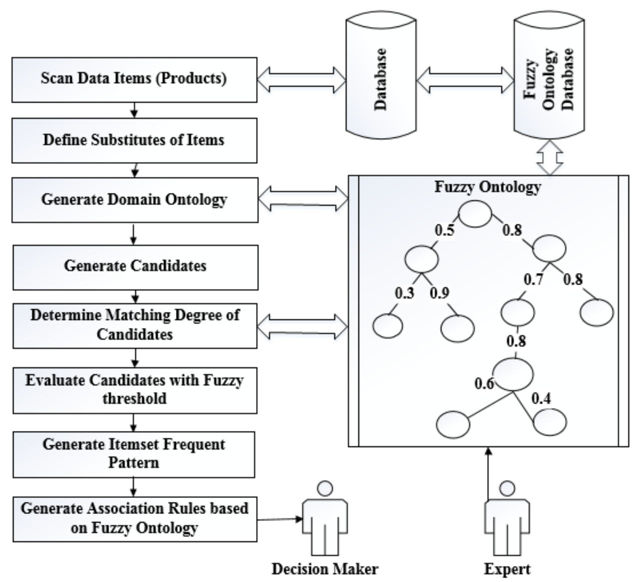

Fig. 1. The phases of the proposed approach

This work uses the fuzzy ontology to compute the similarities between the concepts as a background to Apriori algorithm which is an association rule learning algorithm for mining frequent item sets. The calculation process of frequency will depend on fuzzy rule, which means that: the count of items that happen together in the same transaction will take range from zero to one ($0 \leq \text{count} \leq 1$). The algorithm proceeds by identifying all of the items (concepts) in transactions data-set that match minimum frequency criteria (threshold). The next step is to match the list back to the single item list by transaction to identify associated item groups that meet the support criteria. These processes are repeated extending the associated item list until either the maximum list size is met or the results list is empty.

B. Processing Scenario

The proposed approach incorporates two main steps: (1) pre-processing step and (2) association rule generation step, as shown in Algorithm 1. The pre-processing step uses an ontology and fuzzy logic to determine the alternative items or substitutes for each item or concept in the dataset and the matching degree between each item and its substitutes. The next step is mining process which extracts patterns from these datasets based on fuzzy ontology to enhance the frequent pattern. Finally, in the post-processing step, mined rules may be interpreted and/or filtered, as their terms are generalised according to fuzzy ontology as shown in Algorithm 1.

Algorithm 1: The proposed fuzzy ontology based assosition rule mining algorithm.

Inputs: The domain ontology and the transaction database.

Outputs: Association rules respecting the domain fuzzy ontology.

begin of algorithm
L_1= {frequent items};

for (k= 2; L_{k-1} !=∅; k++) ***do***

begin

C_k= generate candidate itemsets from L_{k-1} (generated by joining L_{k-1} to itself);

for each transaction *T* in the dataset do

Get the matching degree for each generated candidate itemset respecting the domain fuzzy ontology.

Increment the frequency f_J of each candidate itemset *J* in C_k that are included in *T* such that :

$$f_J = f_J + \sum_{I \in T} SIM(\mathrm{J}, \mathrm{I})$$

where, $SIM(\mathrm{J}, \mathrm{I})$ represents the similarity degree between item *I* and any item included in the itemset *J*.

end

Get the frequent itemsets (k-itemset) such that:

L_k = candidate itemsets in C_k that satisfy the predefined threshold value

end

return $\cup_k L_k$;

end of algorithm

V. An Illustrative Case Study

The proposed approach is applied to a dataset of sales order [22]. Table 1 shows the definition of the transaction table. Each row represents an individual item of a transaction, which includes OrderID or TransactionID, ItemID, Quantity, Price and Total. Order no. 1 is depicted in Transaction Dataset State as an example from this case study.

TABLE. I. Transaction Dataset State

OrderID	**ItemID**	**Quantity**	**Price**	**Total**
1	a	5	10.5	55
1	d	10	5	50
1	e	2	8	16

In this case study, the used ontology represents the items and its substitutes in the domain and the relationships between them defined as fuzzy values. Table 2 presents the similarity degrees between items, as matching degrees; it will be used as a base in Apriori algorithm when computing frequent item sets.

TABLE. II. The Matching Degrees between Items

Item1	**Item2**	**Matching Degrees**
a	c	0.8
c	h	0.6
c	e	0.5
e	h	0.3
e	h	0.4

Figure 2 shows a fuzzy ontology specifying the relationhips between items in the underlined domain.

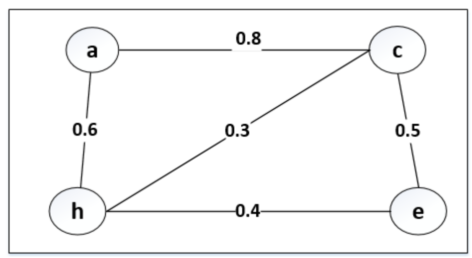

Fig. 2. A fuzzy ontology that defines the relationships between items

In this case study, the considered minimum support is 25% and the minimum confidence degree is 50% in the following cases of association rules mining:

A. Classical Data Mining

Commonly, in classical data mining, the matching degrees between items are neglected. Also, the frequency of items is counted by one. Table 3 shows the result of applying classic data mining technique to generate the association rules from the underlined dataset.

TABLE. III. THE ASSOCIATION RULES FOR CLASSIC DATA MINING

ItemsetID	Item1	Item2	Support (X U Y)	Support (X)	Conf.
1	a	d	0.29	0.37	0.78
2	c	f	0.21	0.45	0.46
3	c	d	0.35	0.45	0.77
4	e	f	0.4	0.54	0.74
5	e	d	0.31	0.54	0.57

As shown in Table 3, the frequent Itemset {a,d} has items a and d which appear together in 29% of the dataset records. Therefore, itemset {a,d} has support of 0.29 and the confidence is 78%, therefore this association rule is accepted. On the other hand, the itemset {c,f} has support and confidence values less than the specified thresholds. Accordingly, the association rule between items c and f is not accepted.

1) Crisp Ontology based Assosiation Rules Mining

In this case, the matching degrees between items and substitutes are considered to be 0 or 1. Also, the frequency of items or substitutes together are counted by one. Table (4) shows the result of association rules mining based on a crisp ontology to consider each items alternatives or substitutes.

TABLE. IV. THE RESULTED ASSOCIATION RULES USING A CRISP ONTOLOGY

ItemsetID	Item1	Item2	Support (X U Y)	Support (X)	Conf.
1	a	d	0.34	0.47	0.72
2	c	f	0.39	0.58	0.67
3	c	d	0.49	0.58	0.84
4	e	f	0.5	0.62	0.81
5	e	d	0.37	0.62	0.59

2) Fuzzy Ontology Based Assosiation Rules Mining

In this case, the matching degrees between items are considered to be in the range [0, 1]. Consequently, the frequency of items and its substitutes are counted by the predefined matching degrees. Table (5) shows the result of applying fuzzy ontology-based data mining technique to generate association rules between items.

TABLE. V. ASSOCIATION RULES BASED ON THE PROPOSED APPROACH

ItemsetID	Item1	Item2	Support (X U Y)	Support (X)	Conf.
1	a	d	0.31	0.41	0.76
2	c	f	0.35	0.52	0.67
3	c	d	0.43	0.52	0.83
4	e	f	0.46	0.65	0.7
5	e	d	0.32	0.65	0.49

Table (6) and Figure 3 shows a comparison between frequencies in three cases: (1) classical data mining, (2) crisp ontology data mining and (3) fuzzy ontology data mining, where frequencies in crisp and fuzzy ontology are greater than the classical data mining case, but in the case of fuzzy ontology the frequencies are flexible and matching human-like interpretation.

TABLE. VI. FREQUENCIES OF ITEMS IN DIFFERENT APPROACHES

ItemsetID	Item1	Item2	Classic	Crisp Ontology	Fuzzy Ontology
1	a	d	3903	3603	3803
2	c	f	2302	3353	3353
3	c	d	3853	4203	4153
4	e	f	3703	4053	3503
5	e	d	2852	2952	2452

The pairs {a,d}, {c,d} and {e,f} all meet or exceed the minimum support of 0.25, so they are frequent in the three cases. The pair {c,f} is not frequent in the case of classical data mining but it is frequent in the case of crisp and fuzzy ontology. Also the pair{e,d} is not frequent in the case of fuzzy ontology but frequent in the other cases.

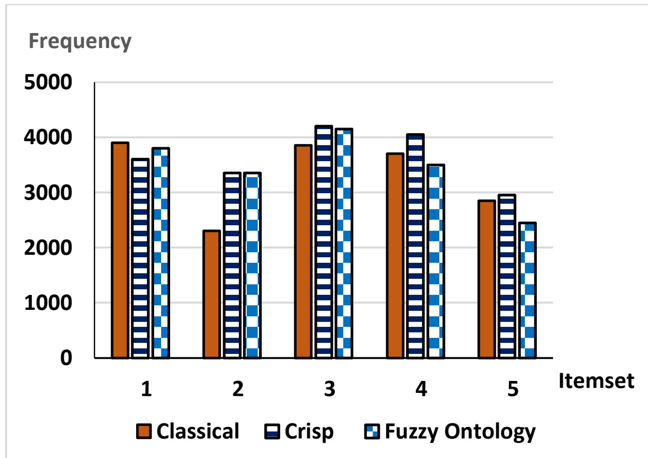

Fig. 3. The Frequencies of the itemsets with three different approaches

TABLE. VII. THE COMPUTED SUPPORT IN THREE DIFFERENT APPROACHES

ItemsetID	Item1	Item2	Classical	Crisp Ontology	Fuzzy Ontology
1	a	d	0.29	0.34	0.31
2	c	f	0.21	0.39	0.35
3	c	d	0.35	0.49	0.43
4	e	f	0.4	0.5	0.46
5	e	d	0.31	0.37	0.32

Table (7) and Figure 4 show a comparison between the computed support for the frequent itemsets in these three cases: (1) classical, (2) crisp ontology and (3) fuzzy ontology. For example, support (X U Y) for item "a" including its substitutes and item "d" with its substitutes, shows the percentage of transactions that contain both products "a" and "d".

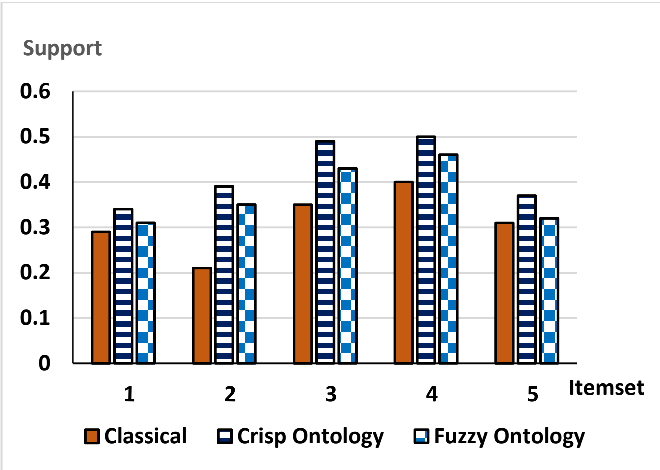

Fig. 4. The support in the three cases

On the other hand, Table (8) and Figure 5 show a comparison between confidences in the three approaches. Since confidence is the strength of implication of a rule (X U Y), so it shows the percentage of transactions that contain Y if they contain X.

TABLE. VIII. THE CONFIDENCE IN THREE CASES

ItemsetID	Item1	Item2	Classical	Crisp Ontology	Fuzzy Ontology
1	a	d	0.78	0.72	0.76
2	c	f	0.46	0.67	0.67
3	c	d	0.77	0.84	0.83
4	e	f	0.74	0.81	0.7
5	e	d	0.57	0.59	0.49

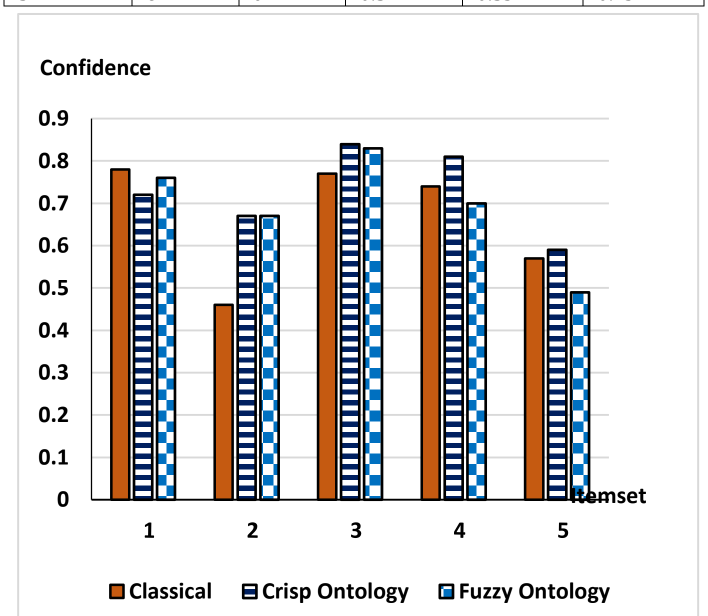

Fig. 5. The Confidence of patterns in the three cases

Based on the computing of support and confidence for the association rules that extracted from fuzzy ontology data mining as a proposed approach, crisp ontology and classical mining, it seems that the support and confidence of crisp is greater than fuzzy ontology in some results and the proposed approach is stronger than classical mining. Although these results, the fuzzy ontology data mining is better than crisp ontology mining because the crisp does not reflect the real case.

VI. THE PROPOSED ALGORITHM VS THE EXTENDED SEMANTICALLY SIMILAR DATA MINER

As mentioned before, the Extended Semantically Similar Data Miner (Extended SSDM) is an algorithm that uses fuzzy ontology in the form of similarity degrees between items to generate data mining rules and generalise terms during the mining process [21].

This section illustrates a comparison between processing scenario of the Extended SSDM algorithm and the proposed algorithm. For the comparison, the case study that was used to test the Extended SSDM [21] is considered. Table 9 shows the transactions of a supermarket that are included in the dataset.

TABLE. IX. TRANSACTIONS OF THE CASE STUDY

Transaction No	Vegetable	Meat
1	Apple	Chicken
2	Kaki	Turkey
3	Tomato	Chicken
4	Apple	Turkey
5	Cabbage	Sausage
6	Apple	Chicken
7	Tomato	Turkey
8	Apple	Chicken
9	Kaki	Chicken
10	Apple	Turkey

On the other hand, Table 10 and Figure 6 illustrate the matching degrees and the constructed fuzzy ontology of food items. The Extended SSDM requires minimum support (0.4), minimum confidence (0.7) and minimum similarity (0.7). This means that the items contained in the association rules must achieve the minimum requirements to be detected in the similarity association.

Figure 6 illustrates that there are direct connections between siblings such as the relation between Apple and Kaki with a matching degree 0.75 and Kaki, Tomato with a matching degree 0.9, also these items are connected to their parent, such as Tomato which is connected to fruit with a matching degree 0.7 and connected to vegetable with a matching degree 0.3.

TABLE. X. FUZZY SIMILARITY DEGREES

Item1	Item2	Matching Degrees
Apple	Kaki	0.75
Apple	Tomato	0.7
Kaki	Tomato	0.90
Tomato	Cabbage	0.15
Chicken	Turkey	0.85
Chicken	Sausage	0.30
Turkey	Sausage	0.10

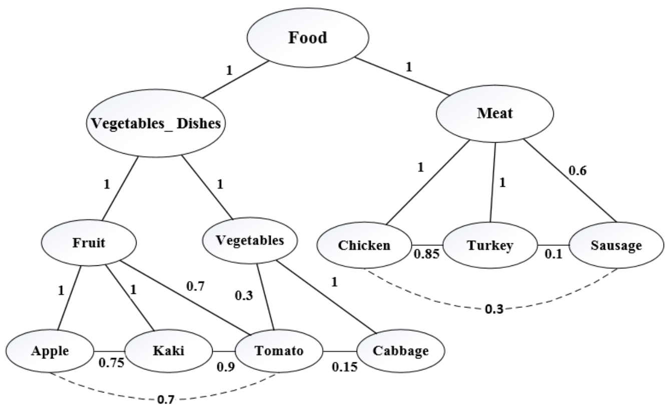

Fig. 6. Fuzzy Similarity Degrees of Food Items

The Extended SSDM considers that only sibling items can be semantically similar to one another, and it does not take into account to evaluate the semantics of non-sibling items. Therefore, the Extended SSDM ignores the matching degrees between each item and its parent.

The result of applying the proposed approach compared with the Extended SSDM is presented in Table 11, Figure 7 and Table 12.

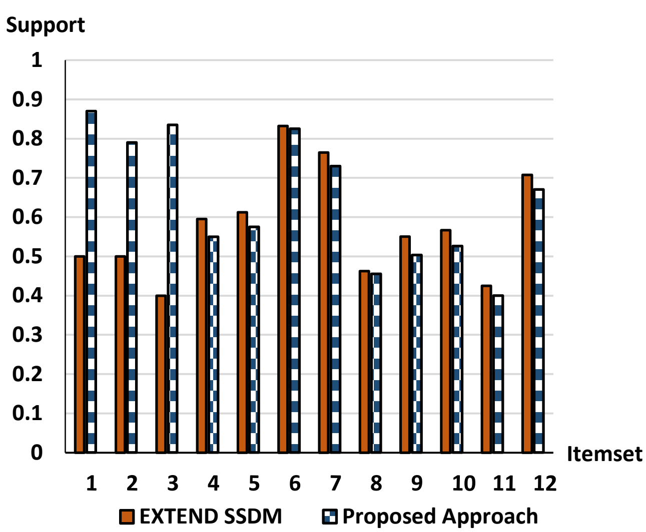

Fig. 7. Supports of Extend SSDM vs the Proposed Approach

TABLE. XI. RESULTS OF EXTENDED SSDM VS PROPOSED APPROACH

ItemsetID	Frequent Itemset	Support	
		Extended SSDM	*Proposed Algorithm*
1	{ Chicken ~ *}	0.5	0.87
2	{Apple ~ *}	0.5	0.79
3	{Turkey ~ *}	0.4	0.835
4	{Tomato ~ Apple}	0.595	0.55
5	{ Kaki ~ Apple}	0.6125	0.575
6	{ Turkey ~ Chicken }	0.8325	0.825
7	{Tomato ~ Kaki ~ Apple}	0.765	0.73
8	{ Turkey ~ Chicken , Apple}	0.4625	0.455
9	{ Turkey ~ Chicken , Tomato ~ Apple}	0.5503	0.5035
10	{ Turkey ~ Chicken , Kaki ~ Apple}	0.5665	0.52625
11	{ Tomato ~ Kaki ~ Apple, Chicken }	0.425	0.4
12	{ Turkey ~ Chicken , Tomato ~ Kaki ~ Apple}	0.7076	0.67

According to the Extended SSDM, there are some itemsets that have been ignored, but the proposed approach involved all itemsets that match the threshold. Table 11 shows all itemsets that are considered from both the Extended SSDM and the proposed approach. On the other hand, Table 12 shows all itemsets that are considered in both approaches and some additional frequent itemsets that are considered in the proposed approach. The itemsets that does not have a value for support are neglected by the Extended SSDM.

The Itemset weight corresponds to the number of its frequencies or occurrences in the transactions, the ~ symbol refers to the fuzzy ontology or similarity relation between items (item1~item2). The ~ * symbol refers to the fuzzy ontology between item and its sibling. For example, the itemset {Chicken ~ *} means the similarity relation between chicken and its brothers {Chicken ~ Turkey ~ Sausage}.

Extended SSDM considers the support of case {Tomato ~ Apple} the same as the support of case {Apple ~ Tomato}. It calculates the average support of the two cases. For example, the support of case {Turkey ~ Chicken, Apple} in proposed algorithm is 0.455 and the support of {Chicken ~ Turkey, Apple} is (0.47), but in Extended SSDM, it is resulted by finding the average of support {Turkey ~ Chicken, Apple} and support of {Chicken ~ Turkey, Apple} which is 0.4625: ((0.455 + 0.47) / 2), but in fact there is a difference between the

TABLE. XII. RESULTS OF PROPOSED APPROACH VS EXTENDED SSDM

Frequent Itemset	Support	
	Extended SSDM	*Proposed Algorithm*
{Chicken ~ *}	0.5	0.87
{Apple ~ *}	0.5	0.79
{Turkey ~ *}	0.4	0.835
{Tomato ~ *}		0.745
{Kaki ~ *}		0.755
{Cabbage ~ *}		0.13
{Sausage ~ *}		0.29
{Tomato ~ Apple}	0.595	0.55
{Apple ~ Tomato}		0.64
{Kaki ~ Apple}	0.6125	0.575
{Apple ~ Kaki}		0.65
{Turkey ~ Chicken }	0.8325	0.825
{ Chicken ~ Turkey }		0.84
{Tomato ~ Kaki ~ Apple}	0.765	0.73
{ Kaki ~ Tomato ~ Apple}		0.755
{ Apple ~ Tomato ~ Kaki }		0.79
{ Turkey ~ Chicken , Apple}	0.4625	0.455
{ Chicken ~ Turkey, Apple}		0.47
{ Turkey ~ Chicken , Tomato ~ Apple}	0.5503	0.5035
{ Turkey ~ Chicken , Apple ~ Tomato}		0.5995
{ Chicken ~ Turkey, Tomato ~ Apple}		0.514
{ Chicken ~ Turkey, Apple ~ Tomato		0.5845
{ Turkey ~ Chicken , Kaki ~ Apple}	0.5665	0.52625
{ Turkey ~ Chicken , Apple ~ Kaki }		0.59375
{ Chicken ~ Turkey, Kaki ~ Apple}		0.5375
{ Chicken ~ Turkey, Apple ~ Kaki }		0.60875
{ Tomato ~ Kaki ~ Apple, Chicken }	0.425	0.4
{ Kaki ~ Tomato ~ Apple, Chicken }		0.415
{ Apple ~ Tomato ~ Kaki, Chicken }		0.445
{ Turkey ~ Chicken , Tomato ~ Kaki ~ Apple}	0.7076	0.67
{ Turkey ~ Chicken , Kaki ~ Tomato ~ Apple }		0.69275
{ Turkey ~ Chicken , Apple ~ Tomato ~ Kaki }		0.72325
{ Chicken ~ Turkey , Tomato ~ Kaki ~ Apple}		0.6805
{ Chicken ~ Turkey , Kaki ~ Tomato ~ Apple }		0.704
{ Chicken ~ Turkey , Apple ~ Tomato ~ Kaki }		0.73825

two cases, because the number of transactions that contain both Turkey and Apple are different from the number of transactions that contain both Chicken and Apple. In this case, the average does not reflect the reality, therefore, the Extended SSDM lead to misunderstanding or unsuitable interpretation of the discovered knowledge.

Consequently, the mining process in proposed algorithm generates more understandable and meaningful association rules based on fuzzy background knowledge which is applied at both siblings and ancestors items, but the Extended SSDM ignores some association rules by applying the concept of average, also, the average of support may include outlier values for support.

If the confidence of a rule is greater than or equal to the required minimum confidence, the rule is considered valid. The Extended SSDM considers the association rule or the fuzzy item to be generalised if the association rule contains all sub-items of an ancestor. For example, the rule {Tomato ∼ Kaki ∼ Apple ⇒ Chicken} can be generalised to the ancestor Fruit, because all its descendants (Tomato, Kaki and Apple) are contained in the fuzzy item. But the fuzzy item Turkey ∼ Chicken can't be generalised to Meat, because it does not contain all Meat descendants. In fact, it is not logic to neglect the generalization of fuzzy item Turkey ∼ Chicken, although, they reflect similarity degrees with the ancestor Meat of 76.9% from all sub-items of meat.

The proposed approach considers the association rule or the fuzzy item to be generalised if the association rule contains some or all sub-items of an ancestor. The weight of this generalization is the percentage of similarity degrees between sibling items and their parent that is contained in the fuzzy item or association rules. For example, the rule {Tomato ∼ Kaki ∼ Apple ⇒ Chicken} can be generalised to the ancestor Fruit with 100% because all its children (Tomato, Kaki and Apple) are enclosed in the fuzzy item. Also, the fuzzy item Turkey ∼ Chicken can be generalised to Meat, because it contains most of Meat descendants with 76.9%, where the weight of this generalization can be calculated by dividing sum of similarity degrees of Turkey and Chicken by sum of similarity degrees of all sub-items of Meat {Chicken, Turkey and sausage}. Equation (5) is used in the proposed approach to compute the weighted generalization of each parent node in the ontology.

$$WG_{P_i} = \frac{1}{N}\sum_{J \in sub\_items\ of\ P_i} SIM(\mathrm{J}, P_i) \qquad (5)$$

where, WG refers to the weighted generalization, $SIM(\mathrm{J}, P_i)$ represents the similarity degree between a sub-item and its parent and N represents sum of all similarity degrees between sub-items and their parent.

For example, the previous rule Turkey ∼ Chicken can be generalised to their parent, i.e. Meat, with weight 0.769 by applying (5) using the similarity degrees from Figure 6.

$$WG = \frac{1+1}{1+1+0.6} = 0.769$$

The Extended SSDM depends on the similarity degrees between sibling leaf-nodes only, and ignores the similarity degrees between sub-items and its ancestor. The proposed approach depends on similarity degrees between sibling items as well as the matching degrees between these items and their parent. Therefore, although the Extended SSDM used fuzzy similarity degrees to generate association rules between items, it does not avoid interpretation mistakes that could be caused by generalization, while the proposed approach avoids these mistakes. Also, the proposed method can perform generalization even if the association rule contains all or some of the sub-items of an ancestor.

VII. CONCLUSION

Generally, data mining represents one of the most important fields of research aiming to discover the more valuable and impacting knowledge that helps in decision making and strategic planning. The association rules mining process aims to find correlations between items, products or concepts. In market analysis and planning such association rules are very crucial for managerial to best organise the correlated products and to set a more accurate ordering and marketing plan.

Some previous works which are based on crisp ontology are done aiming to reach more valuable association rules. Unfortunately, the rigid boundaries of crisp logic used to represent the relationships between concepts make some concepts fully match (in case of matching degree >= 50%) a concept and exclude other concepts (in case of matching degree < 50%). In fact such approximations cause a loss of information, it means that there is inaccuracy in computing support and confidence, where each relationship greater than 0.5 is assumed to count by 1 and others to count by 0. Also, it is not reasonable to assume a 0.5 relationship between two concepts to be fully matching while considering 0.49 relationship degree between two other concepts to be not matching at all.

So, this work presents a fuzzy ontology based approach for association rule mining in a human-like manner that enables and handles partial relationships between concepts. In other words, it considers the real relationships between concepts, specifying how much each concept is similar to other concepts. Such relationship can be represented easily through using a fuzzy ontology. Consequently, it helps to find association rules between a concept and its related concepts from one side and some other concepts and their related concepts from the other side.

The results of applying the proposed approach for fuzzy association rules mining compared with classical and crisp ontology-based mining approaches shows its added value. Commonly, the frequency in classical mining and crisp ontology-based mining is counted by 1s. On the other hand, in fuzzy ontology the frequency of substitutes are computed respecting the relationship degrees μ between the related concepts ($0 \leq \mu \leq 1$). Accordingly, the proposed approach extended the algorithm of Apriori to extract association rules based on fuzzy ontology, which is more flexible, human-like and sufficient for supporting the decision maker. It gives users more flexibility when generating association rules between items or products.

The Extended SSDM expresses semantic similarity between items to generate association rules. Unfortunately, it ignores the variations between some association rules by applying the concept of average, which leads to the problem of outlier values of support. Also, it performs the generalization only when the association rule contains all the sub-items of an ancestor. Therefore, the generalization strategy of Extended SSDM may lead to misunderstanding or unsuitable interpretation of the discovered knowledge. The proposed approach tackles such problems. Also, the proposed approach can perform generalization even if the association rule contains all or some of sub-items of an ancestor. It attempts to find the weight of the generalization using the similarity degrees between the siblings and their ancestors.

REFERENCES

[1] A. Azzalini and B. Scarpa, “Data Analysis and Data Mining”, Oxford University press, New York, 2012.

[2] A. Ali Mohamed, “Intelligent Classifier for Arabic Text Mining”, PHD thesis, Department of Computer Science, Faculty of Computer and Information, Cairo University, unpublished, 2009.

[3] Hou, X., Gu, J., Shen, X., and Yan, W., ”Application of Data Mining in Fault Diagnosis Based on Ontology”, In Third International Conference on Information Technology and Applications (ICITA’05), Sydney, Australia, pp 260–263, 2005.

[4] Chen, X., Zhou, X., Scherl, R. B., and Geller, J., “Using an Interest Ontology for Improved Support in Rule Mining”, In 5th International Conference on Data Warehous-ing and Knowledge Discovery (DaWaK), pp 320–329, 2003.

[5] Quan, T. T., Hui, S. C., and Cao, T. H., “FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web. In ECML/PKDD Workshop on Knowledge Discovery and Ontologies”, Pisa, Italy, pp 37–48, 2006.

[6] Y. Zhao, “R and Data mining: Examples and Case Studies”, Elsevier Inc, First Edition, 2013.

[7] Tan, P.N., Steinbach, M., and Kumar, V. "Introduction to Data Mining", Pearson International Edition - Addison Wesley, 2006.

[8] Agrawal, R. and Srikant, R. , “Fast Algorithms for Mining Association Rules in Large Databases”, In 20th International Conference on Very Large Data Bases, Santiago de Chile, pp 487–499, 1994.

[9] M. Dhar, “Cardinality of Fuzzy Sets: An Overview”, International Journal of Energy, Information and Communications volume 4, Issue 1, February, pp 15-20, 2013.

[10] S. A.Ghallab, N. Badr, M. Hashem, A. M. Salem and M. F. Tolba, "Fuzziness Petroleum Data Consolidation Using A Time Series Forecasting Mode", Faculty of Computer and Information Systems Information System Dept. Ain Shams University, 7th International Conference on Information Technology (ICIT), June, 2015.

[11] M. Alfonse , M. M. Aref, A. M. Salem, “An Ontology-Based System for Cancer Diseases Knowledge Management”, I.J. Information Engineering and Electronic Business, V. 6, pp 55-63, December, 2014.

[12] Pohle, C., “Integrating and Updating Domain Knowledge with Data Mining”, In VLDB PhD Workshop, Berlin, Germany, 2003.

[13] Brisson, L., Collard, M., and Pasquier, N, “Improving Knowledge Discovery Pro-cess Using Ontologies”, In International ICDM Workshop on Mining Complex Data, Houston, Texas, USA, 2005.

[14] Wallace, M. and Avrithis, Y., “Fuzzy relational knowledge representation and context in the service of semantic information retrieval”, In IEEE International Conference on Fuzzy Systems, volume 3, Budapest, Hungary, pp 1397– 1402, 2004.

[15] Chen, G., Wei, Q., and Kerre, E. E., “Fuzzy Data Mining: Discovery of Fuzzy Generalized Association Rules”. In Bordogna, G. and Pasi, G., editors, Recent Issueson Fuzzy Databases, Physica-Verlag, pp 45–66, 2000.

[16] Srikant, R. and Agrawal, R., “Mining Generalized Association Rules”, In 21th International Conference on Very Large Data Bases, Zurich, Switzerland, pp 407–419, 1999.

[17] Chen, G.and Wei, Q., “Fuzzy association rules and the extended mining algorithms”, Information Sciences - Informatics and Computer Science , 147(1-4): 201–228, 2002.

[18] Hong, T.-P., Lin, K.-Y., and Wang, S.-L., “Fuzzy data mining for interesting generalized association rules”, Fuzzy Sets Systems, 138(2):255–269, 2003.

[19] U. Inyaem, P. Meesad, C. Haruechaiyasak and D. Tran, “Construction of Fuzzy Ontology-Based Terrorism Event Extraction”, International Conference on Knowledge Discovery and Data Mining, 3, PP. 391-394, 2010.

[20] B. Ali Dbwan , "AN Enhanced K_way Method In "APRIORI" Algorithm for Mining the Association Rules Through Embedding SQL Commands", M.Sc thesis, Faculty of Information Technology, Middle East University, 2013.

[21] Eduardo L. G. Escovar, Cristiane A. Yaguinuma, Mauro Biajiz, “Using Fuzzy Ontologies to Extend Semantically Similar Data Mining”, Conference: 21st Brazilian Symposium of Databases, Florianópolis, Brazil, October, 2006.

[22] https://fusiontables.google.com/DataSource?docid=121lq6siUIcZGuxa65pxvCaVLmdyWS81psBI7qbQ.