Resolving Context Conflicts Using Association Rules (RCCAR) to Improve Quality of Context-Aware Systems

Asma Abdulghani Al-Shargabi  
Software Technology Research Laboratory (STRL)  
Faculty of Technology  
De Montfort University  
Leicester, UK  
a.alshargabi@gmail.com

Francois Siewe  
Software Technology Research Laboratory (STRL)  
Faculty of Technology  
De Montfort University  
Leicester, UK  
fsiewe@dmu.ac.uk

Abstract- Context-aware systems (CASs) face many challenges to keep high quality performance. One challenge faces CASs is conflicted values come from different sensors because of different reasons. These conflicts affect the quality of context (QoC) and as a result the quality of service as a whole. This paper conducts a novel approach called RCCAR resolves the context conflicts and so contributes in improving QoC for CASs. RCCAR approach resolve context conflicts by exploiting the previous context using Association Rules (AR) to predict the valid values among different conflicted ones. RCCAR introduces an equation that evaluates the strength of prediction for different conflicted context elements values. The approach RCCAR has been implemented using Weka 3.7 and results show the success of the solution for different experiments applied to different scenarios designed to examine the solution according to different possible conditions.

Keywords- RCCAR, Context–Aware System (CAS); Context Conflicts; Quality of Context (QoC); Association Rules(AR); Prediction.

1. INTRODUCTION

Despite the vision of Mark Wiser about ubiquitous computing was very clear for many years ago, Context-Aware Systems (CASs) still face many challenges due to sensors shortages, rapid dynamic environment, lack of harmony between different sensors and difficulties face situations capturing. Quality of Context (QoC) can be considered as the basis the CASs quality story can start with. CAS expects that context data is correct and reliable. However, sensor-rich CASs face inconsistence problems. Context conflicts may occur while collecting data from redundant context sources or while aggregating that data to compose the whole context. These conflicts could affect the produced decisions and consequently lead to undesirable actions. This situation could be serious if the system is critical.

The first definition of QoC was conducted by [1] as: "Any information that describes the quality of information used as context information". Later, the definition of QoC is modified by [2] to involve the subjective nature in the concept and engage user satisfaction to the definition. The definition of [2] was conducted as: "QoC indicates the degree of conformity of the context collected by sensors to the prevailing situation in the environment and the requirements of a particular context consumer". Many researchers have addressed the problems affect quality of context and determined context imperfection aspects. These aspects could be summarized as follows: unknown (no available sensor information), ambiguous (conflicting information from different sources) which is addressed by our paper, imprecise (information with insufficient granularity), and erroneous (sensed or aggregated context not coherent with real situation) [3][4][5][6][7][19][11].

Context conflicts reflect the contradictions within the context [13][20]. Context conflicts are classified by [13] into internal and external conflicts. Referring to [13], internal conflict is "the context conflict/inconsistency that may occur by fusing two or more context elements that characterizes the situation from different dimensions of a same observed entity in a given moment". On the other hand, external conflicts are defined by [13] as "the context conflict/inconsistency that may occur between two or more collected context data that describe the situation of an observed entity from the same point of view".

Some researches tried to resolve conflicts according to QoC parameters. That means the context value has better QoC parameters values should be selected. QoC parameters have been addressed by different approaches [1][2][3][4][7][8][9][12]. For example, the approach in [1] proposed the parameters of precision, probability of correctness, trust-worthiness, resolution, and up-to-dateness. Although there was a great interest given to the QoC parameters, ensuring that the context value is valid has not received the same attention.

The construct in CAS which is responsible of context aggregation with high quality is called Context Management Framework (CMF). CASs should be quality-aware systems in order to adapt correctly. CMF is responsible of the main functions affect context quality such as collecting sensor data, aggregating that information to compose the context and extracting high-level context information by performing reasoning operations [10][13][14][15]. Quality control is a part of quality aggregator element within CMF [10][14].

This paper introduces a novel approach called RCCAR for resolving context conflicts by exploiting the previous context to predict the valid values form conflicted ones. The technique used in this paper for prediction is Association Rules (AR). Based on association rules measures, this paper proposes a mathematical model to calculate the total affirmation from the values of context elements for investigated context element conflicted values. Then simply, the value that has the greater affirmation will be selected among the conflicted values.

The rest of this paper is layout as follows: Section 2 has been devoted to describe and formalize the proposed approach. Implementation part has been described in Section 3.
Results study has been presented in Section 4. Related work of our approach has been presented and analyzed in section 5. Finally, Section 6 conducts the conclusions and future work.

2. RCCAR Approach

Our approach RCCAR has been proposed depending on prediction of which value among the context element values are valid and which of them are not valid. RCCAR discovers all possible affirmations between context elements. These affirmations reflect the patterns that the context elements values are subjected to them according to the previous history of a context.

The philosophy behind our approach is based on some ideas. First, the pervious history of a context can help predicting what seems true or more accurate among the different conflicted values. Each context element can do that alone and also all different possible combinations of them can do that together. Adopting the overall affirmation of context elements together will lead to get more realistic context because of the context is represented by all those elements together.

For prediction purpose, data mining techniques can be utilized to resolve conflicting problem by exploiting the history of context data. One of the best techniques that can resolve this problem is the Association Rules (AR) technique. AR discovers associations which relate the values occur together. If there are three context elements -excluding the context element under investigation-, we will get a total of seven possible combinations of them with maximum value of 7 where the confidence can be within the range [0-1]. According to AR, associations have many measures to judge the strength of produced associations. Basically, association rules are produced with "support" and "confidence" measures values, where support is calculated by equation 1:

\[
Support(YZ \Rightarrow X) = \frac{\text{last-occurrences of } X,Y,Z \text{ together}}{n}
\]

(1)

Where \( n \) is the number of instances in the previous history of a context, and a confidence is calculated by equation 2:

\[
Confidence(YZ \Rightarrow X) = \frac{\text{last-occurrences of } X,Y,Z \text{ together}}{\text{last-occurrences of } X \text{ together}}
\]

(2)

We can see that this measure is very important for our problem because it presents how much the current values of \( Y \) and \( Z \) recommend the current value of \( X \) according to the last occurrences of them.

We want to indicate that the type of context which will be considered in our solution and experiments is that comes from different sources/sensors and not that will be used in upper layers of a CAS. We believe that prediction could be used for any type of conflicts however; more experiments are needed to prove that.

A. The Formulation of Proposed Approach to Resolve Context Conflicts

\( CtxtDB \) is the database contains the context instances (the previous history of context data) ordered by time stamp. \( N \) is the number of instances of \( CtxtDB \).

Let \( X \) is the context element under investigation and \( Y \) is the set contains other context elements which compose the whole context, therefore \( Y \) is derived as in equation 3:

\[
Y = \{ y_1, y_2, y_3, \ldots, y_m \}
\]

(3)

Where \( m \) denotes the number of context elements. As shown in equation 3, each \( y_i \) belongs to \( Y \) represents a particular context element. Element number does not indicate a sequence. It just identifies the context element.

Let expression 4 denotes the association rule recognizes that occurrence of \( y \) affirms occurrence of \( x \).

\[
y \Rightarrow x
\]

(4)

Each produced association rule is associated with a value reflects the strength of this association. In our problem, it is the confidence value. Associations which affirm the context element \( y \) is derived using confidence measure which is calculated by equation 5:

\[
\text{confidence}(y \Rightarrow x) = \frac{\text{support}(yx)}{\text{support}(y)}
\]

(5)

Let \( \text{Total-association} \) is the summation of confidence for all associations affirm the context element under investigation \( x \). By scanning the \( CtxtDB \), equation 6 is used for \( \text{Total-association} \) calculating:

\[
\text{Total-association}(x) = \sum_{i=1}^{\text{m}} \text{confidence}(y_i \Rightarrow x)
\]

(6)

However, equation 6 calculates the \( \text{Total-association}(x) \) by considering the associations that link only two context elements. According to association rules analysis, a collection of zero or more elements is referred to by \( \text{itemset} \). If \( \text{itemset} \) contains \( k \) items, it is called as \( k \text{-itemset} \).

Let \( y(i) \) is a collection of \( k \) context elements belong to \( k \text{-itemset} \) and \( i \) is the identifier of the instance in the respective \( \text{itemset} \). Examples of \( \text{itemset} \) are shown in TABLE 1.

<table>
<thead>
<tr>
<th>( y(i) )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_1(1) )</td>
<td>13</td>
<td>9</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>( y_2(2) )</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>( y_3(3) )</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Where \( a, b, c, \) and \( d \) are context elements. The frequency (occurrences) of context element individually and all possible combinations of them are calculated by scanning \( CtxtDB \) \( d_1(i) \) and \( d_2(i) \) denotes the number of associations for different \( \text{itemset} \). One scan of \( CtxtDB \) is enough to record the occurrences for current context elements values individually and all possible combination of them. Then, we just sum them.

We apply that to all conflicted values and select the context element that has the greater \( \text{Total-association} \) value.
To clarify that, assume Table I contains the occurrences for some context elements. These values are concluded by scanning CtxtDB. Assume that a and b are two conflicted values for a context element. In addition to a and b, the current situation of context is represented by other context elements c and d. According to occurrences in Table I, the associations illustrated in Table II will be produced.

**TABLE II. PRODUCED ASSOCIATIONS**

<table>
<thead>
<tr>
<th>Associations affirm a</th>
<th>Confidence value</th>
<th>Associations affirm b</th>
<th>Confidence value</th>
</tr>
</thead>
<tbody>
<tr>
<td>c ⊃ a</td>
<td>( \frac{8}{10} = 0.80 )</td>
<td>c ⊃ b</td>
<td>( \frac{5}{10} = 0.50 )</td>
</tr>
<tr>
<td>d ⊃ a</td>
<td>( \frac{7}{12} = 0.46 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c, d ⊃ a</td>
<td>( \frac{4}{5} = 0.80 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total − affirmation(a)</strong></td>
<td>2.06</td>
<td><strong>Total − affirmation(b)</strong></td>
<td>0.50</td>
</tr>
</tbody>
</table>

We can conclude that the possibility that the value of a is valid is greater than that for value for b. The table shows that number of associations and the total value of their confidence indicate that a is the recommended value of context element under investigation. Therefore, the Total − affirmation \( (x) \) equation should be as shown in equation (7):

\[
Total − affirmation(x) = \sum_{k=2}^{W} \frac{\sum_{i=1}^{d} \text{Confidence}(y_i(k) = x)}{d} \tag{7}
\]

Where \( W \) is the maximum number of itemsets according to occurrences in CtxtDB for the current values of context, and \( d \) is the number of associations. \( k \) is starting with 2 because the associations start with two elements. The equation shows that the associations which will be considered in the summation are just whose \( x \) in the right side. So, not all associations will be consider.

On other hand, when we talk about previous history of context, we have to be carefully when determining the period which should contribute in the prediction. It depends on the nature of context. We do not necessary need all previous history to resolve the conflicts. It depends strongly on the nature of context, if the patterns change rapidly or it is stable to some extent. For example, if we work with a context-aware system collect weather data and take some actions accordingly, there are different possibilities for the history should be selected in this system. If the climate of the area the system work in is relatively stable area, we can depend on the near history, two weeks could be enough. However, if the climate is not stable, a longer period for history would be recommended. The objective is selecting the adequate period of history which contain the hidden patterns for the context data.

3. RCCAR IMPLEMENTATION AND EXPERIMENTS

This section is devoted for clarifying how implementation of our approach RCCAR was achieved. Experiments were built according to predesigned scenarios using Weka 3.7.7 which is developed at the University of Waikato in New Zealand to implement different experiments which based on machine learning techniques [17]. It is selected as it is used from research community to support solutions related to data mining techniques.

A. Descriptions of Datasets used in Experiments

Many datasets are available, but we looked for different datasets satisfy the following requirements in order to study the successes of proposed approach: (1) we need a data produced during long period of time; (2) we need a large volume of datasets; (3) we need datasets with more than two variables, to examine the combined associations. Two or three variables are not enough.

We used two real datasets available for researchers to examine their solutions. A description of them is in the following.

*The description of Southampton monthly weather historical data (1855-1999)*:

This data is a historical data about some weather variables. It is officially collected and recorded by Southampton Weather Station and published by its website [16]. This dataset contains 1744 instance. Recorded variables are year, month, temperature max degree, temperature min degree, air frost, rainfall, sunshine hours.

*The description of Cardiff climate historical data for five days*:

This data is available on the Internet by The Met Office which is the UK's National Weather Service organization. It has a long history of weather forecasting and has been working in the area of climate change for more than two decades [18]. The variables which are available in this data are date, time, precipitation-probability(%), humidity(%), visibility, temperature, and wind speed.

B. Data Preprocessing

Before applying the proposed solution, datasets were in need for some preprocessing. These preprocessing actions are summarized as follows:

Missing values processing:
When examining data for Southampton weather historical data, a loss is observed for some instances. These missing values represented 0.009\% from 1744 instance. So, we simply eliminate them.

Data transformations:
AR required data to be a nominal. So, we transform it using Weka preprocessing functions. In addition, we transform real variables which are not limited into intervals as known in statistic. We also used this technique when dealing with new investigated value so, we cannot analyze this value definitely but the values around it.

Preparing the data as .arff format file:
Weka accepts data as .arff file and this required us to transform the data which we get into this type. At first, the data was kept as .xls file and then, following Weka instructions it was transformed into .arff file.

C. Scenarios Design
In order to test the proposed solution, we design some different scenarios (test cases) to examine the successes of the solution in different circumstances. Different possibilities were taken into consideration when these scenarios (test cases) were designed. Each scenario includes two values for the next new reading. One of them is valid and the other is not. The proposed solution should discover that according to its expectation based on the previous history of context.

**Factors forming different scenarios:**

Factors that are expected to affect the performance of the proposed solution are different. The first factor which was taken into account when designing the test cases was the depth of the history of the context. Characteristics of investigated value could have its impact; the previous occurrences of the investigated value and if the wrong value has a close value or a far one from the valid value. Number of context elements could affect the prediction strength. Moreover, we examine the results against unilateral and multilateral associations. TABLE IV summarizes the different scenarios according to different factors.

**TABLE II. SUMMARY OF DIFFERENT SCENARIOS**

<table>
<thead>
<tr>
<th>Scenario Num.</th>
<th>Volume of history</th>
<th>State of investigated context element</th>
<th>Kind of associations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>correct value has previous occurrences</td>
<td>Wrong value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Close to correct value</td>
<td>Has previous occurrences of other context elements</td>
</tr>
<tr>
<td>1</td>
<td>25 record</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>25 record</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>25 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>25 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>1744 record</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>1744 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>1744 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>1744 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>25 record</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>10</td>
<td>25 record</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

4. EXPERIMENTS RESULT

According to previous scenarios, many experiments were conducted. All of them prove that the new approach succeeds where the total-affirmation for the correct value was always greater than that for wrong one. We suppose the strength of the prediction as the difference between the values of total-affirmation for correct and wrong values. To study the effect of depth of history and the number of context elements on the result, different experiments were implemented. The dataset of Southampton weather has been used as the data for the weather is available for long history from 1855 to 1999. We applied the solution many times using different periods for history and different numbers of context elements, and finally using different combinations of context elements in the association rules. Beside the investigated element which was the maximum value of temperature, we used three other context elements for these experiments which are the minimum value of temperature, rain fall and month. According to that, the maximum number of AR affirms the context element under investigation would be seven with strength of 7 with 1 value for confidence for each. Thus, the context is represented by four elements, one of them would be the investigated one and the other would be used to predict the validity of it.

**Depth of history and strength of prediction:**

This section is devoted to introduce the results we have got for the relation between the depth of history and the strength of prediction. The following charts demonstrate the result we got Figure 1.

![Figure 1. Results of prediction of different values against different history periods](chart.png)

The previous chart shows that the difference between the predictions of the correct value and the wrong one is positive and it is very close of the prediction of the correct value. The prediction for the wrong value shows a little value very close to zero. What is worth noting here is that the prediction is high in the near past and almost constant over a certain number of years. This could be because of this phenomena change smoothly, so the near history is very close of the current time. Where after a certain number of years the pattern discovered would be constant because the value of confidence (prediction) should take into account all occurrences after this time and this sum of occurrences will affirm the pattern with the same value over years. What we want to indicate here that the prediction using AR for the near history is enough as we see that the last five days was enough taking into account that the data has the average of values for each month so the number of instances is not very large but it was enough. We assume that the number of context elements help for a strong prediction even the past history is not deep.

**Number of context elements and strength of prediction:**

We applied our approach with three context element and also with two elements and recorded our observations as shown in Figure 2.
The figure above shows that there is a big difference in the strength of prediction with using just two elements and when using three elements as we expected. This can be generalized by saying that whenever the number of elements of context higher the better the prediction would be. The strength of prediction is almost doubled when adding just one context element.

Using combinations of possible AR and strength of prediction:

In this section, we studied how much the prediction was affected by the type of AR were selected. As mentioned before, some previous researches used the total affirmation of each context element individually and we prove here that using all possible combinations collectively would be better as Figure 3 shows.

As shown in Figure 3 using all context elements make prediction stronger.

Finally, we would like to indicate that whenever the previous history was greater the prediction will be better, but we are no in need to use all previous context where the results shows good prediction with just 5 previous days (just 25 records) in our experiments; therefore we recommend determining the minimum adequate size of previous context database to reduce the time complexity.

5. RELATED WORK

Conflicts resolving was addressed from research community using different views [10][13][19][14][20]. A solution proposed by [14][20] using the known parameters described above to resolve the conflicts. It achieved that by feeding CASs with what called a context quality policy, so the solution uses different quality parameters for different CASs. Solutions in [10][13] resolve conflicts within context fusion layers of context management frameworks (CMF). This layer detects and resolves conflicts according to two quality attributes: probability of correctness and trustworthiness. In [13][10][19] a way for computing these quality attributes using Bayesian theory was introduced. To resolve conflicts the proposed approach is based on the idea that for a specific situation, the context element is not used alone, almost there is other contexts elements used with it. There is a collection of context data usually occurs together. So, this approach utilizes the previous history of the context and dependencies between context elements to increase the probability that a certain context is correct. Both [13] and [19] proposed close methods to analyze the last occurrence of piece of context and estimate the probability of correctness of that piece of context. Context dependencies relation between two context elements could be either unidirectional or bidirectional. It is affirm/contradict relation with probability rate within the range [0,1] since in real situations the degree of affirmation/contradiction is variable. These relations use a variable called δ degree which is associated with each dependence relation. It is empirically assigned or obtained. To resolve context conflicts.

Our approach is different for some reasons. First of all, using QoC parameters only for conflicts resolving is not enough. This is because of the shortages of sensors which is the basic resource of context elements values. Three basic problems can be addressed with sensors reliability: (1) the default accuracy regarding to the sensor technology type. For example, the spatial position for an object could be captured using different technology (accuracy) such as GPS, infrared and Bluetooth, each one of them have its default according to its technology. (2) The distance between the sensor and the object. Each type of sensors has different spatial range for reliable sensing. (3) Failures could be happen with hardware for different expected or not expected reasons. These failures do not necessary make the sensor out of service, sensor can be continuing with providing the data but unfortunately, with wrong values. This scenario could be happen when context-aware system is working, these systems could be critical and decisions made could be lead to serious problems. We can select the best sensor according to first and second problems. This means the higher value of reliability regarding to the accuracy defined by the manufacture and according to the distance between the sensor and the object. This could be great if the sensor works very well and does not suffer from any hardware failures. But, if the sensor has any undetected hardware problem, these two factors means nothing regarding to the reliability.

Moreover, exploiting the affirmation and contradiction of other context element for conflicts resolving were fine, but there are some reservations regarding that. Probability of correctness which is used to resolve conflicts were based on what called δ value which used to determine the degree of affirmation or contradiction of the relation. This value affects the total probability of correctness while δ is depending on calibration period for CASs. This seems not objective enough. We are in need to get more objective method to predict the best value between the conflicted values.

From another hand, the method used to calculate probability of correctness was confined just on context elements

![Figure 2](image1.png)

**Figure 2.** Prediction using different number of available context elements

![Figure 3](image2.png)

**Figure 3.** Results of prediction for different types of associations
affirmation individually and not collectively. However, we can exploit the different combinations between these context elements which embedded in previous history and know who from these affirms the context element under investigation. This means that we can get the total affirmation of the current context using both affirmations from each context element individually, and also by affirmation from different combinations between these context elements according to previous history. This would be better because there is a CASs where such context element affirm other context element but not with other context elements. It is possible that there are realistic situations where the context elements are supporting a certain context element separately but not collectively. In addition, some previous work adopt calculating the contradict values of the context and its strength to exclude it if it occurs in the current context [10], [13]. We think that we don’t need to use the associations contradict the context element values under investigation. The associations which affirm the values is enough. Moreover, contradicts for context can’t be easily counted. This Solution seems impossible and impractical especially for some types of dynamic context where patterns could change from time to time.

Finally, the method used for calculating total probability of correctness uses the average of conditional probability for context element under investigation given other context elements. Sometimes, the average leads to misleading results. If different affirmation of context elements has extremes, this will affect the average value. So, the probability of correctness for some context elements could be better or worst where it is not. Summation of affirmation would be better. Especially, that we can know the maximum value of the affirmation, if we know how much context elements will be used for prediction.

6. CONCLUSION

This paper introduces our approach RCCAR that exploiting AR technique to predict the best valid value between some conflicted values of a context element. The basic idea is to exploit all possible patterns hidden in the previous context history which combine the context elements together. The paper introduces a mathematical method calculates the total affirmation for each value of investigated context elements based on association rules as a technique. Results show that the proposed solution has succeeded against different conditions. There were some factors improve the outcomes. Using all possible combinations of context elements in AR improve the prediction. Utilizing all context elements recorded in the previous history also improves the result. Depth timeline affect the result, we recommend determining the appropriate depth according to the nature of context as in a certain point the prediction tends to be constant as adding instances just confirm the prediction with the same strength and in the same time increase the complexity where a short history could be enough. We think that this solution can be used easily as many CASs already record their data and have a huge data warehouses. Studying this approach for context management in a comprehensive framework is recommended as the general aim is making the CAS quality-aware systems.

Applying this solution on other types of context within upper layers of CMF could demonstrate good results. Finally, for future work we recommend studying how to improve the algorithm to reduce the time complexity and increase the scalability of algorithm.

REFERENCES


