Resolving Conflict of Wireless Sensor Networks Using Subjective Logic

Abstract

Currently, data sources are widely used in different computing applications. These applications use data sources as input devices and then do some processing in order to get the desired output. The garbage in garbage out principle applies here. In other words, if the source as an input device is not reliable, the output from the related computing application will not be reliable. Therefore, it is becoming increasingly important to monitor the reliability of these data sources. In our work, we assume that multiple time series are coming from multiple sources. Thus, we propose to monitor multiple deviations between multiple data sources as an indication of source reliability. We also propose to represent source reliability using subjective logic formalism.

Keywords: Source reliability; Subjective logic; Change point.

1. Introduction

Currently, data sources such as, wireless sensors that provide continuous data streams, are widely used in different computing applications [1]. These data sources are used in different areas such as health care, infrastructure, environment, military, industry, and research [2]. We assume that continuous numeric data sources send their readings to a central station. These continuous readings can be used to guide the process of decision making [3]. However, data sources can produce incorrect readings that can mislead the decision-making process. Thus, it is critical to monitor the reliability of data sources. Our monitoring system uses subjective logic in representing source reliability, in which each source's reliability is a vector of subjective logic opinions.

One of the common ways for estimating source reliability is comparing the source reading with a reference value and then if this reading equals the reference value, the source reliability value will be updated [4]. In this method, the source will be given an initial weight value that represents its reliability. This weight will be high if the source reading is consistent with the reference value. On the other hand, the weight value will be low if the source reading is not consistent with the reference value. One problem here is, we can't determine which source is the most reliable source during a specific period of time since this method is not time dependent. In our approach, we need to detect the changes in source reliability over the time period.

In our approach, source reliability is defined as the source's consistency with other sources. The source reliability is expressed as a vector of subjective logic opinions. Each opinion reflects the degree of source reliability and the certainty degree within a specified time interval. Time intervals here depend on the sources' consistency change points.

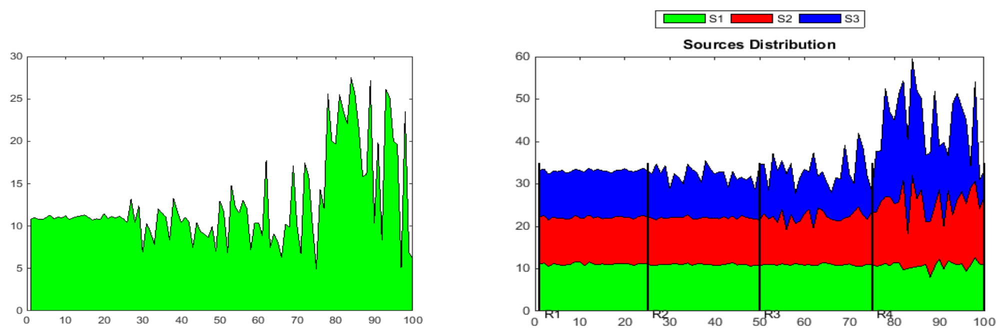
Based on this definition of source reliability, our approach is based on the detection of a change in the source behavior. There is a considerable number of works in change point detection methods. The detection methods aim to detect sudden changes in the source behavior [5]. This detection can be done using different techniques such as density-ratio estimation, comparing two distributions [6], in which past and present distributions are constructed and then compared in order to find if they are significantly different or not. Therefore, change point detection methods will be used in our approach to detect the change in the sources' consistency over the time.

However, the existing change point detection methods deal with individual time series. Figure 1.A shows an area plot for the input of a value change detection method that deals with one time series. We define source reliability as the source's consistency with other sources. Therefore, we need to detect a change in source behavior with other sources. Figure 1. B shows the input of our proposed approach, which is multiple time series coming from multiple sources.

Our reliability assessment and monitoring approach detects the change in reliability based on the change in sources consistency with respect to other sources instead of detecting the change in the reading values of each source. Our approach provides the following contributions:

- We propose a method to monitor multiple deviations between multiple data sources as an indication of source reliability.
- We also propose to represent the source reliability using subjective logic formalism. In our approach, each source's reliability is a vector of subjective logic opinions rather than just one accumulated value for each source.

- We demonstrate that our proposed approach can efficiently describe the behavior of the source reliability and can be efficiently used for monitoring source reliability.



A **B**

Fig. 1. A) Area plot for the input of value change detection method.

B) Area plot for the input for our proposed approach.

The paper is organized as the follows; in section 2; we discuss the background and related work, including source reliability, subjective logic, and change points detection methods, in section 3; we discuss our proposed approach, in section 4; we provide an experimental study, and finally in section 5; we provide our conclusion and future work.

2. Background and Related Work

2.1 Source Reliability

Source reliability can be defined as the trust of the information provided by the source, or, in other words, the probability of getting correct information from this source [7]. The source here is anything that can collect data such as sensors, publications, human observers, and so on. In order for a source to be reliable, the information provided by this source should be replicated [7].

Data reliability and data confidence are very important components of data analysis and decision-making process. Thus, unreliable data can weaken the truthfulness of the conclusion. During data collection, reliability is ensured by measurement device calibration, adapted experimental design, and statistical repetition [8]. However, the growth of big data approach and the growth of using the web allow an opportunity to collect extra details from multiple sources. Therefore, it is becoming increasingly important to develop some methods to estimate the reliability of the source before using it.

Source reliability can be inferred from several criteria including [9]; (1) source type, (2) experimental protocol and methods used to collect the data, (3) statistical procedures such as repetition, uncertainty quantification, and experimental design. There are several studies arguing that an information source is totally reliable if and only if the information it delivers is true in the real world [10, 11]. In other studies, reliable source has been defined as pedigree information [12], confidence [13], multi-criterion feature ranking [14].

Several methods are available to estimate source reliability such as, (1) evidence theory method that evaluates reliability based on choosing reliability scores that minimize the error function [15]; (2) comparing the source assessments with reference values [16], this method requires the definition of an objective error function and a fair amount of data with a known reference value.

Estimating source reliability becomes more difficult in the case of trying to quantify the reliability of data collected from human sources, in which humans work as sensors. The difficulty here is coming from the uncertain nature of human measurement that is less reliable than well calibrated and tested infrastructure sensors [17, 18].

In our approach, source reliability has the following properties:

- Source reliability is the source consistency with other sources. The reliability is expressed as a vector of subjective logic opinions.
- Every opinion reflects the degree of source reliability and the certainty degree within a specified time interval. Time intervals here depend on the consistency change points.
- Our approach doesn’t require a reference value.

2.2 Subjective Logic

Subjective logic can be defined as a type of probabilistic logic that explicitly takes uncertainty into account [19]. In general, subjective logic is suitable for modeling and analyzing situations that involve uncertainty and incomplete knowledge. For example, if you toss a coin you will be certain that you will get one out of two outcomes, which are head or tail. On the other hand, if someone says that there is a life on Venus planet. The possible outcomes can be either yes or no. However, no one can be certain that either outcome is correct since there is no complete evidence that there is a life on Venus planet. This means that there is uncertainty, and this uncertainty can't be represented by the traditional probabilistic logic. Subjective logic opinion can be used whenever there is uncertainty about the argument [20].

Subjective logic is better than the traditional probabilistic logic in real world situations [19]. Subjective logic opinion can be applied when dealing with continuous numeric stream data sources. For example, if there is a change point in the source reading stream, then the possible justification options for this change is either; an error from a faulting sensor, or a correct change that is captured by the sensor. No one can be certain which option is the real source of the change. Thus, there is an uncertainty in this situation and that is why subjective logic opinion can be applied in our approach.

The subjective logic opinion can be represented using the interior of an equal-sided triangle (Figure 2). In which the opinion about a state x is represented by a metric $< b_x, d_x, u_x >$ where bx, dx, and ux are; belief, disbelief, and uncertainty respectively. b_x, d_x, and ux ϵ [0 1] and bx + $d_x + u_x = 1$. For example, when you roll a normal six-sided die with numbers from one to six, then the subjective logic opinion about getting a number that is less than 7 is: $\omega_x<7 = <1, 0, 0>$. In this case, we are 100 % certain (uncertainty= 0) that we will get a number that is less than 7. On the other hand, the subjective logic opinion of getting a number that is greater than or equal 7 is $\omega_x>=7 = (0, 1, 0)$. In this case, we are 100 % certain (uncertainty= 0) that we will not get a number that is greater than or equal 7. One more example is about the life in Venus. In this example, no complete evidence that there is a life in Venus or not. So, the opinion about the life on Venus can be represented by $\omega_{life\_Venus} = (0.4, 0.35, 0.25)$. There is uncertainty about the life on Venus.

As shown in Figure 2, the triangle vertices represent the uncertainty, belief, and disbelief. The belief line starts from the middle of the edge that connects the other two parameters vertices, which is the edge that connects the uncertainty and the disbelief. This also applies to the uncertainty and disbelief lines. Each line for belief, disbelief, and uncertainty starts with 0 and ends with 1 at the vertex that corresponds to this line. Figure 1 shows how to put $\omega_x = (0.7, 0.1, 0.2)$[21].

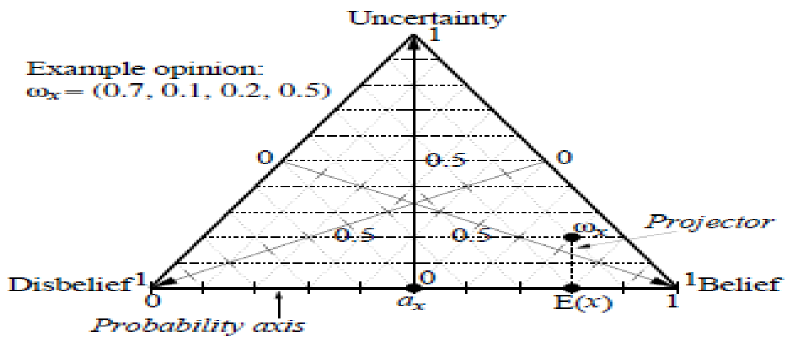


Fig. 2. Opinion triangle with example [21]

The parameter ax is the base rate of x. For example, in tossing an unbiased coin, the probability of getting a head is 0.5. Thus, a_{head} is 0.5. However, if the coin is biased and has a certain cutting edge that increases the probability of getting a head, then a_{head} can be 0.7. In most cases, the state space is binary. So, a_x = 0.5. E(x) is the expectation value of x that equals to b_x + (a_x)(ux) [21].

In the subjective logic opinion, if b_x = 1, this means an absolute TRUE, if d_x = 1, this means an absolute FALSE. On the other hand, if u_x = 1, this means an empty opinion, if u_x = 0, this means a traditional probability. When $0 < u_x < 1$, this means an opinion with some uncertainty [21].

The subjective logic opinions can be combined using operators such as recommendation, consensus, and conjunction [22]. Recommendation operator can be used when entity A has no opinion about statement P, and there is an entity B that has an opinion about statement P. If A has an opinion about B, then A can get an indirect opinion about statement P. The conjunction operator can be used to combine two opinions, in which each opinion is about a different statement. Consensus operator can be used when there are multiple entities that have opinions about statement P in order to come up with one opinion about statement P [22].

2.3 Change Points Detection Methods

Change points detection can be defined as the process of finding sudden changes in time-series [23]. One effective change point detection approach is to use the divergence between the probability distributions of data in the past and the corresponding data in the future at an instance of time t. The time t point can be considered as a change point if the divergence between the two distributions is significantly large [23]. The generalized likelihood ratio (GLR) [24] and the cumulative sum [25] approaches basically rely on the concept of divergence between the probability distributions to detect the change point. In these two approaches, the logarithm of the likelihood ratio between two probability distributions is used as a measure of change point detection. Each probability density is estimated independently by density estimation. The problem with this kind of approaches is the reliance on the density estimation [26, 27], which always performs poorly. Additionally, the density estimation accuracy declines with noise.

The density-ratio based change point detection approaches have been proposed in order to avoid using the density estimation [5, 28]. Examples of these methods include; kernel mean matching (KMM) [27], the Kullback-Leibler importance estimation procedure (KLIEP) [29], WKV [30], the unconstrained least-squares importance fitting (uLSIF) and its robust extension called relative uLSIF (RuLSIF) [31, 32]. These direct density-ratio estimation methods have offered the optimal convergence rate for nonparametric density-ratio estimation. However, the accuracy of the density-ratio estimation is likely to be declined by noise features.

Stationary subspace analysis (SSA) is a dimensionality reduction method for multivariate timeseries data [33]. SSA factorizes a multivariate time-series data into stationary and nonstationary sources. The change point can be detected in a non-stationary subspace. SSA can reduce the dimensionality of data without losing the sudden change characteristic. Therefore, it can significantly improve change point detection performance. On the other hand, SSA needs to compute the log of a covariance matrix. It also needs a large number of training samples to accurately factorize the stationary and non-stationary sources. Thus, an SSA based change point detection algorithm is not applicable to high-dimensional change point detection problems.

The additive Hilbert-Schmidt Independence Criterion (aHSIC) method can be defined as the weighted sum of HSIC values between each feature and its corresponding pseudo binary labels [23]. One advantage of this method is that it can be estimated using features that are important for a sudden change, which means that it is more robust to noisy features. One problem with this method is the complexity of computing the estimators for HSIC [34].

As mentioned in the introduction, common change detection methods are not suitable for our purposes since we define source reliability as the source consistency with other sources. Therefore, we need to detect a change in the source behavior with respect to other sources.

3. Proposed Approach

Our reliability assessment and monitoring approach detects the change in reliability based on the change in source consistency with respect to other sources instead of detecting the change in the reading values of each source. The approach is composed of three steps: 1) Generating subjective logic opinions. 2) Comparing opinions. 3) Making decisions about reliability change points.

3.1 Generating Subjective Logic Opinions

The subjective logic opinion consists of three parameters; b, d, and u (ω =<b, d, u>). This means that we have to generate b, d, and u for each source's opinion at every instance of time t in the time series. Given that b + d + u = 1, then the degree of freedom is 2. Since the degree of freedom is 2, we need to generate two parameters (d, u), and then calculate the third parameter (b) as; b = 1 - d – u.

3.1.1 Calculating disbelief (d)

The disbelief (d) in our approach is related to how far the source reading is from other sources' readings. For example, if we have three sources S1, S2, and S3 and their readings at time t are; 10, 12, and 20 respectively (As shown in Table 1), then S1 reading is 12 units far from other sources (2 unit far from S2 and 10 units far from S3), S2 reading is 10 units far from other sources (2 unit far from S1 and 8 units far from S3), and S3 reading is 18 units far (10 units far from S1 and 8 units far from S2). And the differences matrix at time t= [12 10 18].

Table 1. Three sources readings during three time units

Time	Source 1 Value	Source 2 Value	Source 3 Value
t-1	10	10	17
t	10	12	20
t+1	11	11.5	20

One challenge here is that the differences between the sources' readings are heavily related to the type of the measured variable. For example, if two sources measuring human body temperature have 2 units difference between their readings, and another two sources measuring atmosphere temperature have 2 units difference between their readings, then the meaning of value of 2 here is different in the two cases. To address this challenge, we transform the difference in readings to a standardized value by using the equivalent z score value (As shown in equation 1). In other words, the difference value is represented by how many standard deviations this value is far from the mean of the differences. In this example, the 2 units of difference in human body temperature will be transformed to z1 and the 2 units of difference in atmosphere temperature will be transformed to z2. z1 and z2 are different according to the standard deviation in each case.

$$\mathbf{Z= (value - \mu) / \sigma} \qquad (1)$$

Where μ is the mean of the differences' matrix, and σ is the standard deviation of the differences' matrix.

If one source is far from other sources by a threshold value (For example, 6 standard deviations), then this source will get a high disbelief value, which means that we can't trust this source. We use exponential function property here because e^{-x} approximately saturates for x>6. The source disbelief at time t is given by equation 2:

$$\mathbf{Disbelief\ (d_i) = 1 - e^{-k|(diff\_avgi - \mu)/\sigma|}} \qquad (2)$$

Where di is the disbelief of source i. $diff\_avg_i$ is the averaged source i difference from all other sources' readings. μ is the sources averaged differences matrix mean, and σ is the sources averaged differences matrix standard deviation, k is a control factor to specify at which point the uncertainty saturates since this value scales the $| diff\_avg - \mu)/\sigma |$ and then the power of the exponential. For example, to calculate the disbelief for source 1 at time t using the data from Table 1:

$$\text{Differences\_Matrix} = \begin{bmatrix} 0 & 2 & 10 \\ 2 & 0 & 8 \\ 10 & 8 & 0 \end{bmatrix}$$

As it is a symmetric matrix, we can sum either; row 1 or column 1 values in order to get the total difference value for source 1. To find the source 1 differences average, we divide the total differences of source 1 by the (number of sources -1). Thus, source 1 differences average = (0 + 2 + 10)/ (3-1) = 6.

Sources averaged differences matrix= [6 5 9], μ = 6.667, and σ = 2.0817

By applying equation 2 to source 1 using k=0.5, then

$d1 = 1 – e\text{-}0.5^{|(6 – 6.667)/2.0817|} = 0.1480$

3.1.2 Calculating Uncertainty (u)

The uncertainty (u) in our approach is related to how far each source reading is from the mean of all sources' readings after excluding the extreme readings. For example, if we have four sources S1, S2, S3, and S4 and their readings at time t are; 10, 11, 13 and 30 respectively. In this case, S4 reading is an extreme value. If S4 is included in the mean calculation, then the mean will be 16. However, if we exclude the S4 reading, then the mean will be 11.33. Having extreme readings will lead to misleading results. As in the previous example, having a mean value of 16 means that sources 1, 2, and 3 are far from the mean, and thus their disbelief values will increase. On the other hand, having a mean value of 11.33 means that sources 1, 2, and 3 are close to the mean, and thus their disbelief values will be lower than the disbelief values when using extreme case. In both cases, the extreme value will be far from the mean and will get the highest disbelief value. In the next step, we calculate how many standard deviations each source reading is far from the mean.

The uncertainty of the reading is related to the distance between this reading and the mean. Thus, when the reading is far from the mean it will gain more uncertainty. If the source is far from the mean of the readings by a threshold value or more such as 6 standard deviations, then the source will gain a higher uncertainty score. We use exponential function property here because e^{-x} approximately saturates for x>6. The source uncertainty value at a time t is given by equation 3:

$$\text{Uncertainty } (u_i) = 1 – e^{-k|(xi – \mu)/\sigma|} \quad (3)$$

Where u_i is the uncertainty of source i, x_i is the reading of source i, μ is the readings mean after excluding the extreme readings, and σ is the readings standard deviation after excluding the extreme readings, k is a control factor to specify at which point the uncertainty saturates since this value scales the | xi – μ)/σ | and then the power of the exponential. For example, to calculate the uncertainty for source 1 at time t using the data from Table 1:

Readings _Average = (10 + 12 + 20) / 3 = 14

By using k=0.5, the Source 1 uncertainty (u_1) = $1 – e^{-0.5|(10 – 14)/5.92|} = 0.343$

3.1.3 Calculating belief (b)

As b + d + u = 1 then

$b_i = 1 - d_i – u_i$ (4)

The value of disbelief (d_i) ranges from 0 to 1. Additionally, the value of uncertainty (u_i) ranges from 0 to 1. As a result, the value of ($d_i + u_i$) can be greater than 1. Therefore, it is important to check the sum of disbelief and uncertainty. So, if the value of ($d_i + u_i$) > 1, then the sum needs to be scaled down to ($d_i + u_i = 1$). Equation 5 will be used to scale down the value of disbelief and uncertainty.

$new\_ u_i = u_i / (d_i + u_i)$ and $new\_ d_i = d_i / (d_i + u_i)$ (5)

In order to find the belief value for source 1 at time t using the data from Table 1:

Check if ($d_1 + u_1$) > 1, ($d_1 + u_1$) = (0.1480 + 0.3430) = 0.491, then no need to scale d_1 and u_1.

$b_1 = 1 - d_1 - u_1 = 1 – 0.1480 – 0. 3430= 0.509$

When the source reading is far from other sources' readings, then it will have a high uncertainty value and a high disbelief value. Thus, it will have a low belief value. On the other hand, when the source reading is close to the other sources' readings, then it will have a low disbelief value and a low uncertainty value, which will result in a high belief value.

3.1.4 Special cases in the proposed approach

To illustrate our approach, we consider some special cases of reliability opinions. An opinion can be represented as ω= <b, d, u> b, d, u and ϵ [0 1], b + d + u =1. If either one of b, d, or u equals one, then there will be a special case. If there are n sources, then all sources readings matrix can be defined as S= [s_1, s_2… s_n]. The followings are the special cases:

- **Case 1: ω= < 1, 0, 0>.**
  When all sources provide the same value such as S = [10 10 10], then the opinion for each source will be the same and equals to: ω= <1, 0, 0>.

- **Case 2: ω= < 0, 1, 0>.**

When the source reading is far from all other sources' readings by at least a threshold value, then the opinion for this source is: ω= < 0, 1, 0>. For example, If S= [10 15 100 22] and the threshold value is 50, then source 3 will have an opinion of ω= < 0, 1, 0>.

- **Case 3: ω= < 0, 0, 1>.**
  When the source reading is an extreme value and far from other sources readings mean by at least a threshold value, then the opinion of this source will be ω= < 0, 0, 1>. For example, If S= [10 12 700] and the threshold value is 100, then source 3 will have an opinion of ω= < 0, 0, 1>.

3.2 Comparing Subjective Logic Opinions

In order to compare the subjective logic opinions that consist of <b, d, u>, it is important to combine each opinion's parameters, and get one value to represent this opinion. In our approach, we are interested in the change of opinions not the expectation value of the opinions. The combination of b, d, and u for each opinion is done using equation 6:

$$\text{Combination\_b\_d\_u} = \left| k*\log(b) / (\log(u) - \log(d)) \right| \tag{6}$$

Where b, d, and u are the opinion belief, disbelief, and uncertainty. k is the scaling factor.

The logarithm is used in equation 6 since b, d, and u ϵ [0 1]. Thus, the logarithm of b, d, and u ϵ [-∞ 0]. The absolute of the logarithm of b, d, and u will be greater than the values themselves, which will result in a more significant difference for comparing. For example, if b changes from 0.6 to 0.8, then its logarithm will change from -0.5108 to -0.2231; the change in b value is 0.2 but the change in its logarithm is 0.29, which is more helpful in detecting the change since the difference is higher.

Assume that opinion ω1 = <0.6, 0.25, 0.15>, ω2 = <0.68 0.20, 0.12>, and k=1, then the combination of ω1 (using equation 6) will be (ω1_Comb_b_d_u = 1) and the combination of ω2 will be (ω2_Comb_b_d_u = 3.0813). The difference between the two combinations is 2.0813. In this example, in order to compare the two opinions, then the combined values (1, and 3.0813) will be used in the comparison.

3.3 Making decision about the reliability change points

After converting each source opinion to one value, then each source will have its combined opinions for each instance of time. For each source, we detect the change point of reliability using the two sliding windows method [35]. The two windows are; reference window and current window. These two windows have the same width (w). The reference window extends from time t to time t + w, while the current window extends from time t + w + 1 to time t + 2w + 1. For each window, we calculate the mean and standard deviation for each window and then compare the two means and the two standard deviations. If there is any significant change in the means or in the standard deviations, then we regard that there is a reliability change point at time t + w + 1. After that, the reference window will be updated. So it will extend from time t + 2w +2 to time t + 3w + 2. Additionally, the current window will extend from time t + 3w +3 to time t + 4w + 3. If there is no significant change in the means and in the standard deviations of the reference and current windows, then each window will be shifted by one time unit, and the comparison will be done again. The comparison between the two windows will be done as the following:

| log (Ref_Mean / Curr_Mean) | > Mean_Threshold, OR

| log (Ref_Std / Curr_Std) | > Std_Threshold

We use log (a/b) here because we are interested in either (a/b > threshold value) or (b/a > threshold value). Using the absolute value of log (a/b) will give the same result for both a/b and b/a.

Since we are comparing the change in reliability for multiple sources, we may get some clustered reliability change points. This is because each source may provide a reliability change point that is close to another source reliability change point. In the case of clustered change points, we filter these points using a scanning window that has the same width of the reference and current windows.

In the filtering process, if there is more than one reliability change point within the scanning window, then a new change point will be created based on the weighted average of the clustered change points. The new change point will be located close to the strongest reliability change point. The location of this new point is identified by taking the weighted average time of the clustered change points' times. For example, if source 1 combined reliability value has changed at time t from the value 20 to 12, and source 2 combined reliability value has changed from the value 20 to 8 at time t + 3, and the scanning window width is 10-time units. In this case, there will be two reliability change points; one at time t, and one at time t+3. Therefore, after doing the filtering process there will be just one reliability change point that is closer to source 2 time (t + 3). This is because the change in the combined reliability value of source 2 is stronger than the change in the combined reliability value of source 1.

Table 2. Reliability monitoring algorithm

Reliability monitoring algorithm

Input: readings from multiple sources (S1, S2, . . ., Sn) for each time (t1, t2, …, tm)
Output: Reliabilities R*={ωij, i=1:m,j=1:n}
for i=1 to n
initialize reference average μref and σref for the period t=1 to t=window width
for j=1+w to m-w
get the current average μcur and σcur for the period t=j to t=j+window width
compare the μcur and μref averages as well as the σcur and σref
if the comparing (μcur and μref)> μthreshold or comparing (σcur and σref)> σthreshold
add new change point(time j, source n, comparing result value) to the change points matrix
else
slide the reference and current windows one slot.
end if
update the reference and current window.
end for
end for
filter the change point matrix

4. Experimental study

In this section we perform a simulation-based study of our approach. We applied our approach to small- and large-scale scenarios. MATLAB R2014b software was used in this experiment. The experiment was performed using a machine with the following characteristics: AMD Phnom™ II N850 Triple-core Processor 2.20 GHz processor and 4 GB RAM.

4.1 Small scale experiment setup

In order to detect the change in reliability we use simulated data, in which the changes in the sources' readings are known. Figure 3 shows a data set of three sources (S1, S2, and S3) for a time series (t1, t2. . . t100). In this data set, the three sources are consistent during the period (t1...t25). At time t26, source 3 deviates from the other two sources. At time t50, source S2 also starts to deviate, and at time t76, source 2 and source3 start to deviate in more severe patterns than the previous period of deviations (t26...t50, t51...t75). The time for this dataset is divided into four intervals; each interval has different values of mean and standard deviation for each source as shown in Table 3.

Table 3. Three sources with different distributions

Region	Region Interval	Sources Mean	Source Standard Deviations
R1	t1...t25	$\mu1 = \mu2 = \mu1 = 11$	$\sigma1 = \sigma2 = \sigma3 = 0.2$
R2	t26...t50	$\mu1 = \mu2 = \mu3 = 11$	$\sigma1 = \sigma2 = 0.2$ $\sigma3 = 2$
R3	t51...t75	$\mu1 = \mu2 = \mu3 = 11$	$\sigma1 = 0.2$ $\sigma2 = 2$ $\sigma3 = 4$
R4	t76...t100	$\mu1 = 11$ $\mu2 = 15$ $\mu3 = 20$	$\sigma1 = 1$ $\sigma2 = 3$ $\sigma3 = 6$

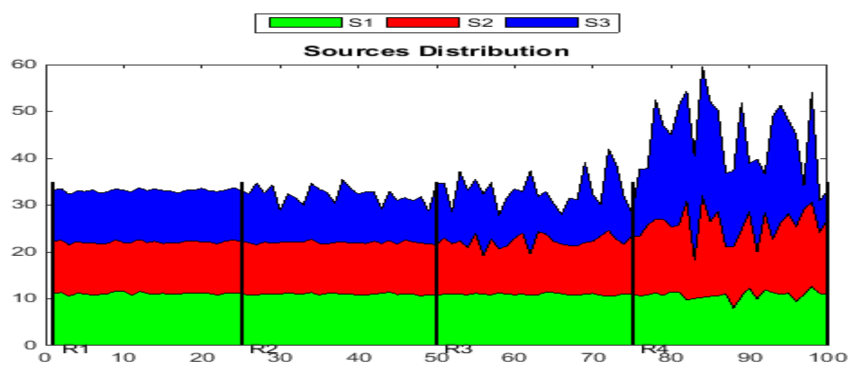
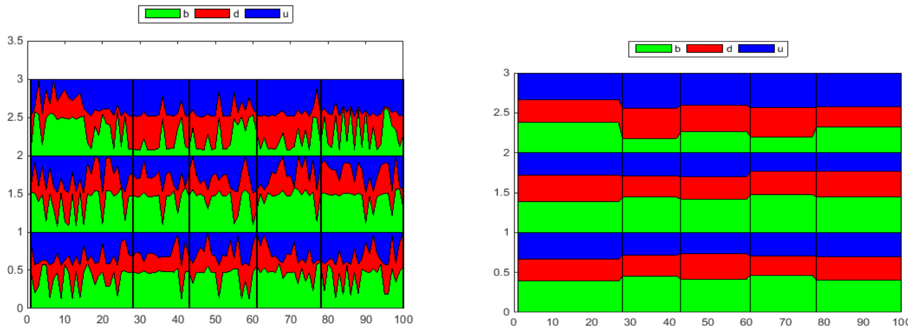


Fig. 3. Area plot for the three sources readings

Figure 3 is an area plot for the three sources (as shown in Table 3) after generating 100 records for each source using its μ and σ. Source 1 distribution is shown in green, source 2 distribution is shown in red, and source 3 distribution is shown in blue. The vertical lines demarcate the changing behavioral patterns of the sources. At t25, source 3 starts to deviate from other sources. At t50, source 2 starts to deviate and source 3 continues to deviate in a more severe pattern. At t75, source 1 starts to deviate, while sources 2 and 3 start to deviate in more severe patterns than the previous deviation.

4.2 Small Scale Experiment Results

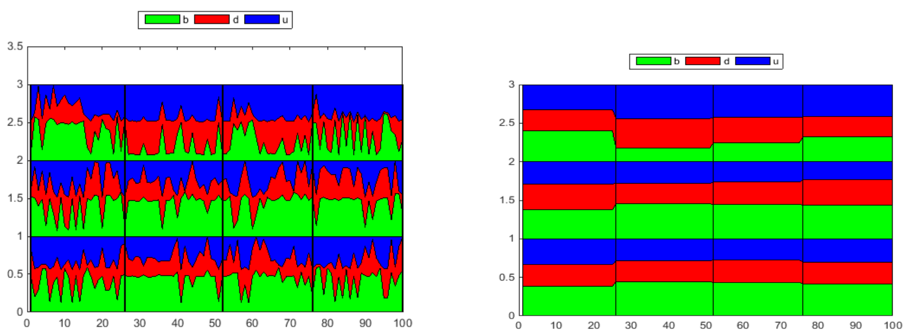
After applying the proposed approach 200 times on a predefined dataset, we attained consistent results with approximately similar outputs in the 200 times. The following Figures illustrate the result we attained from one case (the identical case in Table 3).



A **B**

Fig. 4. A) Area plot for subjective logic opinions for three sources.

B) Area plot for averaged values. Window width=11.



A **B**

Fig. 5. A) Area plot for subjective logic opinions for three sources.

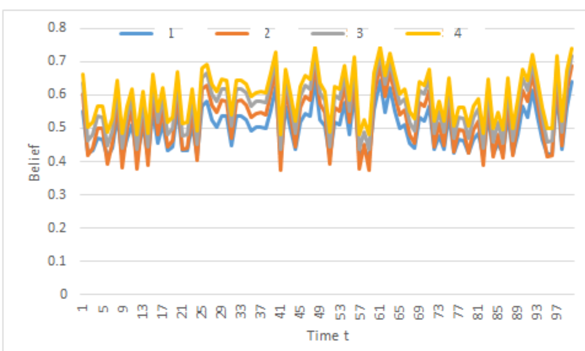
B) Area plot for averaged values. Window width=14.

Figures 4 and 5 show the reliability of three sources using window width of 11 and 14 time units respectively. The source's reliability is represented by subjective logic opinions, in which each opinion is related to a different region. In Figure 5, each source has a reliability opinion for each region: region1 [t1:t26), region 2 [t26:t52), region 3 [t52:t76), and region 4 [t76:t100). The belief is represented by green, disbelief is represented by red, and uncertainty is represented by blue.

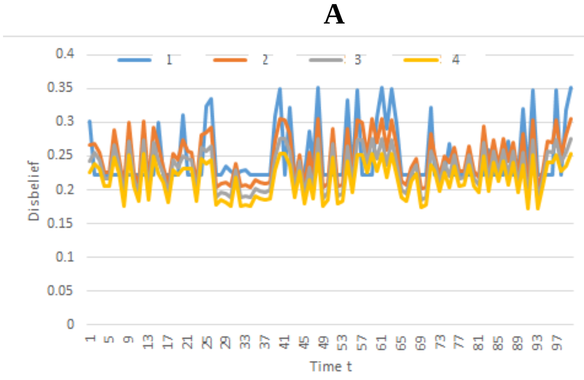
As b + d + u = 1 for each opinion, then source 1 opinion area extends vertically from 0 to 1, while source 2 opinion area extends from 1 to 2 and source 3 opinion area extends from 2 to 3. The vertical lines represent the change points in reliability. Between each two consecutive lines there is a different region of reliability. Figures 4(b) and 5 (b) show the averaged belief, disbelief, and uncertainty for each source within each region.

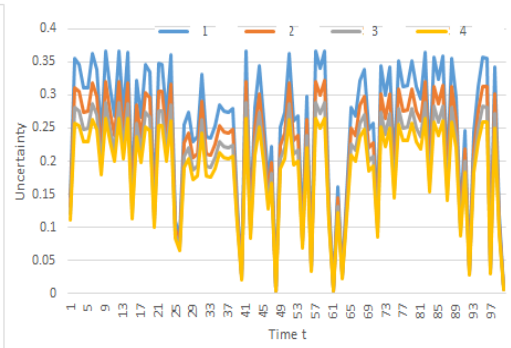
Our proposed approach can accurately detect the reliability change points. Based on the input data set shown in Table 3, there should be change points at t25, t50, and t75. If we look at the output result shown in Figure 5 the change points at these times can clearly be seen.

At time (t26), source 3 got less reliability level because of its deviation from the other sources. At time t52, source 2 started to deviate. As a result, source 3 got more reliability level since it has become more consistent with source 2. Additionally, source 1 and source 2 got lower reliability levels since they are no longer as consistent as before. In general, if the source is more consistent with other sources it will have a higher reliability value (expressed as a subjective logic opinion) as shown in Figure 6. Deviation doesn't necessarily indicate low reliability because if this deviation makes the source consistent with other sources, it will give the source more reliability.



A





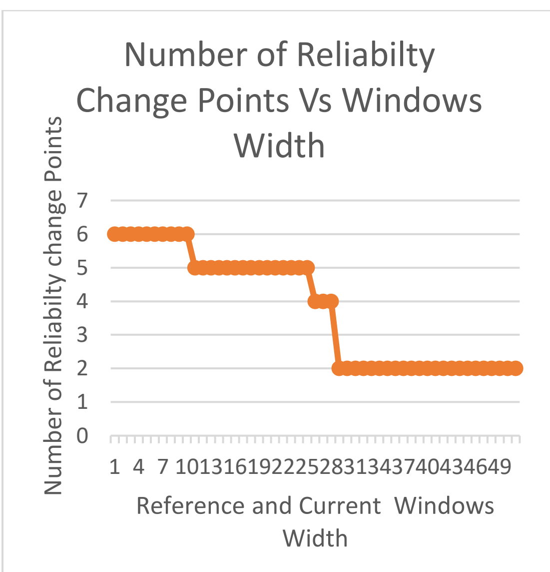
B **C**

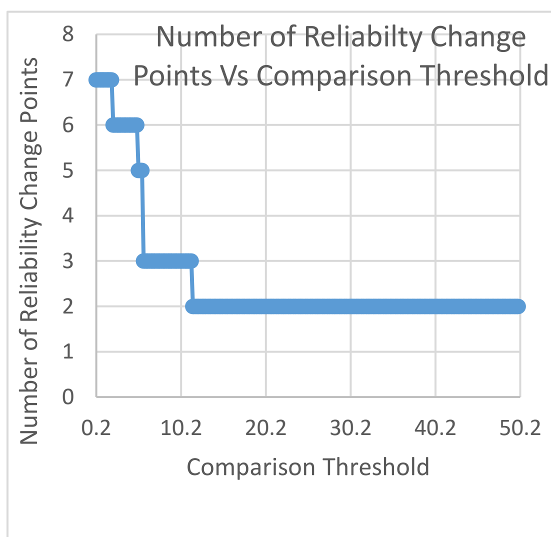
Fig. 6. A) Source 1 belief vs. different numbers of consistent sources with source 1
B) Source 1 disbelief vs. different numbers of consistent sources with source 1
C) Source 1 uncertainty vs. different numbers of consistent sources with source 1

Figure 6 shows the belief, disbelief, and uncertainty for source 1 in different cases. In case one, there is one source consistent with source 1. In case two, there are two sources consistent with source 1. In case three, there are three sources consistent with source 1, and in case four, there are four sources consistent with source 1.

From Figure 6.A we can see that the source belief value increases when the number of consistent sources with this source increases. However, Figure 6.B and Figure 6.C show that the disbelief and uncertainty values decrease when the number of consistent sources with this source increases. The decrease in the uncertainty value is larger than the decrease in the disbelief value for any source as shown in Figure 6.B and Figure 6.C. This is because | (Source1_Reading – μ)/σ | is smaller than | (Source1_AvgDiff – μ)/σ |.

Figure 7.A shows the relationship between the window width and the number of reliability change points. Figure 7.B shows the relationship between the comparison threshold and the number of reliability change points. Using wider windows will result in smaller number of regions. However, the sensitivity to detect the change will be low. Additionally, there is another factor that affects the number of reliability change points, which is the comparison threshold that is used in comparing the means and the standard deviations between the reference and current windows. Higher threshold will ignore small changes, and this will lead to a smaller number of reliability change points as shown in Figure 7.B.





A **B**

Fig. 7. A) Number of Reliability Change Points vs. windows width.

B) Number of Reliability Change Points vs. comparison threshold.

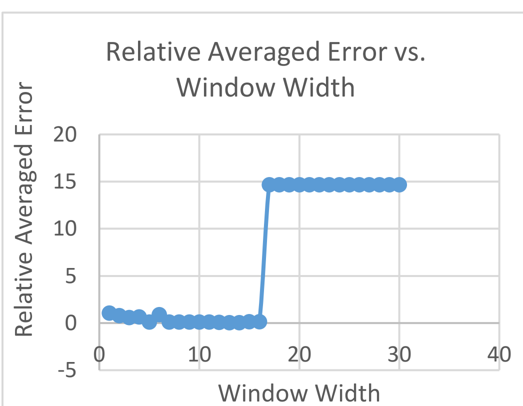
As shown in Figure 7.A, when the window width is 7, then the number of reliability change points is 6. However, when the window of width is wider such as 14, then the number of reliability change points is 5. Using lower value for window width makes the detection of change points more sensitive to any change, while using higher values for window width makes the detection less sensitive to change points.

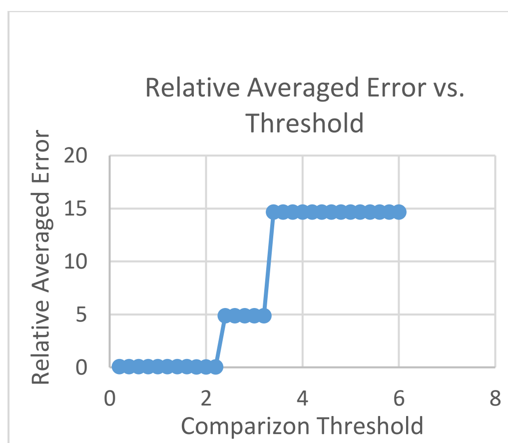
Using lower threshold value in the comparison of two distributions will result in more reliability change points. For example, 0.2 threshold will result in 7 reliability change points, while larger threshold value such as 8 will result in 3 reliability change points. Using lower value for threshold makes the detection of change points more sensitive to any change. However, using higher values for threshold makes the detection less sensitive to change points.

Figure 8.A shows the relationship between the relative averaged error and the window width. Figure 8.B shows the relationship between the relative averaged error and the comparison threshold. The relative averaged error can be calculated using equation 7.

$$Relative\_Averaged\_Error = 1/n \sum_{k=1}^{n} \left|(Detected\_Point_k - Ref\_Point_k) / Ref\_Point_k\right| \qquad (7)$$

Where n is the Max (number of detected change points, number of reference points). Each detected point is mapped to the nearest reference point. For example, if a reference set of points R = {20, 36, 55}, then the detected point 33 is mapped to the point of value 36 from R.





A **B**

Fig. 8. A) Relative averaged error vs. windows width.

B) Relative averaged error vs. comparison threshold.

As shown in figure 8.A, using high window width and high comparison threshold can result in a large error of reliability change detection. For example, using window width of 30 and comparison threshold of 5 will result in a large error in detection. The error here doesn't mean that we have incorrect reliability change points. Instead, it means that we have incorrect number of the reliability change points. That is either, lower number of change points than expected or larger number of change points than expected. For example, if there is an actual set of points R = {20, 36, 55, 76}, and the detected set of points D = {51, 77}. In this example, the two detected points are correct. However, the number of the detected points is incorrect.

Table 4 shows three cases of reliability change point detection. In case 1, the number of detected change points is less than the number of actual points. In case 2, the number of detected change points equals the number of actual points. In case 3, the number of detected points is larger than the number of reference points. In case 1 and case 3, the relative averaged error is high since there is a difference in the number between the detected change points and the actual points, which results in an error in the mapping between the detected and the actual points. Each point is mapped to the nearest point in the other set. For example (As shown in Table 4), X (20) in case 1 is mapped to Y (51), and X (18) in case 3 is mapped to Y(55). On the other hand, if the number of detected change points equals the number of actual points, then the relative averaged error will be low. For example, X (20) in case 2 is mapped to Y (20).

Table 4. Three cases of reliability change detection

Case 1
R = {20, 36, 55,76}
D ={51,77}
X = Larger set of R,D
Y =Smaller set of R,D

X	Y	$\lvert(Y-X)/X\rvert$
20	51	31/20
36	51	15/36
55	51	4/55
76	77	1/76
Total		2.053
Total / n		0.5133

Case 2
R = {20, 36, 55,76}
D ={20, 35, 51,77}
X = Larger set of R,D
Y =Smaller set of R,D

X	Y	$\lvert(Y-X)/X\rvert$
20	20	0/20
36	35	1/36
55	51	4/55
76	77	1/76
Total		0.1137
Total / n		0.0284

Case 3
R = {55,76}
D ={18, 30, 51,77}
X = Larger set of R,D
Y =Smaller set of R,D

X	Y	$\lvert(Y-X)/X\rvert$
18	55	37/18
30	55	25/30
51	55	4/51
77	76	1/77
Total		2.98
Total / n		0.745

To come up with a one final result from multiple sources, we use the consensus operator in order to combine the multiple opinions about the sources readings average into one opinion

[22]. As shown in Figure 9.B, the average of readings from multiple sources within each region is stated in each region.

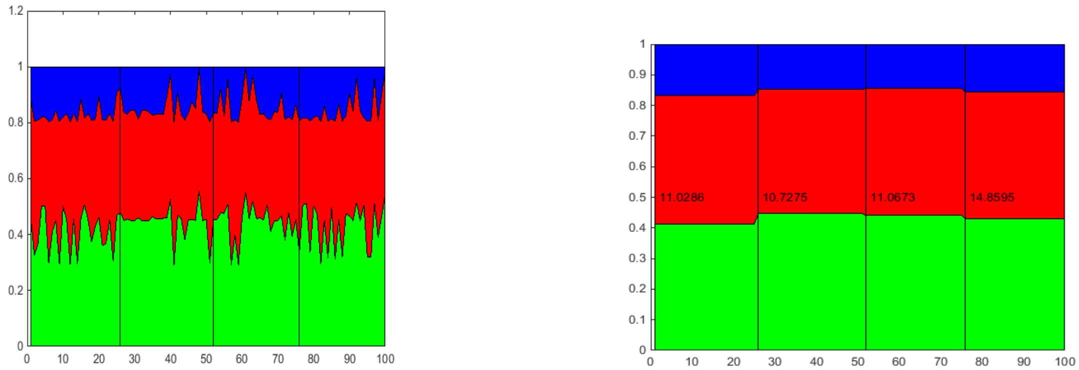


Fig. 9. A) Subjective logic opinion about the sources' readings average for each region. B) Averaged subjective logic opinion about the sources' readings average for each region.

4.3. Large scale experiment set up

In this experiment, there are 25 data sources. Each data source has different mean and standard deviation, which are assigned randomly using MATLAB. We just present this large-scale experiment to show that our approach can be applied to any number of sources. All discussion from the small-scale experiment will be applied here.

We randomly generated 25 means and 25 standard deviations. After that, we assigned a mean and a standard deviation to every source, and then we generated 100 reading values for each source using its μ and σ. After applying the approach, we got the following results. Figure 10 shows the subjective logic opinions for the 25 sources

4.4 Large scale experiment results

Figure 11 shows the area plot for the subjective logic opinions for the 25 sources. Figure 12 shows a zoom in view of source 9. We can see that source 9 has a low belief value and high disbelief and uncertainty values. This low belief value means that source 9 readings are far from other sources. As shown in Figure 11, source 13 and source 15 have high belief values since their readings are consistent with the other sources' readings.

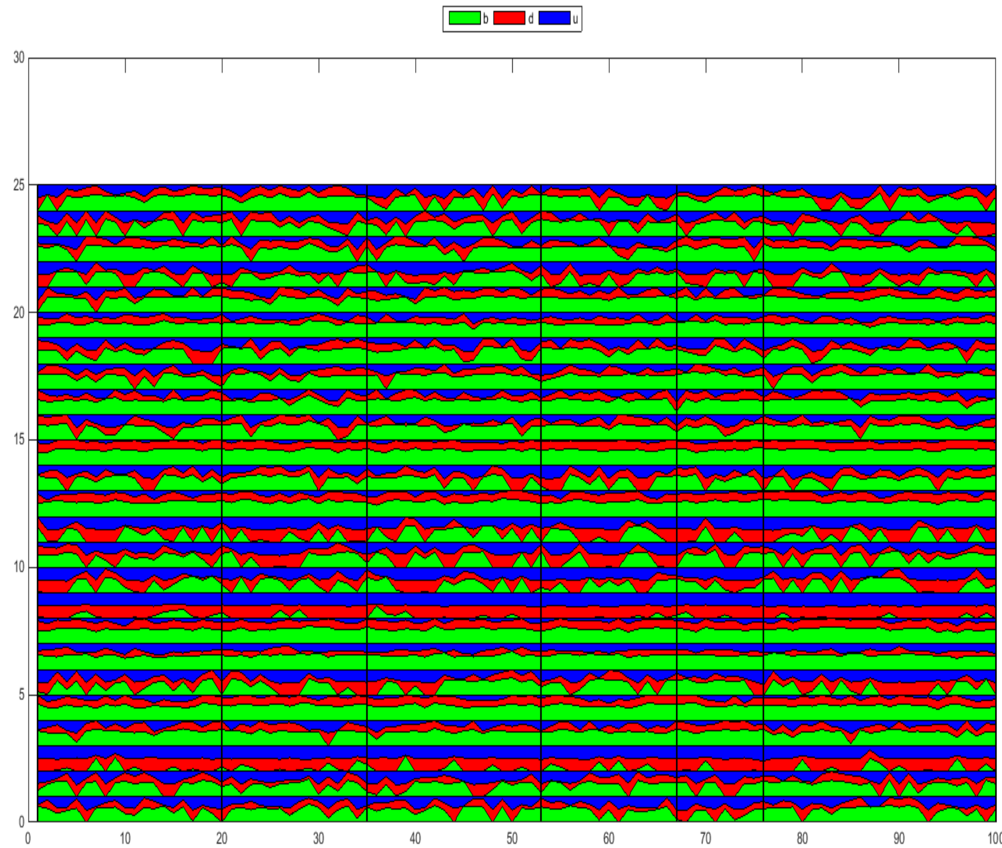


Fig. 10. Area plot for subjective logic opinions for 25 sources.

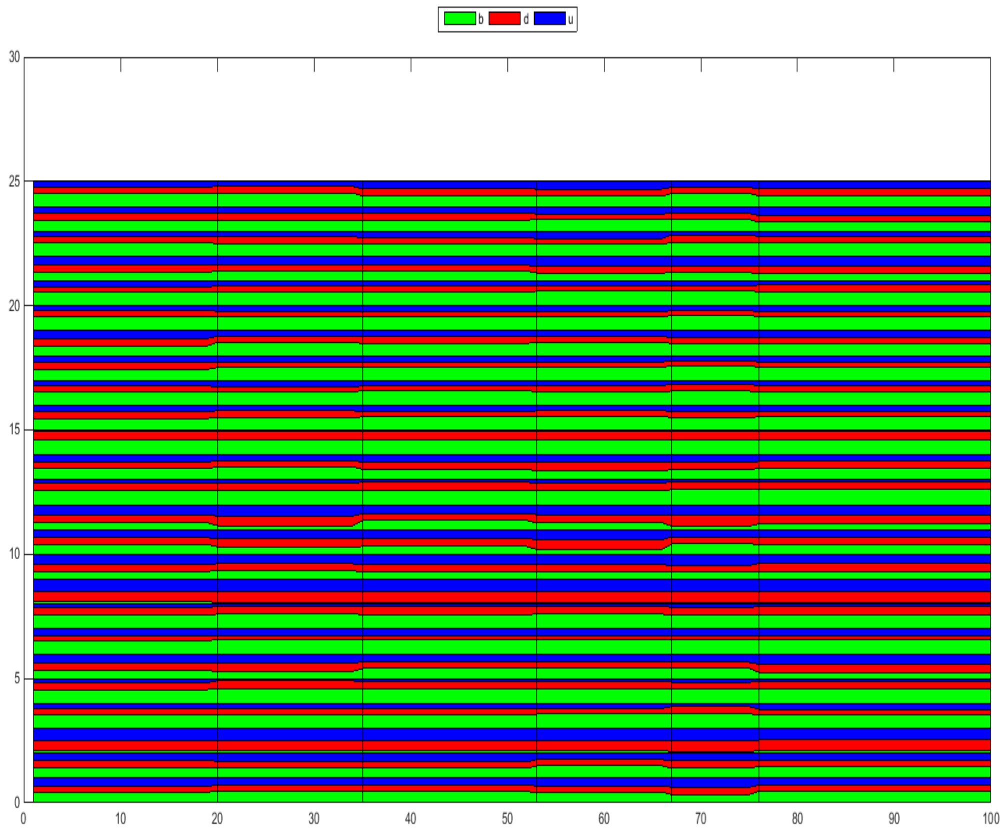


Fig. 11. Area plot for averaged subjective logic opinions for 25 sources.

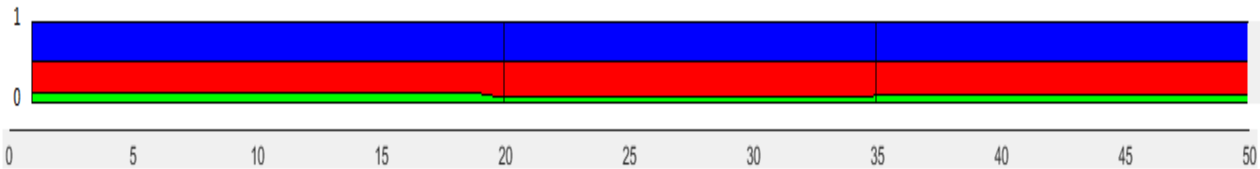


Fig. 12. Area plot for averaged subjective logic opinions for source 9.

Table 5 shows 30 readings from sources 9, 13, and 15 and their belief values. It is clear that source 9 readings are far from the other two sources. Source 9 values are in the range (18:23), while source 13 and source 15 values are in the range (12:17).

Table 5: Source 9, 13, and 15 readings, and their corresponding belief values

Time t		S9	S13	S15		S9 _b	S13 _b	S15 _b
1		19.77507	13.65453	13.39954		0.08913	0.501393	0.514336
2		20.72783	15.59529	13.4814		1.00E-10	0.587713	0.513649
3		23.31965	14.75939	12.99166		1.00E-10	0.606329	0.67203

4		22.56084	13.49526	12.34677		1.00E-10	0.588026	0.655517
5		19.96574	14.48278	13.27402		0.201681	0.542224	0.621861
6		18.69248	13.46591	13.57264		0.283943	0.58755	0.578307
7		20.9463	12.55796	13.24514		0.072907	0.532551	0.589044
8		22.03072	13.71037	13.34898		1.00E-10	0.551947	0.565168
9		23.03662	13.23019	13.28271		1.00E-10	0.656599	0.651451
10		21.86463	12.63286	12.94307		1.00E-10	0.577778	0.604574
11		20.18528	15.81529	13.19923		0.040758	0.515249	0.59391
12		22.158	13.50594	13.39566		1.00E-10	0.53173	0.537045
13		20.11619	17.24487	13.33196		0.098729	0.499834	0.638194
14		20.40809	12.73488	13.7037		0.254077	0.570584	0.659971
15		20.83238	14.01709	13.50783		0.303186	0.634138	0.653232

Our approach can capture the inconsistency between different sources and reflect the degree of this inconstancy. It also can determine the period of this inconstancy. For example, As shown in Figure 11, sources 13, and 15 have high belief values in the period [t1:t100]. This means that these two sources are consistent with the other sources. Source 12 has a low belief value in the period [t21:t35]. In the other hand, it has a high belief value in the period [t36:t50]. If we go to Table 6, we can find that source 12 is far from sources 13 and 15 within the period [t21:t35]. On the other hand, source 12 is relatively close to sources 13 and 15 within the period [t36:t50]. If we calculate the total difference between source 12 and sources 13 and 15 in the period [t21:t35], we find that it equals to 234.3668, while the total difference between source 12 and sources 13 and 15 in the period [t36:t50] is 143.6051. Therefore, the change points and the sources subjective logic opinions shown in Figure 11 are compatible with the data in Tables 5 and 6. This ensures that our approach can accurately work on any number of sources.

Table 6. Source 12, 13, and 15 readings, and their corresponding belief values

Time t		S12	S13	S15		S12 _b	S13 _b	S15 _b
21		20.86666	11.20937	13.65205		1.00E-10	0.542678	0.622147
22		17.84667	12.76814	13.15242		4.91E-01	0.656834	0.651143
23		21.97382	13.58133	13.00024		1.00E-10	0.647482	0.608106
24		20.14622	15.39586	13.55171		7.85E-02	0.559692	0.677297
25		19.9442	15.31776	12.673		1.00E-10	0.491627	0.607042
26		30.87008	13.13514	13.8152		1.00E-10	0.637721	0.620209
27		17.80486	14.553	13.30997		0.4105	0.498908	0.525411
28		19.93364	12.69746	13.7723		0.137865	0.632965	0.614879
29		29.89884	14.16138	13.29002		1.00E-10	0.703787	0.718038
30		15.91891	16.03028	13.70429		4.95E-01	0.491804	0.654287

31		22.12343	13.40981	13.3874		1.00E-10	0.668511	0.669149
32		16.17529	12.72837	13.45757		5.16E-01	0.563139	0.5093
33		23.867	12.43109	13.64552		1.00E-10	0.617303	0.648352
34		21.30038	14.24386	14.01773		5.25E-02	0.573584	0.585332
35		22.48527	15.16896	12.90519		1.00E-10	0.557447	0.663148
36		19.08682	16.34378	12.2981		0.588373	0.438104	0.54535
37		20.99981	13.25675	12.20369		0.19781	0.595437	0.508974
38		25.48001	13.41916	13.35244		1.00E-10	0.659607	0.663304
39		11.79978	12.23339	13.16369		0.614573	0.622672	0.575008
40		14.14264	15.57831	13.34125		6.18E-01	0.581759	0.69038
41		21.62139	13.06378	13.1121		1.00E-10	0.588619	0.585992
42		17.71461	12.47102	13.5329		4.76E-01	0.689702	0.641314
43		19.02299	16.43104	13.95623		0.489039	0.482677	0.62688
44		14.64045	14.38305	13.15965		5.59E-01	0.568285	0.635554
45		16.80178	14.24367	13.6625		5.93E-01	0.512934	0.563278
46		18.62935	11.67777	13.94902		1.68E-01	0.538587	0.578993
47		17.45533	13.82291	12.13984		0.619123	0.640453	0.589856
48		16.94177	13.9513	13.24294		6.08E-01	0.582229	0.601552
49		18.56487	11.54861	12.97337		1.07E-01	0.705197	0.595602
50		16.66368	12.83975	12.63978		0.545392	0.679265	0.667904

5. Conclusion and future work

We presented a reliability assessment and monitoring approach for multiple data sources using subjective logic. We define source reliability as the source consistency with other sources. Each source's reliability is a vector of subjective logic opinion rather than just one weighted value for each source. The proposed approach can describe the behavior of the source reliability, which means that one can see where the source gains more reliability and where it loses some reliability.

We conducted a small and a large-scale simulation-based study of our approach. We demonstrated that, our approach accurately detects the reliability change points, regardless of the number of sources. The accuracy of detection depends on the size of sliding window and the comparison threshold. In future work, we will explore optimal size of sliding window. Additionally, we will investigate methods for predicting source reliability, as well as, methods for data conflict resolution based on the source reliability.

References

1. Nittel, S., Real-time Sensor Data Streams. SIGSPATIAL Newletter, Spec, 2015. 7(2): p. 22-28.

2. Minaie, A. and A. Sanati-Mehrizy, Application of Wireless Sensor Networks in Health Care System. 120th ASEE Annual Conference and Exposition, 2013.

3. Elnahrawy, E., Research Directions in Sensor Data Streams: Solutions and Challenges. Rutgers University, Tech. Rep. DCIS-TR-527, 2003.

4. Qi, L., et al., Resolving Conflicts in Heterogeneous Data by Truth Discovery and Source Reliability Estimation. SIGMOD, 2014.

5. Kawahara, Y. and S. M., Change-point detection in time-series data by direct density-ratio estimation. SDM, 2009: p. 389–400.

6. Liu, S., et al., Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 2013. 43: p. 72-83.

7. Lombardo, J. Assessing the Reliability and Validity of Sources. [cited 2015; Available from: http://study.com/academy/lesson/assessing-the-reliability-and-validity-of-sources.html.

8. Vijayaganesh, M., et al., An Evaluating Enhanced Data Reliability to a Web-Enabled Data Warehouse. International Journal of Innovative Research in Computer and Communication Engineering, 2015. 3(2).

9. Destercke, S., P. Buche, and B. Charnomordic, Evaluating data reliability: an evidential answer with application to a Web-enabled data warehouse. IEEE Computer Society, 2013. 25(1): p. 92105. 10. Nimier, V. A formalisation of operational recommendations. in Proceedings of the Seventh International Conference on Information Fusion. 2004. Mountain View, CA.

11. Marchetti, L., D. Nobili, and L. Iocchi, Improving Tracking by Integrating Reliability of Multiple Sources, in Information Fusion: 11th International Conference on. 2008, IEEE. p. 1-8.

12. Ceruti, M., et al., Pedigree information for enhanced situation and threat assessment, in Information Fusion: 9th International Conference. 2006. p. 1-8.

13. Young, S. and J. Palmer, Pedigree and confidence: Issues in data credibility and reliability, in Information Fusion: 10th International Conference. 2007. p. 1-8.

14. Yan, W., Fusion in multi-criterion feature ranking, in Information Fusion: 10th International Conference. 2007. p. 1-6.

15. Buche, P., D. lemy, and H. Chebil, Flexible SparqlQuerying of Web Data Tables Driven by an Ontology. Flexible Query Answering Systems (FQAS), 2009: p. 345-357.

16. Mercier, D., B. Quost, and T. Denoeux, Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting. Information Fusion,, 2008. 9: p. 246-258.

17. Wang, D., et al., On Quantifying the Accuracy of Maximum Likelihood Estimation of Participant Reliability in Social Sensing. DMSN 2011: 8th International Workshop on Data Management for Sensor Networks, 2011. 31(6).

18. Wang, D., T. Abdelzaher, and L. Kaplan, A maximum likelihood estimation approach. UIUC Technical Report. 2011.

19. Jøsang, A. Probabilistic Logic Under Uncertainty. in Proceedings of Computing: The Australian Theory Symposium (CATS'07). 2007. Ballarat.

20. Druzdzel, M., Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense. 1993.

21. Pope, S. and A. Jøsang, Analysis of competing hypotheses using subjective logic. Proceedings of the 10th International Command and Control Research and Technology Symposium, 2005.

22. Oleshchuk, V. and V. Zadorozhny, Trust-Aware Query Processing in Data Intensive Sensor Networks, in International Conference on Sensor Technologies and Applications. 2007.

23. Yamada, M., et al. Change-Point Detection with Feature Selection in High-Dimensional TimeSeries Data. in Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence 2013.

24. Gustafsson, F., The marginalized likelihood ratio test for detecting abrupt changes. Automatic Control, IEEE Transactions on,, 1996. 41(1): p. 66–78.

25. Basseville, M. and I.V. Nikiforov, Detection of abrupt changes: theory and application. 1993, NJ: Prentice Hall Englewood Cliffs.

26. Hardle, W., et al., Nonparametric and Semiparametric Models. Springer Verlag, 2004.

27. Huang, J., et al., Correcting sample selection bias by unlabeled data. NIPS, 2007: p. 601–608.

28. Liu, S., et al., Change-point detection in time-series data by relative densityratio estimation. Neural Networks, 2013. 43: p. 72–83.

29. Sugiyama, M., et al., Direct importance estimation for covariate shift adaptation. Annals of the Institute of Statistical Mathematics, 2008. 60(4): p. 699–746.

30. Nguyen, X.L., M.J. Wainwright, and M.I. Jordan, Estimating divergence functionals and the likelihood ratio by convex risk minimization. Information Theory, IEEE Transactions on, 2010. 56(11): p. 5847–5861.

31. Kanamori, T., S. Hido, and M. Sugiyama, A least-squares approach to direct importance estimation. Journal of Machine Learning Research, 2009. 10: p. 1391–1445.

32. Yamada, M., T. Suzuki, and T. Kanamori, Relative density-ratio estimation for robust distribution comparison. NIPS, 2011: p. 594–602.

33. Blythe, D., et al., Feature extraction for change-point detection using stationary subspace analysis. Neural Networks and Learning Systems, IEEE Transactions on, 2012. 23(4): p. 631–643.

34. Song, L., et al., Feature Selection via Dependence Maximization. Journal of Machine Learning Research, 2012.

35. Kifer, D., S. Ben-David, and J. Gehrke, Detecting Chnage in Data Streams. VLDB 04, 2004. 30: p. 180-191.

1.
Nittel, S., *Real-time Sensor Data Streams.* SIGSPATIAL Newletter, Spec, 2015. **7**(2): p. 22-28.

2. Minaie, A. and A. Sanati-Mehrizy, *Application of Wireless Sensor Networks in Health Care System.* 120th ASEE Annual Conference and Exposition, 2013.
3. Elnahrawy, E., *Research Directions in Sensor Data Streams: Solutions and Challenges.* Rutgers University,. Rep. DCIS-TR-527, 2003.
4. Kawahara, Y. and S. M., *Change-point detection in time-series data by direct density-ratio estimation.* SDM, 2009: p. 389–400.
5. Liu, S., et al., *Change-point detection in time-series data by relative density-ratio estimation.* Neural Networks, 2013. **43**: p. 72-83.
6. Lombardo, J. *Assessing the Reliability and Validity of Sources*. [cited 2015; Available from: http://study.com/academy/lesson/assessing-the-reliability-and-validity-of-sources.html.
7. Vijayaganesh, M., et al., *An Evaluating Enhanced Data Reliability to a Web-Enabled Data Warehouse.* International Journal of Innovative Research in Computer and Communication Engineering, 2015. **3**(2).
8. Destercke, S., P. Buche, and B. Charnomordic, *Evaluating data reliability: an evidential answer with application to a Web-enabled data warehouse.* IEEE Computer Society, 2013. **25**(1): p. 92-105.
9. Nimier, V. *A formalisation of operational recommendations*. in *Proceedings of the Seventh International Conference on Information Fusion*. 2004. Mountain View, CA.
10. Marchetti, L., D. Nobili, and L. Iocchi, *Improving Tracking by Integrating Reliability of Multiple Sources*, in *Information Fusion: 11th International Conference on*. 2008, IEEE. p. 1-8.
11. Ceruti, M., et al., *Pedigree information for enhanced situation and threat assessment*, in *Information Fusion: 9th International Conference*. 2006. p. 1-8.
12. Young, S. and J. Palmer, *Pedigree and confidence: Issues in data credibility and reliability*, in *Information Fusion: 10th International Conference*. 2007. p. 1-8.
13. Yan, W., *Fusion in multi-criterion feature ranking*, in *Information Fusion: 10th International Conference*. 2007. p. 1-6.
14. Buche, P., D. lemy, and H. Chebil, *Flexible SparqlQuerying of Web Data Tables Driven by an Ontology.* Flexible Query Answering Systems (FQAS), 2009: p. 345-357.
15. Mercier, D., B. Quost, and T. Denoeux, *Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting.* Information Fusion,, 2008. **9**: p. 246-258.
16. Cooke, R., *Experts in Uncertainty*. 1991, Oxford Univ. Press,.
17. Wang, D., et al., *On Quantifying the Accuracy of Maximum Likelihood Estimation of Participant Reliability in Social Sensing.* DMSN 2011: 8th International Workshop on Data Management for Sensor Networks, 2011. **31**(6).
18. Wang, D., T. Abdelzaher, and L. Kaplan, *A maximum likelihood estimation approach. UIUC Technical Report.* 2011.
19. Jøsang, A. *Probabilistic Logic Under Uncertainty*. in *Proceedings of Computing: The Australian Theory Symposium (CATS'07)*. 2007. Ballarat.
20. Druzdzel, M., *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense.* 1993.
21. Pope, S. and A. Jøsang, *Analsysis of competing hypotheses using subjective logic.* Proceedings of the 10th International Command and Control Research and Technology Symposium, 2005.
22. Yamada, M., et al. *Change-Point Detection with Feature Selection in High-Dimensional Time-Series Data*. in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* 2013.
23. Gustafsson, F., *The marginalized likelihood ratio test for detecting abrupt changes*. Automatic Control, IEEE Transactions on,, 1996. **41**(1): p. 66–78.
24. Basseville, M. and I.V. Nikiforov, *Detection of abrupt changes: theory and application*. 1993, NJ: Prentice Hall Englewood Cliffs.
25. Hardle, W., et al., *Nonparametric and Semiparametric Models.* Springer Verlag, 2004.
26. Huang, J., et al., *Correcting sample selection bias by unlabeled data.* NIPS, 2007: p. 601–608.

27. Liu, S., et al., *Change-point detection in time-series data by relative densityratio estimation.* Neural Networks, 2013. **43**: p. 72–83.
28. Sugiyama, M., et al., *Direct importance estimation for covariate shift adaptation.* Annals of the Institute of Statistical Mathematics, 2008. **60**(4): p. 699–746.
29. Nguyen, X.L., M.J. Wainwright, and M.I. Jordan, *Estimating divergence functionals and the likelihood ratio by convex risk minimization.* Information Theory, IEEE Transactions on, 2010. **56**(11): p. 5847–5861.
30. Kanamori, T., S. Hido, and M. Sugiyama, *A least-squares approach to direct importance estimation.* Journal of Machine Learning Research, 2009. **10**: p. 1391–1445.
31. Yamada, M., T. Suzuki, and T. Kanamori, *Relative density-ratio estimation for robust distribution comparison.* NIPS, 2011: p. 594–602.
32. Blythe, D., et al., *Feature extraction for change-point detection using stationary subspace analysis.* Neural Networks and Learning Systems, IEEE Transactions on, 2012. **23**(4): p. 631–643.
33. Song, L., et al., *Feature Selection via Dependence Maximization.* Journal of Machine Learning Research, 2012.
34. Oleshchuk, V. and V. Zadorozhny, *Trust-Aware Query Processing in Data Intensive Sensor Networks*, in *International Conference on Sensor Technologies and Applications*. 2007.