

Wagner, Christian and Miller, Simon and Garibaldi, Jonathan M. and Anderson, Derek T. and Havens, Timothy C. (2014) From interval-valued data to general type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems*, 23 (2). pp. 248-269. ISSN 1063-6706

**Access from the University of Nottingham repository:**

[http://eprints.nottingham.ac.uk/29176/1/TFS\\_IntervalsToT2Sets\\_Final\\_06March14\\_TwoColumns.pdf](http://eprints.nottingham.ac.uk/29176/1/TFS_IntervalsToT2Sets_Final_06March14_TwoColumns.pdf)

**Copyright and reuse:**

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

- Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners.
- To the extent reasonable and practicable the material made available in Nottingham ePrints has been checked for eligibility before being made available.
- Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
- Quotations or similar reproductions must be sufficiently acknowledged.

Please see our full end user licence at:

[http://eprints.nottingham.ac.uk/end\\_user\\_agreement.pdf](http://eprints.nottingham.ac.uk/end_user_agreement.pdf)

**A note on versions:**

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact [eprints@nottingham.ac.uk](mailto:eprints@nottingham.ac.uk)

# From Interval-Valued Data to General Type-2 Fuzzy Sets

Christian Wagner, *Senior Member, IEEE*, Simon Miller, *Member, IEEE*, Jonathan M. Garibaldi, *Member, IEEE*,  
Derek T. Anderson, *Member, IEEE*, Timothy C. Havens, *Senior Member, IEEE*,

**Abstract**—In this paper a new approach is presented to model interval-based data using Fuzzy Sets (FSs). Specifically, we show how both crisp and uncertain intervals (where there is uncertainty about the endpoints of intervals) collected from individual or multiple survey participants over single or repeated surveys can be modelled using type-1, interval type-2, or general type-2 FSs based on zSlices. The proposed approach is designed to minimise any loss of information when transferring the interval-based data into FS models, and to avoid, as much as possible assumptions about the distribution of the data. Furthermore, our approach does not rely on data pre-processing or outlier removal which can lead to the elimination of important information. Different types of uncertainty contained within the data, namely intra- and inter-source uncertainty, are identified and modelled using the different degrees of freedom of type-2 FSs, thus providing a clear representation and separation of these individual types of uncertainty present in the data. We provide full details of the proposed approach, as well as a series of detailed examples based on both real-world and synthetic data. We perform comparisons with analogue techniques to derive fuzzy sets from intervals, namely the Interval Approach (IA) and the Enhanced Interval Approach (EIA) and highlight the practical applicability of the proposed approach.

**Index Terms**—Survey data, zSlices, uncertainty, Computing With Words, Type-2, agreement, Interval Agreement Approach

## I. INTRODUCTION

Surveying groups of individuals on an area of interest is an important means of eliciting information and knowledge. By combining people’s opinions, overall opinions of groups and sub-groups of those people can be produced. Further, the dynamics (and variations) in decision-making among those surveyed can be studied, represented statistically, and models of the underlying relationships in the data and between those surveyed can be created. Employing survey tools that enable the capturing of the uncertainty in the responses of individuals (e.g., [1], [2], [3] and [4]) and capturing this uncertainty in the resulting models allows, for example, the creation of expert

systems for automatic reasoning which take uncertainty levels into account, as well as decision support systems which provide recommendations while highlighting relevant uncertainty levels in the underlying data. Fuzzy logic, in particular Type-2 (T2) fuzzy logic, has proven to be both a useful and popular method of modelling the opinions and perceptions of groups of people. As part of the recent surge in interest in the Computing with Words (CW) paradigm, a number of techniques have been investigated in order to capture user opinions of words expressed through surveys, including [1], [3]–[7].

In general terms, uncertainty in survey data takes many forms including the uncertainty that a specific participant has about their answer (for example, through lack of information/knowledge), variation amongst the individual opinions of a group of participants (*inter-expert uncertainty*), and variation in the opinions of a particular participant (*intra-expert uncertainty*). The latter type of uncertainty can itself be split into two independent sources of uncertainty. First, variation in the participants’ responses over time, i.e. participants provide different responses when queried repeatedly (they may have new or better information about the problem domain, may have forgotten specific details, may be in a different mood, etc.). Second, participants may be uncertain about the endpoints of a given interval response, i.e., there may be indecision about the exact level of uncertainty.

While previous approaches to model survey-based data using T2 FSs, such as the Interval Approach (IA) [5] and Enhanced Interval Approach (EIA) [1], [2], have shown great promise, they require data pre-processing (e.g., outlier removal), and/or depend on specific FS models, such as triangular or trapezoidal membership functions. Furthermore, most current approaches combine all different types of uncertainty (e.g., inter- and intra-expert uncertainty) in one uncertainty model, such as the Footprint of Uncertainty (FOU) of Interval Type-2 (IT2) FSs.

In this article, we extend our previous work [3] [4] by detailing how zSlices-based General Type-2 (zGT2) FSs [8] can be used to model the different aspects of inter- and intra-expert uncertainty. Specifically, we propose a novel approach to capture and model survey-based uncertainty which requires no pre-processing of the data, nor the prior definition of a specified membership function (MF) type. While no pre-processing is required in our model, this does not preclude the removal of data known to be ‘bad,’ if desired, to protect the veracity of the constructed model. However, in our experience with real-world survey data, it is often hard to distinguish whether an outlier is ‘bad’ or represents a valid view which

This work was part funded by RCUK’s Horizon Digital Economy Research Hub grant, EP/G065802/1, the EPSRCs Towards Data-Driven Environmental Policy Design grant, EP/K012479/1 and CESG - the UK Government’s National Technical Authority for Information Assurance (IA).

Christian Wagner, Simon Miller and Jonathan M. Garibaldi are with the Intelligent Modelling and Analysis Group and the Horizon Digital Economy Research Institute, School of Computer Science, University of Nottingham, Nottingham, UK. (email: christian.wagner@nottingham.ac.uk, s.miller@nottingham.ac.uk, jon.garibaldi@nottingham.ac.uk)

Derek T. Anderson is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. (email: anderson@ece.msstate.edu)

Timothy C. Havens is with the Department of Electrical and Computer Engineering, Department of Computer Science, Michigan Technological University, Houghton, MI, USA. (email: thavens@mtu.edu)

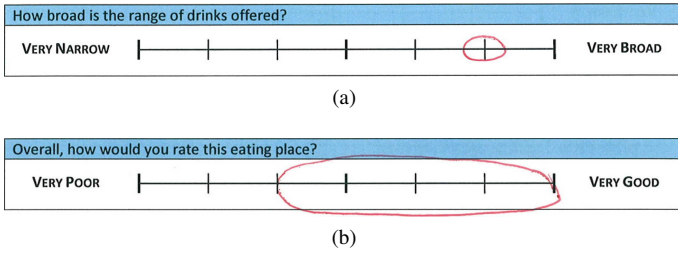


Fig. 1. Example interval responses (from a real survey), where ‘a’ is a less uncertain response and ‘b’ is a more uncertain response

is different from the consensus (e.g., novel insight). Thus, our method provides flexibility to the model creator by making pre-processing an option, rather than a requirement.

Throughout this article, in order to capture uncertainty during data collection, we employ a survey design in which participants can express their uncertainty about a given response by specifying an interval, rather than specifying or choosing a crisp point such as on a Likert scale. The width of the interval denotes the participant’s certainty in their answer, a narrow interval is used when they are sure where on the scale the answer lies, and a wider one where they are less certain. Figure 1 provides examples of less and more uncertain responses collected in our surveys. Our approach leverages the different degrees of freedom of zGT2 FSs to individually model the different types of uncertainty cited above. We describe which types of uncertainty the proposed approach aims to model (i.e., intra- and inter-expert uncertainty) and show how each type is captured in the resulting FSs. For the remainder of this article we refer to our proposed method as the Interval Agreement Approach (IAA).

We show how to construct zGT2 FSs [8] from data based on both synthetic data (to demonstrate the advanced features of the IAA method) and real-world data samples, and detail how the resulting FS models can be used in practical applications. Specifically, we illustrate the use of IAA in three case studies / worked examples before detailing a practical application. Case one features a real-world data set collected as part of an ongoing research project between the University of Nottingham and CESG, the UK Government’s National Technical Authority for Information Assurance, which is exploring the uncertainty in cyber-security experts’ assessments of real-world security threats. Case two features data collected within the University of Nottingham concerning the rating of local restaurants. This data set is then synthetically extended to incorporate uncertain endpoints (of the intervals) and is used in the third case. Finally, we describe a set of sample applications (also discussed in [4]) which employ the proposed IAA approach in the context of linguistic summarization and CW.

The paper is structured as follows. Section II provides background on type-2 FSs and the IA/EIA [1], [5] methods. Section III introduces and details the IAA method. Section IV shows a series of numeric examples for different data including comparisons to the results generated when employing the IA/EIA approaches. It is followed by a real world application example of the IAA in Section V. Finally, Section VI provides a discussion of the IAA approach in the general context of transforming interval data to fuzzy sets while Section VII

explores the conclusions and considers directions of future work.

## II. BACKGROUND

In this section we succinctly review IT2 and GT2 FSs [9], zGT2 FSs [8], and current methods employed to design IT2 FSs from survey data.

### A. General and Interval Type-2 Fuzzy Sets

GT2 FSs [9] were initially introduced by Zadeh [10] as a generalisation of type-1 (T1) FSs. Instead of a single point of membership  $\mu$  at each point on the domain  $x$ , the membership is defined itself as a T1 FS. This increased level of complexity of GT2 FSs, while enabling a more fine-grained and potentially more accurate modelling of uncertainty, results in a considerable increase in computational complexity which has only recently been addressed through the inception of new forms of representation for GT2 FSs, including [8], [11], [12] and [13]. IT2 FSs [14] were introduced as a simplification of GT2 FSs where the primary membership is defined as an interval  $[\underline{y}, \bar{y}]$ , where  $\underline{y}$  and  $\bar{y}$  are the degrees of membership of  $x$  in the lower membership function (LMF) and upper membership function (UMF), respectively [15]. It is worth noting that both the UMF and LMF are T1 FSs.

It is common practice in the fuzzy logic community to refer to the membership of a T1 FS as  $\mu$  while for T2 FSs, the primary membership is referred to as  $u$  and the secondary membership is referred to as  $\mu$  [15]. As we are employing the individual dimensions of the FSs to model different types of uncertainty and are employing both T1 and T2 FSs, we generally refer to the primary membership of all FSs as  $y$  (being on the  $y$ -axis) and the secondary membership as  $z$  (being on the  $z$ -axis) to avoid any confusion. An IT2 set  $\tilde{A}$  is fully described by its UMF and LMF and can be written as:

$$\tilde{A} = \int_{x \in X} \int_{y \in [\underline{y}_x, \bar{y}_x]} 1/(x, y), \quad (1)$$

where the integral notation describes the union.

### B. zSlices and zSlices based General Type-2 Fuzzy Sets

A zSlice [8] is formed by slicing a GT2 FS in the third dimension ( $z$ ) at level  $z_i$ . This slicing action results in an interval set in the third dimension with height  $z_i$ . As such, a zSlice  $\tilde{Z}_i$  is equivalent to an IT2 FS with the exception that its membership grade in the third dimension is not fixed to 1 but is equal to  $z_i$  where  $0 \leq z_i \leq 1$ . Thus a zSlice  $\tilde{Z}_i$  can be written as follows:

$$\tilde{Z}_i = \int_{x \in X} \int_{y_i^x \in [\underline{y}_i^x, \bar{y}_i^x]} z_i/(x, y_i^x), \quad (2)$$

where at each  $x$  value (shown in Fig. 2a), zSlicing creates an interval set with height  $z_i$  and domain  $[y_i^x, \bar{y}_i^x] \subset [0, 1]$  as shown in Fig. 2b,  $1 \leq i \leq I$ , where  $I$  is number of zSlices (excluding  $\tilde{Z}_0$ ) and  $z_i = i/I$ . A GT2 FS  $\tilde{F}$  can be seen as equivalent to the collection of an infinite number of zSlices (3),

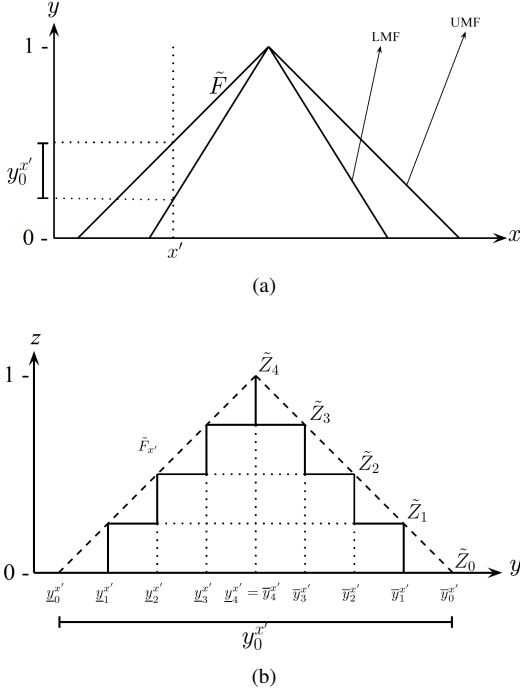


Fig. 2. zSlices based GT2 FSs, where (a) is a front view of a zGT2 set  $\tilde{F}$  with  $I = 4$  and (b) is the third dimension (vertical slice) at  $x'$ . Note, in (b) the horizontal axis represents  $y$  while in (a) the horizontal axis represents  $x$ .

$$\tilde{F} = \int_{0 \leq i \leq I} \tilde{Z}_i, \quad I \rightarrow \infty. \quad (3)$$

In the discrete case, (3) can be rewritten as

$$\tilde{F} = \sum_{i=0}^I \tilde{Z}_i \quad (4)$$

We will refer to the discrete version at (4) throughout the paper. It should be noted that the summation sign in (4) does not denote arithmetic addition but the fuzzy union operation as discussed in [15]. Figure 3 shows a 3-dimensional plot of a GT2 FS (shown in Fig. 3a and Fig. 3c) that is represented as a zGT2 FS (Fig. 3b and Fig. 3d) with  $I = 3$ . More details on the operations of zSlices based GT2 FSs are available in [8].

### C. The Interval and Enhanced Interval Approaches

The IA [5] introduced by Liu and Mendel is a procedure for taking intervals describing words and building IT2 FSs that represent those intervals/words. It is part of a recent research effort towards CW which aims to employ FSs to model words (generally adjectives) and thus make them suitable for computation and approximate reasoning (e.g., see [6], [16], [17]). The IA method is composed of two distinct parts, the *data part* and the *fuzzy set part*.

In the data part, intervals are pre-processed and data statistics are produced for the intervals that ‘survive’ the pre-processing phase. The pre-processing phase eliminates outliers, non-sensical data, data that do not fall within a tolerance threshold, and intervals that do not overlap with another interval. In the next stage the remaining intervals are assigned

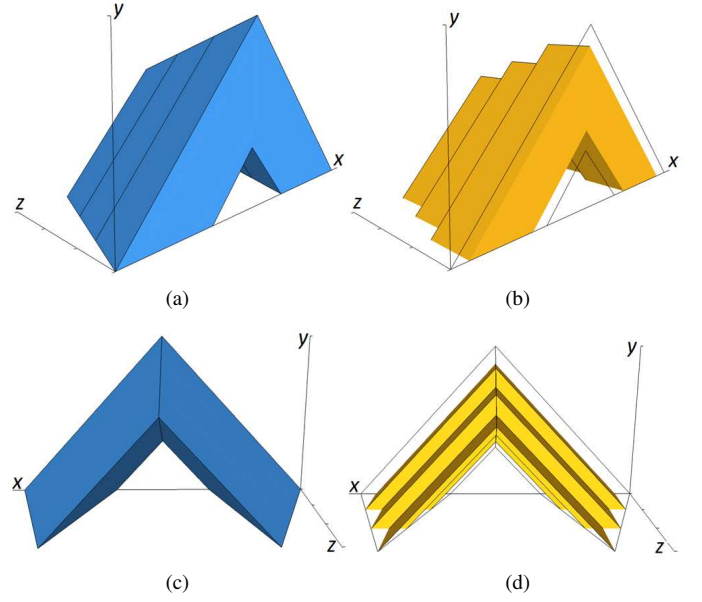


Fig. 3. (a) Side view of a GT2 FS, indicating three zLevels on the third dimension. (b) Side view of the zSlices version (zGT2 set) of the set in (a), with  $I = 3$ . (c) Tilted rear/below view of the GT2 set in (a). (d) Tilted rear/below view of the zGT2 set in (b).

a probability distribution and data statistics are produced [5]. The mean and standard deviation are used to calculate data statistics which are used in the FS part.

The FS part of the IA method [5] involves nine steps:

- 1) Select a T1 FS model - either a left-shoulder, symmetrical triangle or right shoulder T1 MF.
- 2) Establish FS uncertainty measures - In this case the uncertainty measures are the mean and standard deviation.
- 3) Compute uncertainty measures for T1 FS models - Compute the mean and standard deviation for the T1 FS.
- 4) Compute general formulas for parameters of T1 FS models - Equate the mean and standard deviation of a T1 FS to each data interval.
- 5) Establish nature of FOU - Classify the set of data intervals as either an interior, left-shoulder or right-shoulder FOU, depending on the mean values of the left and right end points of the data intervals.
- 6) Compute Embedded T1 FSs - Map the remaining data intervals to their respective T1 FSs.
- 7) Delete inadmissible T1 FSs - Discard embedded T1 FSs with a support that falls outside the desired range.
- 8) Compute an IT2 FS using the ‘union’ - Create an IT2 FS by taking the ‘union’ of the T1 FSs.
- 9) Compute mathematical model for FOU - Approximate the UMFs and LMFs of the FOU.

In [5], the IA method is demonstrated with survey data containing peoples’ opinions on the meaning of a set of linguistic terms such as ‘very little’ and ‘a lot.’ An Enhanced version of the IA method (EIA) is presented in [1]. In this version, the pre-processing stage is altered so that a narrower FOU is produced, and the calculation of the interior LMF is altered. Pre-processing has been changed as (while using data from a large online survey of people not necessarily

familiar with Fuzzy Logic or the purpose of the survey) the authors found that the IT2 FSs produced for words using the standard IA were much wider than those observed originally in [5]. The EIA is designed to produce IT2 sets that capture the uncertainty of words while retaining the basic shape of the trapezoidal or shoulder FOU. We focus on the IA/EIA approaches as a basis for comparison in this paper as both IA/EIA have a similar aim of translating interval-based data into FSs as the proposed IAA approach.

### III. THE INTERVAL AGREEMENT APPROACH

This section introduces a novel process of generating FSs from interval-valued data, called the Interval Agreement Approach (IAA). The approach is built on top of previously published work in [3], [4], [18]. Section III-A reviews the specific aims and motivation behind the IAA, followed in Sections III-B, III-C and III-D by in-depth descriptions of the individual stages involved, namely:

- capturing the information from one or more intervals (with crisp or uncertain interval endpoints) collected from a *single source* (e.g., single-expert responses over one or multiple surveys)
- capturing the information from one or more intervals (with crisp or uncertain interval endpoints) collected from *multiple sources* (e.g., multi-expert responses over one or multiple surveys).

We consider *crisp* intervals (no uncertainty about the interval endpoints) and *uncertain* intervals (each endpoint is uncertain and itself modelled as a crisp interval), as shown in Fig. 4. Throughout the paper we denote crisp intervals with an overbar, e.g.,  $\bar{A}$ , with the left and right endpoints as  $l_{\bar{A}}$  and  $r_{\bar{A}}$ . We denote uncertain intervals with an umlaut to indicate the interval nature of the left and right endpoints, e.g.,  $\ddot{B}$ . The crisp intervals representing the left and right ends of the uncertain interval are denoted as  $[l_{\bar{l}_{\ddot{B}}}, r_{\bar{l}_{\ddot{B}}}]$  and  $[l_{\bar{r}_{\ddot{B}}}, r_{\bar{r}_{\ddot{B}}}]$ , respectively (see Fig. 4).

Numeric examples and comparisons of the IAA and the IA/EIA are provided in Section IV.

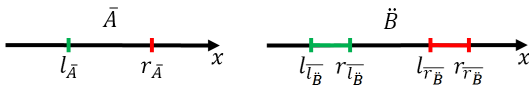


Fig. 4. Crisp ( $\bar{A}$ ) and uncertain ( $\ddot{B}$ ) intervals (where for the latter each endpoint is itself an interval).

#### A. Motivation

The aim of the IAA is to accurately model information captured through the collection of crisp or uncertain intervals (e.g., through surveys). Considering the latter, it is clear that one could continue the process of modelling the uncertainty about endpoints indefinitely, i.e., model the uncertainty about each endpoint of the interval that itself captures the uncertainty about an interval endpoint. However, it invariably becomes both challenging to provide an intuitive interpretation for such an approach and difficult to justify the required effort (computational and in terms of data-collection), particularly

when considering the diminishing returns in terms of relating modelling quality to modelling effort. Hence, we present a framework which allows the modelling of:

- individual interval-based data items such as survey responses (rather than just crisp points);
- uncertainty about the endpoints of those intervals (i.e., the modelling of uncertain intervals);
- uncertainty about the above from a single source over one or more data capture cycles (e.g., surveys);
- uncertainty arising from responses of multiple sources over one or more data capture cycles / surveys.

The IAA enables these capabilities while minimising the number and scope of assumptions during the model-creation process. In concrete terms, it generates non-parametric FS models of the interval-based data without preselecting a specific FS type (such as triangular or Gaussian) and minimises the loss of information during the model creation process (from the original, potentially uncertain intervals). Finally, the IAA leverages the different degrees of freedom of zGT2 FSs to represent and to distinguish the different types of uncertainty. Specifically, agreement/variation over multiple interval samples (e.g., survey responses) from the same source (e.g., expert) is commonly captured using the primary memberships ( $y \in [0, 1]$ ), with lower and upper bounds of primary membership modelling potential uncertainty about interval endpoints. Uncertainty originating from the agreement/variation between multiple sources (e.g., experts) of information is modelled through the secondary memberships of the zGT2 FSs ( $z \in [0, 1]$ ). Note that the choice of in which dimension to model which type of uncertainty is arbitrary and either choice is valid - as long as data collection and the interpretation of the resulting sets is modified accordingly. To illustrate, consider the case where only a single sample/survey was administered to a large number of subjects; here, it is preferable to model inter-expert uncertainty using the primary membership, resulting in simpler and easier interpretable (type-1 FS) models.

#### B. Method

Using multiple sets of interval-valued data in a survey context, we can model to what extent participants agree with their own previously surveyed responses (intra-expert uncertainty), and to what extent they agree with the opinions of other candidates in the survey group (inter-expert uncertainty). As noted, the two types of uncertainty are represented individually through the primary and secondary memberships of (GT2) FSs. Two stages comprise the overall method, the first to model the intra-expert uncertainty using T1 and IT2 FSs, and the second to combine the individual participant-specific FSs into a single GT2 FS while incorporating inter-expert uncertainty information. The resulting FSs capture both intra- and inter-expert uncertainty. Both stages are described in Sections III-C and III-D. As stated before, it is worth noting that the decision of which type of uncertainty is modelled using which dimension is a design choice made based on a given application.



### C. Modelling interval-based data from a single source

In this section, we consider two variations of modelling interval-based data from a single source (e.g., an expert): the case of crisp intervals and the case of “uncertain” intervals, i.e. where each endpoint of the interval is expressed itself as an interval to represent the uncertainty about the specific endpoint. Each of the two cases is considered for single and multiple samples (e.g., survey iterations) per information source.

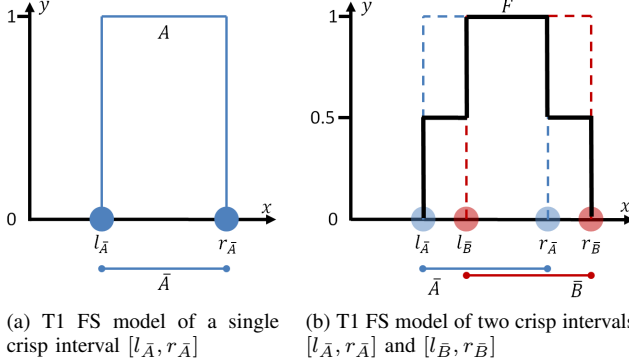


Fig. 5. Modelling crisp interval(s) from a single source using the IAA (no uncertainty about interval endpoints). Note that the order of the intervals does not matter as the operation extracts n-tuple agreement.

1) *Modelling one or more crisp intervals from a single source:* For a given information source (e.g., an expert), a T1 FS is created on the basis of the provided crisp interval(s) (see Fig. 5a) which represents the expert’s opinion/response over one or more survey iterations. In other words, the resulting T1 FS encodes the intra-expert uncertainty present in the survey data over multiple survey iterations. The degree of membership  $y$  of the set over the survey domain  $x$  captures the number of intervals that are overlapping (are in agreement) at a particular point. For example, consider a series of intervals  $\bar{A}_n$ ,  $n \in \{1, \dots, N\}$ , where  $N$  is the number of intervals/survey iterations. In order to combine the intervals  $\bar{A}_n$  we create a T1 FS  $A$  defined by the MF  $\mu(A)$  as described at (5).

$$\begin{aligned} \mu(A) = & y_1 / \bigcup_{i_1=1}^N \bar{A}_{i_1} \\ & + y_2 / \left( \bigcup_{i_1=1}^{N-1} \bigcup_{i_2=i_1+1}^{N-1} (\bar{A}_{i_1} \cap \bar{A}_{i_2}) \right) \\ & + y_3 / \left( \bigcup_{i_1=1}^{N-2} \bigcup_{i_2=i_1+1}^{N-1} \bigcup_{i_3=i_2+1}^N (\bar{A}_{i_1} \cap \bar{A}_{i_2} \cap \bar{A}_{i_3}) \right) \\ & + \dots \\ & + y_N / \left( \bigcup_{i_1=1}^1 \dots \bigcup_{i_N=N}^N (\bar{A}_{i_1} \cap \dots \cap \bar{A}_{i_N}) \right), \end{aligned} \quad (5)$$

where  $y_i = \frac{i}{N}$

Equation (5) employs the common notation of membership for FSs where  $/$  refers to the degree of membership, not division. Also, note that  $\cup$  and  $\cap$  denote the union, respectively

the intersection of intervals and  $+$  denotes the traditional union operation (using the *max* t-conorm) of the T1 FSs resulting from each part of the equation. Thus, the overall T1 FS is formed by the union of a sequence of components (each a T1 FS), where the first one is formed as the union of all intervals, associated with a primary membership of  $y_1$ , the second one as the union of all possible two-tuple intersections of intervals associated with  $y_2$ , etc. Figures 5(a,b) show the case of generating a T1 FS for single and multiple (two in this case) intervals respectively using the IAA. Note that when source intervals do not overlap, the resulting IAA model may be both non-convex and non-normal (see Fig. 21 for examples of such models). Algorithm 1 is a formal description of the T1 FS model generation. In [19], we proved that (5) can be efficiently calculated in a recursive manner (see Remark 1, page 3 of [19]). In practical applications the actual generation of the complete (continuous) T1 FS model can also often be avoided as the resulting set will be discretized. If this is the case, the discrete membership function can be formulated as

$$\begin{aligned} \mu_A(x') &= \frac{\left( \sum_{i=1}^N \mu_{\bar{A}_i}(x') \right)}{N}, \\ \mu_{\bar{A}_i}(x') &= \begin{cases} 1 & l_{\bar{A}_i} \leq x' \leq r_{\bar{A}_i}, \\ 0 & \text{else} \end{cases} \end{aligned} \quad (6)$$

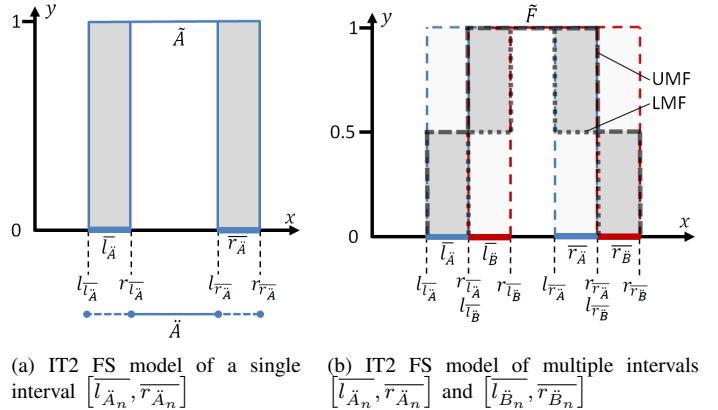


Fig. 6. Modelling uncertain interval(s) from multiple sources using the IAA - Uncertainty about interval endpoints (shaded)

2) *Modelling one or more uncertain intervals from a single source:* When information is provided in the form of an uncertain interval, an IT2 FS is employed to model this information gathered over one or more (survey) iterations from a single source. For example, consider a series of uncertain intervals  $\bar{A}_n$ ,  $n \in \{1, \dots, N\}$ , where  $N$  is the number of intervals/survey iterations. In order to combine the individual uncertain intervals we proceed similarly as for crisp intervals, however, we employ (5) independently for all “outer” and “inner” endpoints, resulting in the (T1) UMF and LMF of the IT2 FS model. Figures 6(a,b) illustrate the case of using the IAA to generate an IT2 FS for single and multiple (two in this case) uncertain intervals respectively, while numeric examples are provided in Section IV. It is worth noting that the crisp interval case can be considered a special case of the more general uncertain interval case where the left and right

---

**Algorithm 1** Generation of type-1 FS models from crisp-interval data
 

---

- 1: Input the  $\bar{A}_n, n \in \{1, \dots, N\}$  ▷ Where  $N$  is the number of intervals. We refer to the set of all intervals as  $S$
  - 2: Compute the primary degrees of membership  $y_i = i/N, i \in \{1, \dots, N\}$
  - 3: **for** each  $y_i$  **do**
  - 4: Calculate the intersections  $INT_b^i, b \in \{1, \dots, B_i\}$  for all  $i$ -combinations of  $S$  ▷ Where  $B_i$  is the binomial coefficient  $\binom{N}{i}$ . For example, consider the set  $S = \{\bar{A}, \bar{B}, \bar{C}\}$ ,  $B_1 = 3, INT_1^1 = \bar{A}, INT_2^1 = \bar{B}, INT_3^1 = \bar{C}; B_2 = 2, INT_1^2 = \bar{A} \cap \bar{B}, INT_2^2 = \bar{A} \cap \bar{C}, INT_3^2 = \bar{B} \cap \bar{C}$  and  $B_3 = 1, INT_1^3 = \bar{A} \cap \bar{B} \cap \bar{C}$ .
  - 5: Calculate the union  $\bar{U}_i$  over all intersections  $INT_b^i$ . ▷ Note that the resulting interval may be non-continuous.
  - 6: Associate  $\bar{U}_i$  with the primary membership degree  $y_i$ , i.e.  $\forall x \in \bar{U}_i, \mu_{\bar{U}_i}(x) = y_i$  and  $\forall x \notin \bar{U}_i, \mu_{\bar{U}_i}(x) = 0$  to create the T1 FS  $U_i$ .
  - 7: **end for**
  - 8: Calculate the union over all interval FSs  $U_i$  to generate the output/overall T1 FS.
- 

endpoints defining the endpoints of a given uncertain interval are equal and therefore, so are the UMF and LMF of the IT2 FS; thus the IT2 FS simplifies to a T1 FS (considering that the secondary membership of IT2 FSs can be omitted).

#### D. Modelling interval-based data from multiple sources

While the previous section focused on using the IAA to model data collected from individual sources (experts) (i.e. intra-source/intra-expert uncertainty), this section focuses on how the IAA can be used to aggregate those single-source models to capture inter-source (inter-expert) uncertainty. Similarly to the previous section, we consider the case of crisp and uncertain intervals separately in the respective subsections below.

1) *Modelling one or more crisp intervals from multiple sources:* In order to capture the intra-source uncertainty (variation and agreement) between multiple sources, all individual-source based T1 FS models can be aggregated into a GT2 FS which subsequently provides an overall model of the intra- and inter-source uncertainties across the (surveyed) sources. The resulting zGT2 FS represents both the intra- and inter-source variability in two separate domains, specifically the primary and secondary memberships of the zGT2 FS. The process of creating this set shares many similarities with the single-source model creation process illustrated in Section III-C. The set is computed using (7), where each zSlice is calculated separately and the number of zSlices is equal to the number of experts. Again,  $/$  refers to the degree of membership,  $\cup$  and  $\cap$  denote the union, respectively the intersection of intervals and  $+$  denotes the union operation for zGT2 FSs (using the *max*-t-conorm) [20]. The corresponding algorithm is provided in Algorithm 2. For more information on zSlices or zGT2 FSs, see [8]. Similarly as in (6), a recursive calculation can be employed and for discrete values of  $x$ , a rapid form to compute (7) is possible but for space consideration in this paper we will describe this in more detail in a future publication.

$$\begin{aligned}
 \mu(\tilde{A}) = & z_1 / \left( \bigcup_{i_1=1}^N A_{i_1} \right) + z_2 / \left( \bigcup_{i_1=1}^{N-1} \bigcup_{i_2=i_1+1}^N (A_{i_1} \cap A_{i_2}) \right) \\
 & + z_3 / \left( \bigcup_{i_1=1}^{N-2} \bigcup_{i_2=i_1+1}^{N-1} \bigcup_{i_3=i_2+1}^N (A_{i_1} \cap A_{i_2} \cap A_{i_3}) \right) \\
 & + \dots \\
 & + z_N / \left( \bigcup_{i_1=1}^1 \dots \bigcup_{i_N=N}^N (A_{i_1} \cap \dots \cap A_{i_N}) \right),
 \end{aligned} \tag{7}$$

where  $z_i = \frac{i}{N}$

For example, consider a series of T1 FSs  $A_n$ , where  $n \in \{1, \dots, N\}$  and  $N$  is the number of sources/sets/experts. In order to fuse the T1 FSs into a GT2 FS  $\tilde{A}$ , we employ the agreement principle [18] and associate a higher secondary membership (zLevel) to areas where the T1 FSs overlap, effectively weighting “areas” of high agreement highly. Thus, as shown in (7), the lowest zLevel  $z_1$  is associated with the union of all T1 FSs, the union of all 2-tuple intersections of T1FSs is associated with  $z_2$ , that of 3-tuple intersections with  $z_3$ , etc. Figure 14 provides a visual example of zGT2 FS model produced using the IAA based on multiple crisp intervals from multiple sources.

The zGT2 FSs produced using the IAA employ all of the information available to represent the agreement between a group of sources/individuals. No information is removed or added, and any assumptions (such as on the membership function type, e.g., triangular or Gaussian) are kept to a minimum. The agreement-based model focuses on agreement between sources where all sources (experts) are individually given an equal weighting, though giving extra weight to more reliable sources (and vice versa) would be possible. Outliers are purposely not removed but modelled at a low level of agreement, reducing their impact but preserving their input. As stated before, if we know that elements of a data set are bad we will remove them; however, in our work it is often not possible to tell whether data is bad or whether the participant is contributing new, previously unknown knowledge. If required, data pre-processing (as seen in [5] and [1]) could be applied to remove intervals regarded as non-sensical or not useful. An example of this has been included in Section V-A. Finally,

---

**Algorithm 2** Generation of zGT2 FS models from T1 FSs
 

---

- 1: Input the  $A_n, n \in \{1, \dots, N\}$  ▷ Where  $N$  is the number of T1 FSs. We refer to the set of all T1 FSs as  $S$
  - 2: Compute the secondary degrees of membership  $z_i = i/N, i \in \{1, \dots, N\}$
  - 3: **for** each  $z_i$  **do**
  - 4: Calculate the intersections  $INT_b^i, b \in \{1, \dots, B_i\}$  for all  $i$ -combinations of  $S$  ▷ Where  $B_i$  is the binomial coefficient  $\binom{N}{i}$ . For example, consider the set  $S = \{A, B, C\}$ ,  $B_1 = 3$ ,  $INT_1^1 = A$ ,  $INT_2^1 = B$ ,  $INT_3^1 = C$ ;  $B_2 = 3$ ,  $INT_1^2 = A \cap B$ ,  $INT_2^2 = A \cap C$ ,  $INT_3^2 = B \cap C$  and  $B_3 = 1$ ,  $INT_1^3 = A \cap B \cap C$ .
  - 5: Calculate the union  $U_i$  over all intersections  $INT_b^i$ . ▷ Note that the resulting T1 FS may be non-convex.
  - 6: Associate  $U_i$  with the secondary membership degree  $z_i$ , i.e.  $\forall x \in U_i, \mu_{U_i}(x) = y_i$  and  $\forall x \notin U_i, \mu_{U_i}(x) = 0$  to create the zSlice (GT2 FS)  $\tilde{U}_i$ .
  - 7: **end for**
  - 8: Calculate the union over all zSlices (GT2 FSs)  $\tilde{U}_i$  to generate the output zGT2 FS.
- 

it is worth mentioning that the described process may result in non-convex FSs. Where this is the case, this is a direct result of the data and we believe it is important to maintain this characteristic. In [4] we showed how such sets can be used in real world applications (we do not show additional examples here for space restrictions). In the future, we foresee significant potential in characterising the resulting sets as the respective distributions (e.g., bimodal vs. unimodal) contain valuable information about the nature of the consensus in the data.

2) *Modelling one or more uncertain intervals from multiple sources*: To generate a model capturing multiple source-specific uncertain intervals from multiple sources, the IT2 FSs generated for each source (see Section III-C2) are aggregated to create a zGT2 FS. As part of the aggregation process, (7) is applied twice, once for all source-specific UMFs (resulting in the UMFs of the respective zSlices) and once for all source-specific LMFs (resulting in the LMFs of the respective zSlices). The resulting zGT2 FS provides a model of both the intra-source (intra-expert) and inter-source (inter-expert) uncertainty for the given set of uncertain intervals. Figure 19 provides an example of a zGT2 FS model produced using the IAA based on multiple uncertain intervals from multiple sources.

#### IV. NUMERIC EXAMPLES

We provide three numeric examples illustrating the modelling of different types of interval-based survey data as discussed in Section III. Specifically, we focus on demonstrating the following:

- 1) modelling of individual crisp intervals (i.e., from a single survey) from multiple sources (Section IV-A);
- 2) modelling of multiple crisp intervals (i.e., from repeated surveys) from multiple sources (Section IV-B);
- 3) modelling of multiple uncertain intervals (i.e., from repeated surveys) from multiple sources (Section IV-C).

For each of the three cases we provide an overview of the specific dataset employed (based on real-world and synthetic survey data) as well as a comparison of the FS models produced with the IAA and the IA/EIA methods.

TABLE I  
EXAMPLE 1 - SINGLE INTERVALS FROM MULTIPLE SOURCES - DATA.

Expert	Interval	Interval Mean
<b>A</b>	[0.20,0.50]	0.35
<b>B</b>	[0.40,0.80]	0.60
<b>C</b>	[0.20,0.60]	0.40
<b>D</b>	[0.19,0.64]	0.415
<b>Mean of experts</b>	[0.248,0.66]	0.454

#### A. Example 1: Individual Crisp Intervals from Multiple Sources

In this example we use data collected during an exercise carried out in collaboration with CESG, the UK Government's National Technical Authority for Information Assurance. A group of cyber-security professionals was asked to rate 26 IT network components with regard to the difficulty of compromising or bypassing them. This was done using a series of questions relating to the security level provided by the component. For example, questions included "How mature is this type of technology?" and "How likely is it that there will be a publicly available tool that could help with this attack?" Each expert provided a single response to each question in the form of an interval contained in  $[0,1]$ . Based on the approach detailed in Section III-C, we show how the intervals provided by four individual experts can be modelled as T1 FSs and (employing the approach detailed in Section III-D1) how a cross-expert zGT2 FS model (encompassing inter-expert variability) can be created from the individual T1 FSs. For this example, the opinions of four (rather than all) experts were used as this allows demonstration of the IAA while limiting the overall complexity, allowing presentation of the actual calculations and visualisation of the results. Table I shows an example of the data/intervals collected for the question "Overall, how difficult would it be for an attacker to do this?" for one of the 26 components. Intervals are in the range 0 (Very Easy) to 1 (Very Hard).

Section III-C describes how one can create T1 FSs from the original crisp interval-based data. In this case, as shown in Table I, the data do not include uncertainty information about the interval endpoints, i.e., the intervals are crisp. Further, as each expert has only been surveyed once, only a single interval is available per expert, avoiding the need (or potential) to model intra-expert uncertainty by computing the agreement over multiple intervals collected from the same expert for the



same question. We address intra-expert uncertainty modelling in Section IV-B.

Figure 7 shows the T1 FSs generated using the IAA at Equation (5) for the intervals provided by each of the four experts (Table I). The models generated by employing the IA/EIA approach are included for comparison. Note that some assumptions (discussed in Appendix B) had to be made in order to generate models based on the same data using all three approaches.

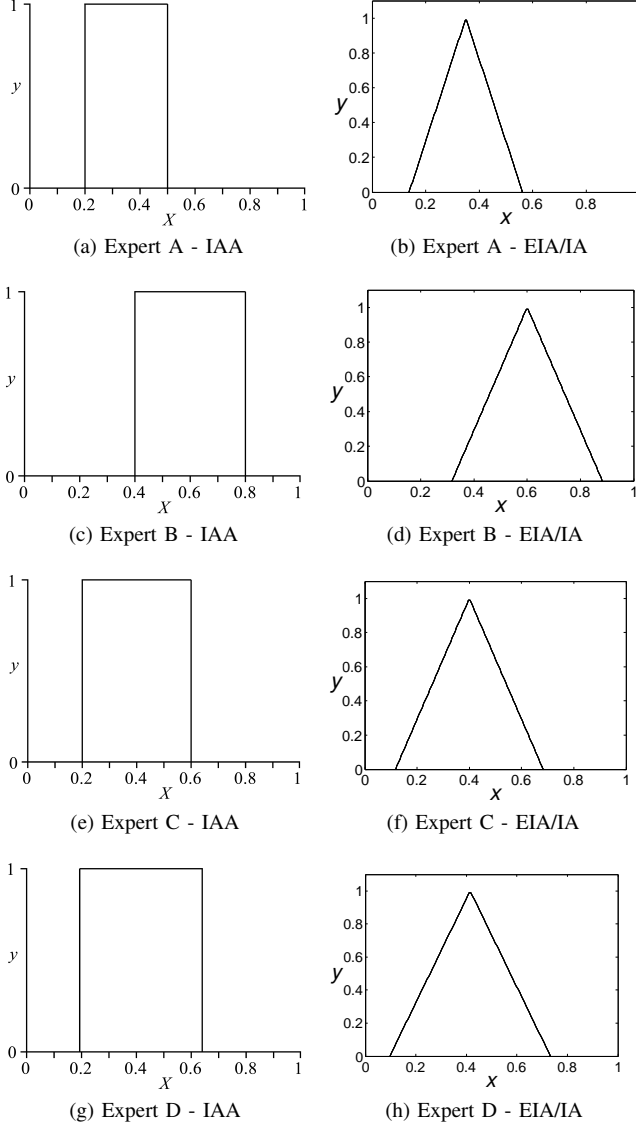


Fig. 7. Example 1 - T1 FSs for each individual expert using the IAA (left column) and the EIA/IA (right column). (For this dataset the EIA/IA result in the same models.)

From Fig. 7 it is worthwhile to note that the T1 FSs are a direct (FS based) representation of the interval data in Table I, i.e. no assumptions (such as a converting the interval into a distribution such as a triangular or Gaussian FS) have been made. Clearly, one could argue that assuming an equally weighted interval is itself making an assumption but for the limited information available (i.e. interval endpoints), this approach adheres to the principle of least commitment [21], [22]. Further, it can be argued that the resulting in-

terval T1 FSs are not “fuzzy” at all (i.e. their degree of membership is either 0 or 1). This is true by design — the intervals by themselves contain no (known) uncertainty, hence no uncertainty is present in the equivalent FS based model. As uncertainty is introduced in the data, such as through variation encountered based on repeated sampling or through the sampling of multiple sources (experts), *this* uncertainty information will be captured in the T1 and T2 FSs respectively as is detailed in the examples below.

After generating T1 FS models for each expert, we proceed to combine these T1 sets to produce a zGT2 FS that represents the inter-expert uncertainty as detailed in Section III-D1. As described, in order to create a model capturing the outputs of all four experts, the secondary membership domain is divided into four levels, one for each level of agreement between the four experts, giving secondary membership degrees of  $z_1 = 1/4, z_2 = 2/4, z_3 = 3/4$  and  $z_4 = 1$ . Clearly, different secondary memberships could be chosen to weight different levels of agreement in different ways; however, in the current paper we proceed with the most simple case of evenly distributing degrees of secondary membership across levels of agreement. The actual degree of secondary membership at a given  $x$  can be computed using the number of T1 sets (representing the separate expert’s opinions) that intersect at that particular point in the domain of  $x$  and  $y$ , following (7).

In our example, the GT2 FS  $\tilde{Z}$  consists of four zSlices  $\tilde{Z}_1, \tilde{Z}_2, \tilde{Z}_3$  and  $\tilde{Z}_4$ . Equations (8), (9), (10) and (11) describe how each of these slices are calculated using (7). Equation (12) shows the combination of the individual zSlices into the complete zGT2 FS  $\tilde{Z}$ . Table II provides the numeric details of all of the zSlices, as well as the centroids of each zSlice and the defuzzified value of the overall zGT2 set  $\tilde{Z}$ . Note that all of the centroids are effectively a crisp number (rather than an interval) as the original intervals were crisp, leading to T1 sets and consequently to zSlices where the UMF and LMF are identical. When the end points of the initial intervals collected are associated with uncertainty, this uncertainty can be modelled using the FOU as is shown in the subsequent examples. Figure 8 shows 2D representations of the zSlices at the respective secondary membership degrees of 0.25, 0.5, 0.75 and 1. Figure 9 shows the same slices in 3D and Fig. 10a shows a rear view of the full zGT2 FS  $\tilde{Z}$ , representing all of the experts opinions. Figure 10b shows a front view of the same set.

$$\begin{aligned} \tilde{Z}_1 &= 0.25 / (1 / ([0.20, 0.50] \cup [0.40, 0.80] \cup [0.20, 0.60] \\ &\quad \cup [0.19, 0.64])) \\ &= 0.25 / (1 / [0.19, 0.80]) \end{aligned} \quad (8)$$

$$\begin{aligned} \tilde{Z}_2 &= 0.5 / (1 / (([0.20, 0.50] \cap [0.40, 0.80]) \cup ([0.20, 0.50] \\ &\quad \cap [0.20, 0.60]) \cup ([0.20, 0.50] \cap [0.19, 0.64]) \\ &\quad \cup ([0.40, 0.80] \cap [0.20, 0.60]) \cup ([0.40, 0.80] \\ &\quad \cap [0.19, 0.64]) \cup ([0.20, 0.60] \cap [0.19, 0.64]))) \\ &= 0.5 / (1 / [0.20, 0.64]) \end{aligned} \quad (9)$$

$$\begin{aligned}
\tilde{Z}_3 &= 0.75 / (1 / (([0.20, 0.50] \cap [0.40, 0.80] \cap [0.20, 0.60]) \\
&\cup ([0.20, 0.50] \cap [0.40, 0.80] \cap [0.19, 0.64]) \\
&\cup ([0.20, 0.50] \cap [0.20, 0.60] \cap [0.19, 0.64]) \\
&\cup ([0.40, 0.80] \cap [0.20, 0.60] \cap [0.19, 0.64]))) \\
&= 0.75 / (1 / [0.20, 0.60])
\end{aligned} \tag{10}$$

$$\begin{aligned}
\tilde{Z}_4 &= 1 / (1 / ([0.20, 0.50] \cap [0.40, 0.80] \cap [0.20, 0.60] \\
&\cap [0.19, 0.64])) \\
&= 1 / (1 / [0.40, 0.50])
\end{aligned} \tag{11}$$

$$\tilde{Z} = \tilde{Z}_1 \cup \tilde{Z}_2 \cup \tilde{Z}_3 \cup \tilde{Z}_4 \tag{12}$$

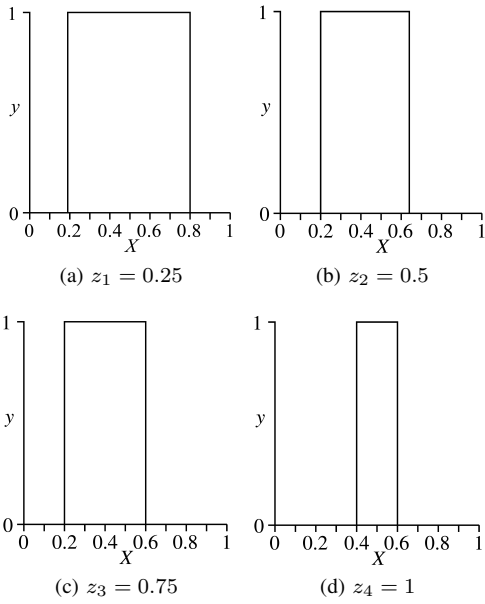


Fig. 8. Example 1 - 2D view of the zSlices produced using T1 sets and the agreement process for zGT2 FSs [18].

TABLE II  
EXAMPLE 1 - zSLICE PARAMETERS WITH INTERVALS AND ASSOCIATED PRIMARY AND SECONDARY MEMBERSHIPS.

	$z_1 = 0.25$	$z_2 = 0.5$	$z_3 = 0.75$	$z_4 = 1$
$y = 1$	[0.19,0.80]	[0.20,0.64]	[0.20,0.60]	[0.40,0.50]
<b>Centroid</b>	0.25/0.495	0.5/0.42	0.75/0.40	1/0.45
<b>Defuzzified</b>	0.434			

*Comparison:* In order to provide a comparison between the IAA and the IA/EIA, we employed the software accompanying the research papers of Liu and Mendel [5] and Coupland, Mendel and Wu [1] to generate FS models of the data. In order to use the provided code ‘as is’ some assumptions had to be made which are detailed in Appendix B. The sets generated for each individual expert using the IAA and IA/EIA approaches are shown in Fig. 7, while Figs. 10 and 11 show the sets generated to capture all four experts. Table III shows the centroids (and defuzzified values where applicable) of the FSs generated by the IAA and IA/EIA, both for the individual expert, as well as for the ‘‘all expert’’ models.

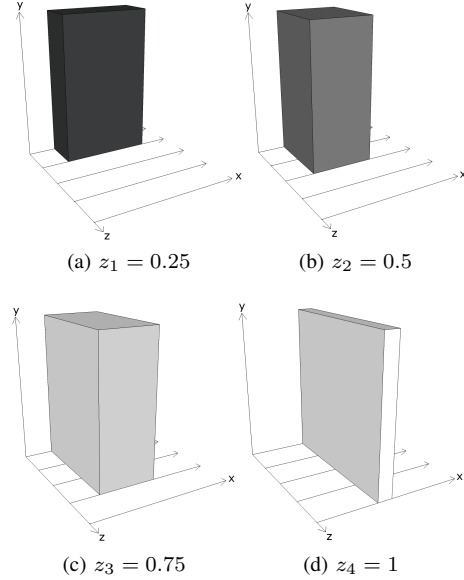


Fig. 9. Example 1 - 3D view of the zSlices produced using T1 sets and the agreement process for zGT2 FSs [18].

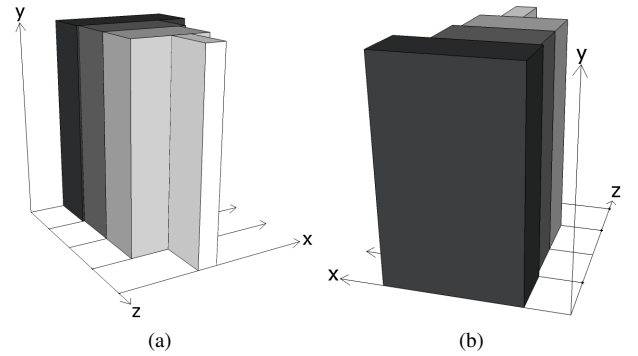


Fig. 10. Example 1 - zGT2 FS for all four experts, (a) is a rear view and (b) is a front view.

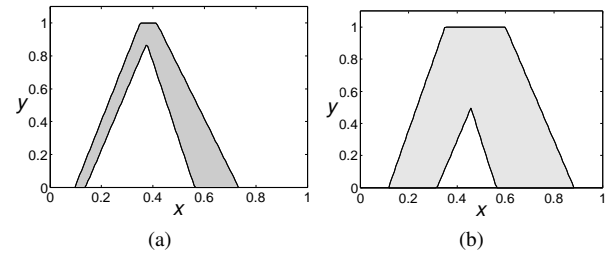


Fig. 11. Example 1 - IT2 FS for all four experts created using the (a) EIA and (b) IA.

Performing a one-to-one comparison, beyond a visual and numeric comparison, between the outputs of both methods is not trivial. While the individual sets for all four experts mainly differentiate each other through the type of membership function (the IA/EIA mapping the intervals to a triangular function), the overall models (Figs. 10 and 11) are significantly different. The IAA generates a zGT2 FS based on the agreement principle [18], indicating how at each increasing zLevel, the agreement diminishes (in terms of its support) but is valued

more highly (in terms of the secondary membership). Also, it can be noted that in the resulting set the primary degree of membership is 1 over the whole support, indicating that there is no intra-expert uncertainty. Only the secondary membership varies indicating the inter-expert uncertainty/variability. For the IA/EIA approaches, an overall output set based on a triangular IT2 FS membership function is generated, modelling uncertainty in its FOU.

### B. Example 2: Multiple Crisp Intervals from Multiple Sources

In this second example, we demonstrate how the extra uncertainty introduced by having multiple crisp intervals from each expert can be represented using the proposed IAA method. In this case, the extra information is provided by repeated surveys of the same individuals with the same questions, resulting in both intra-expert variation (across surveys) as well as inter-expert variation (across experts). As a real-world example, we use a survey conducted within the Intelligent Modelling and Analysis (IMA) research group at the University of Nottingham. The survey was conducted on students which, in contrast to the GCHQ security expert data, resulted both in the lack of confidentiality requirements and the straightforward possibility of repeated surveying. Participants were asked to rate various aspects of a series of restaurants in the Nottingham city centre area. For example, questions included “*How polite are the staff?*” and “*Overall, how would you rate this eating place?*” Ratings were provided using intervals to define where on a scale they thought the answer lay, and how certain/confident they were in their opinion (see example in Fig. 1). Participants’ ratings of restaurants were based on their memories of eating in each of the restaurants (i.e., they hadn’t necessarily visited each restaurant recently) which is clearly a contributory factor of uncertainty; if a participant has visited a restaurant once, a long time ago, unless the experience was extreme in some way, they are unlikely to have a clear recollection of all of the factors they are questioned about.

In order to provide an example of the survey results here that allows demonstrating the modelling approach while still being intuitive, we provide a small subset of the collected data, using the answers given by three participants on two separate occasions (approximately eight weeks apart) to one of the questions for one of the restaurants. Though using only a small amount of the available data, the example demonstrates the use of our method of representing crisp interval-based survey data from multiple sources collected over multiple surveys. We focus on the answers to the question ‘*How good is the food?*’, where the possible answers ranged from 0 (Very Poor) to 1 (Very Good). Table IV contains the (crisp interval) answers given by the participants on each occasion/survey.

TABLE IV  
EXAMPLE 2 - MULTIPLE INTERVALS FROM MULTIPLE SOURCES

Participant	1 <sup>st</sup> Answer	2 <sup>nd</sup> Answer
A	[0.51,0.84]	[0.42,0.77]
B	[0.23,0.87]	[0.17,0.87]
C	[0.62,0.70]	[0.31,0.91]

In the first stage, as in the previous example, the intervals for each participant are modelled with T1 FSs, capturing intra-participant uncertainty for each participant. However, unlike in the previous example, we now have multiple intervals (one from each survey iteration) for each expert. Therefore, the number of intervals for a particular participant that overlap at a particular point determines the degree of primary membership in the resulting T1 FS of this point. Specifically, as there are two sets of survey data, the primary membership domain ( $y$ ) is divided into two, giving membership degree levels of  $y_1 = 1/2$  and  $y_2 = 1$ . The membership function of the FS can then be calculated using (5). Equation (13) gives an example, using Participant A, of how (5) is used with the intervals from Table IV to create a T1 FS incorporating both surveys. Figure 12 shows the T1 FS created for each participant in the group, providing also a visual model of intra-participant variability. Table V provides the numeric detail of the sets where the intervals are associated with their respective primary membership grades (where a point is associated with more than one primary membership degree, the maximum degree applies). Also shown is the defuzzified centroid for each set which is provided for information and comparison purposes.

Using (5) with the intervals  $[0.51, 0.84]$  and  $[0.42, 0.77]$ , ( $y_1 = 0.5$  and  $y_2 = 1$ ), results in:

$$\begin{aligned} \mu(A) &= (y_1 / ([0.51, 0.84] \cup [0.42, 0.77]) \\ &\quad + y_2 / ([0.51, 0.84] \cap [0.42, 0.77])) \quad (13) \\ &= (0.5 / [0.42, 0.84] + 1 / [0.51, 0.77]) \end{aligned}$$

TABLE V  
EXAMPLE 2 - T1 FS PARAMETERS WITH INTERVALS AND ASSOCIATED PRIMARY MEMBERSHIP.

Participant	$y = 0.5$	$y = 1$	Defuzzified (overall)
A	[0.42,0.84]	[0.51,0.77]	0.637
B	[0.17,0.87]	[0.23,0.87]	0.54
C	[0.31,0.91]	[0.62,0.70]	0.643

After the creation of the T1 sets from the intervals during the first stage, we combine these T1 FSs to produce a zGT2 FS representing both the intra- and inter-expert uncertainty in the primary and secondary membership domains respectively. The secondary membership domain is divided into three zLevels, one for each level of agreement between the three participants, at membership degrees of  $z_1 = 1/3 \approx 0.33$ ,  $z_2 = 2/3 \approx 0.66$ , and  $z_3 = 3/3 = 1$  (see (7), [18]). The resulting zGT2 FS  $\tilde{Z}$  consists of three zSlices  $\tilde{Z}_1$ ,  $\tilde{Z}_2$  and  $\tilde{Z}_3$ . Equations (14), (15) and (16) describe how each of these slices are calculated using (7). Equation (17) shows their combination, producing the complete zGT2 FS  $\tilde{Z}$ . Table VI provides the numeric details of all of the zSlices, as well as the centroids of each zSlice and the defuzzified value of the overall zGT2 set  $\tilde{Z}$ . Again, all of the centroids are effectively a crisp number as the UMFs and LMFs of each zSlice are identical—there is no uncertainty about the interval endpoints in this example. When the end points of the initial intervals collected are associated with uncertainty, this uncertainty can be modelled using the FOU. This will be explored further in the next example. Figure 13 shows 2D representations of the zSlices at the respective

TABLE III  
EXAMPLE 1 - CENTROIDS AND DEFUZZIFIED VALUES OF FSS (SEE FIGS. 7, 10 AND 11) PRODUCED WITH THE EIA [1], IA [5] AND IAA

Expert	EIA		IA		IAA	
	Centroid	Defuzzified Centroid	Centroid	Defuzzified Centroid	Centroid	Defuzzified Centroid
A	0.35	0.35	0.35	0.35	0.35	0.35
B	0.60	0.60	0.60	0.60	0.60	0.60
C	0.40	0.40	0.40	0.40	0.40	0.40
D	0.415	0.415	0.415	0.415	0.415	0.415
All	[0.34,0.43]	0.38	[0.32,0.62]	0.47	[0.434,0.434]	0.434

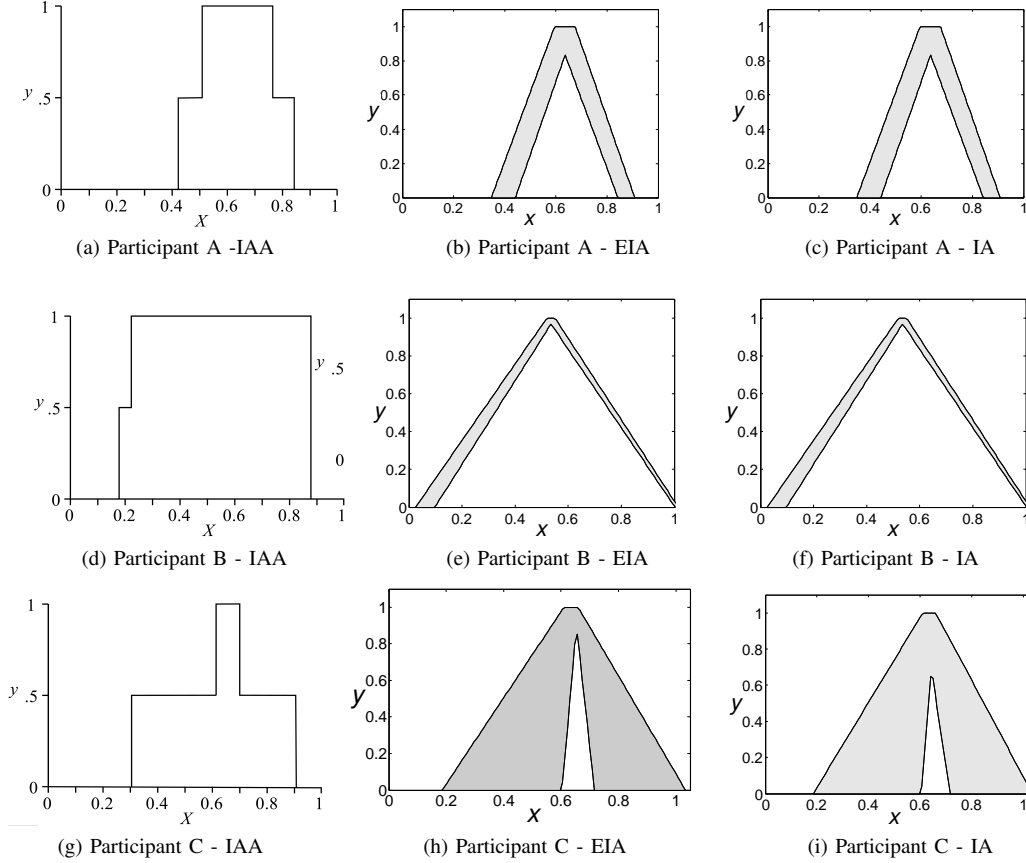


Fig. 12. Example 2 - T1 FSS for each participant over both surveys using the IAA (column 1), EIA (column 2) and the IA (column 3).

zLevels (secondary membership degrees) of 0.33, 0.66 and 1. Figures 14(a,b) show a rear and front view of the complete set  $\tilde{Z}$ .

$$\begin{aligned} \tilde{Z}_1 &= 0.33 / (y1 / ([0.42, 0.84] \cup [0.17, 0.87] \cup [0.31, 0.91]) \\ &+ y2 / ([0.51, 0.77] \cup [0.23, 0.87] \cup [0.62, 0.70])) \\ &= 0.33 / (0.5 / [0.17, 0.91] + 1 / [0.23, 0.87]) \end{aligned} \quad (14)$$

$$\begin{aligned} \tilde{Z}_2 &= 0.66 / (y1 / (([0.42, 0.84] \cap [0.17, 0.87]) \cup ([0.42, 0.84] \\ &\cap [0.31, 0.91]) \cup ([0.17, 0.87] \cap [0.31, 0.91])) \\ &+ y2 / ((([0.51, 0.77] \cap [0.23, 0.87]) \cup ([0.51, 0.77] \\ &\cap [0.62, 0.70]) \cup ([0.23, 0.87] \cap [0.62, 0.70]))) \\ &= 0.66 / (0.5 / [0.31, 0.87] + 1 / [0.51, 0.77]) \end{aligned} \quad (15)$$

$$\begin{aligned} \tilde{Z}_3 &= 1 / (y1 / [0.42, 0.84] \cap [0.17, 0.87] \cap [0.31, 0.91] \\ &+ y2 / ([0.51, 0.77] \cap [0.23, 0.87] \cap [0.62, 0.70])) \\ &= 1 / (0.5 / [0.42, 0.84] + 1 / [0.62, 0.70]) \end{aligned} \quad (16)$$

$$\tilde{Z} = \tilde{Z}_1 \cup \tilde{Z}_2 \cup \tilde{Z}_3 \quad (17)$$

TABLE VI  
EXAMPLE 2 - zSLICE DETAILS WITH INTERVALS AND ASSOCIATED PRIMARY AND SECONDARY MEMBERSHIPS.

zSlice	$\tilde{Z}_1$	$\tilde{Z}_2$	$\tilde{Z}_3$
	$z_1 = 0.33$	$z_2 = 0.66$	$z_3 = 1$
$y = 0.5$	[0.17,0.91]	[0.31,0.87]	[0.42,0.84]
$y = 0.1$	[0.23,0.87]	[0.51,0.77]	[0.62,0.70]
Centroid	0.33 / 0.5467	0.66 / 0.6233	1 / 0.65
Defuzzified $\tilde{Z}$	0.6239		

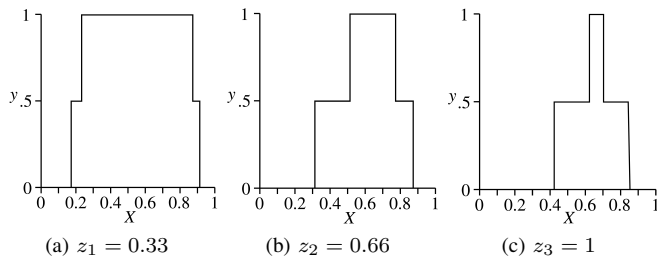


Fig. 13. Example 2 - 2D view of the zSlices produced with the IAA.

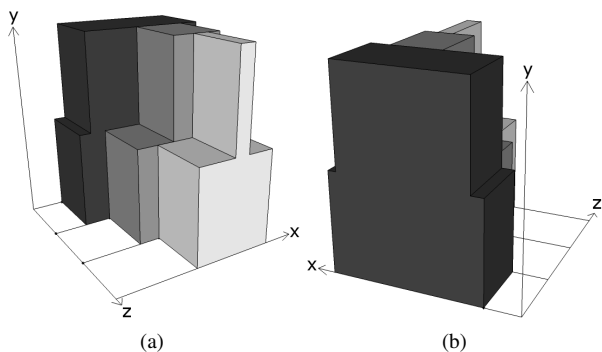


Fig. 14. Example 2 - zGT2 FS produced with the IAA where, (a) is a rear view and (b) is a front view.

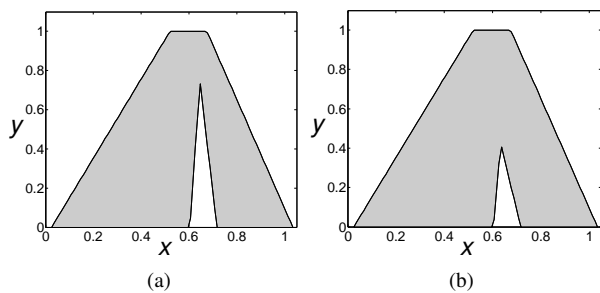


Fig. 15. Example 2 - IT2 FS produced with (a) EIA and (b) IA for all individuals.

It can be seen from the results that the IAA method allows the representation of multiple survey, multiple expert data with minimal loss or addition of information, i.e., the resulting representation uses all of the data available and is solely determined by the data. There is also a clear distinction in the representation of intra- and inter-expert uncertainty in the primary and secondary degrees of membership respectively.

*Comparison:* Again, we provide a comparison of IAA and IA/EIA. As in the previous example it was necessary to address some specific requirements of the software provided by the creators of IA/EIA [5], [1]—full details can be found in Appendix B. Figures 12 and 15 show the resulting sets for IA/EIA, and Table VII shows the centroids and the defuzzified centroid values of the FSs produced by IA/EIA and IAA respectively.

Comparing the individual models generated for each participant by the three approaches, it is clear how the IA/EIA approaches model the intra-participant uncertainty in the FOU of the resulting IT2 FSs while the IAA models it using the primary degree of membership of the T1 FSs. In the overall

cross-participant models (i.e., Figs. 14 and 15), as in the previous example, the difference in terms of type of FS (zGT2 FS and IT2 FS) is visualised. While the IA/EIA approaches model both intra- and inter-participant uncertainty using the FOU of the output set, from Fig. 14, it can be seen how the IAA method captures the intra-participant uncertainty arising from the two survey iterations in the primary membership and the inter-participant uncertainty arising from the three participants surveyed in the secondary membership.

### C. Example 3: Multiple Uncertain Intervals from Multiple Sources

In this synthetic example we simulate the case of four experts being asked to provide their answer to a given question using intervals on three separate occasions. The endpoints of the individual intervals have been associated with a 5% (of the total range of [0,1]) uncertainty evenly spread around the original interval endpoints. In a real-world scenario, this uncertainty information might be sourced directly from the candidates (who would be asked to provide a range capturing their uncertainty about each endpoint) or existing knowledge about participants (such as the level of experience) which could then be factored in through an injected level of uncertainty (e.g., low spread for highly experienced candidates and high spread for novices). Alternatively, direct collection of this information is possible by for example asking participants to provide minimum and maximum bounds using ellipses on a scale (as in Fig. 4). We will address such expanded survey designs in future work and focus on the described synthetically augmented dataset here.

Equation 18 shows how each pair of intervals is calculated, where  $\tilde{p}$  is the resulting pair of intervals,  $u$  is the uncertainty value and  $[a, b]$  is an expert's interval-valued opinion. Table VIII shows the intervals before and after the endpoint uncertainty has been applied. Further, Fig. 16 shows how the FS generated from the interval provided by candidate A for the first survey looks before (T1 FS) and after (IT2 FS) the endpoint uncertainty has been injected.

$$\tilde{p} = [[a - u, a + u], [b - u, b + u]] \quad (18)$$

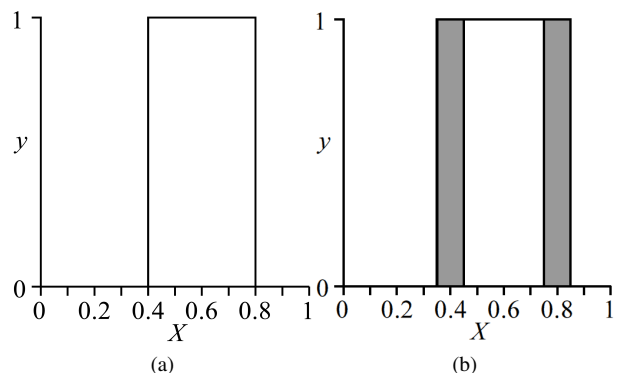


Fig. 16. Example 3 - FS based on intervals provided by Candidate 1, where, (a) is the original (crisp) interval and (b) is the (uncertain) interval after endpoint uncertainty has been applied.

TABLE VII  
EXAMPLE 2 - CENTROIDS AND DEFUZZIFIED VALUES OF FSS PRODUCED WITH THE EIA [1], IA [5] AND IAA

Expert	EIA		IA		IAA	
	Centroid	Def. Centroid	Centroid	Def. Centroid	Centroid	Def. Centroid
A (i.e. Figs. 12b, 12c and 12a)	[0.59,0.68]	0.635	[0.59,0.68]	0.635	0.6367	0.6367
B (i.e. Figs. 12e, 12f and 12d)	[0.52,0.55]	0.535	[0.52,0.55]	0.535	0.54	0.54
C (i.e. Figs. 12h, 12i and 12g)	[0.47,0.79]	0.631	[0.46,0.80]	0.628	0.6433	0.6433
All (i.e. Figs. 15a, 15b and 14)	[0.36,0.80]	0.579	[0.31,0.83]	0.570	[0.6239,0.6239]	0.6239

TABLE VIII  
EXAMPLE 3 - CRISP INTERVALS AND UNCERTAIN INTERVALS

<i>Crisp Intervals</i>			
Candidate	1 <sup>st</sup> Survey	2 <sup>nd</sup> Survey	3 <sup>rd</sup> Survey
A	[0.40,0.80]	[0.30,0.60]	[0.35,0.70]
B	[0.40,0.85]	[0.50,0.80]	[0.45,0.95]
C	[0.20,0.80]	[0.30,0.85]	[0.25,0.75]
D	[0.25,0.75]	[0.35,0.75]	[0.30,0.70]
<i>Uncertain Intervals (5% uncertainty (=0.05)) (see Fig. 4)</i>			
Candidate	1 <sup>st</sup> Survey	2 <sup>nd</sup> Survey	3 <sup>rd</sup> Survey
A	[[0.35,0.45],[0.75,0.85]]	[[0.25,0.35],[0.55,0.65]]	[[0.30,0.40],[0.65,0.75]]
B	[[0.35,0.45],[0.80,0.90]]	[[0.45,0.55],[0.75,0.85]]	[[0.40,0.50],[0.90,1]]
C	[[0.15,0.25],[0.75,0.85]]	[[0.25,0.35],[0.80,0.90]]	[[0.20,0.30],[0.70,0.80]]
D	[[0.20,0.30],[0.70,0.80]]	[[0.30,0.40],[0.70,0.80]]	[[0.25,0.35],[0.65,0.75]]

In the present example, the uncertain intervals are converted into IT2 FSs as described in Section III-D2 during Stage 1, whereas in the previous examples the crisp intervals were converted into T1 FSs. The resulting sets capture the intra-participant uncertainty/variability across the three surveys and include the uncertainty information available on the uncertain endpoints of the original intervals. As discussed in Section III-D2, the individual IT2 FSs are created by performing the procedure seen in previous examples (for the T1 case and crisp intervals) twice, once for the UMF and once for the LMF. The LMF is computed using the interior endpoints of the intervals (i.e., the right hand side of the left intervals and the left hand side of the right intervals), and the UMF is created using the exterior endpoints of intervals (i.e., the left hand side of the left and the right hand side of the right intervals).

Equations (19) and (20) give examples (for Candidate A) on how (5) is used with the uncertain intervals from Table VIII to create the UMF ( $\bar{\mu}(A)$ ) and LMF ( $\underline{\mu}(A)$ ) which together completely describe the IT2 FS  $\tilde{A}$  for Candidate A. In the primary domain,  $y_1 \approx 0.33$ ,  $y_2 \approx 0.66$  and  $y_3 = 1$ . A more detailed view of the calculations can be found in Appendix A-A. Figure 17 shows the IT2 FSs created for each of the experts in this example over the three surveys; Table IX shows the numeric detail of the sets where the intervals for the UMF and LMF are associated with their respective primary membership grades (where a point is associated with more than one primary membership degree, the maximum degree applies). Also shown is the centroid for each set, which is provided for information and comparison purposes. Note that as the uncertainty has been injected into the source examples symmetrically, the centroids of the UMF and LMF are equal.

$$\bar{\mu}(\tilde{A}) = (y_1 / [0.25, 0.85] + y_2 / [0.30, 0.75] + y_3 / [0.35, 0.65]) \quad (19)$$

$$\underline{\mu}(\tilde{A}) = (y_1 / [0.35, 0.75] + y_2 / [0.40, 0.65] + y_3 / [0.45, 0.55]) \quad (20)$$

TABLE IX  
EXAMPLE 3 - T1 FS PARAMETERS WITH INTERVALS AND ASSOCIATED PRIMARY MEMBERSHIP

Candidate	$y = 0.33$	$y = 0.66$	$y = 1$	Centroid
<b>UMF</b>				
A	[0.25,0.85]	[0.30,0.75]	[0.35,0.65]	0.5167
B	[0.35,1]	[0.40,0.90]	[0.45,0.85]	0.6542
C	[0.15,0.90]	[0.20,0.85]	[0.25,0.80]	0.525
D	[0.20,0.80]	[0.25,0.80]	[0.30,0.75]	0.5208
<b>LMF</b>				
A	[0.35,0.75]	[0.40,0.65]	[0.45,0.55]	0.5167
B	[0.45,0.90]	[0.50,0.80]	[0.55,0.75]	0.6542
C	[0.25,0.80]	[0.30,0.75]	[0.35,0.70]	0.525
D	[0.30,0.70]	[0.35,0.70]	[0.40,0.65]	0.5208

After the participant-specific uncertain intervals have been captured using IT2 FSs in Stage 1, we proceed to Stage 2 where we combine the IT2 FSs to create a zGT2 FS representing the intra- and inter-expert uncertainty. As previously shown, the secondary membership domain is divided into four levels, one for each level of agreement between the four participants, at membership degrees of 0.25, 0.5, 0.75 and 1. We employ the IT2 sets generated in the first stage to compute each level of secondary membership (by computing the agreement over all IT2 FSs, see (7) and [18]). In our example, the zGT2 FS  $\tilde{Z}$  consists of four zSlices  $\tilde{Z}_1, \tilde{Z}_2, \tilde{Z}_3$  and  $\tilde{Z}_4$ . Equations (21) to (24) provide the detail of the zSlices which are calculated using (7). Equation (25) shows the combination of the individual zSlices, producing the actual zGT2 FS  $\tilde{Z}$ . The calculations can be found in Appendix A-B. Table X provides the numeric details of all of the zSlices, as well as the centroids of each zSlice and the overall defuzzified value of the zGT2 set. Notice that, unlike in previous examples, the LMF and UMF are not equal as



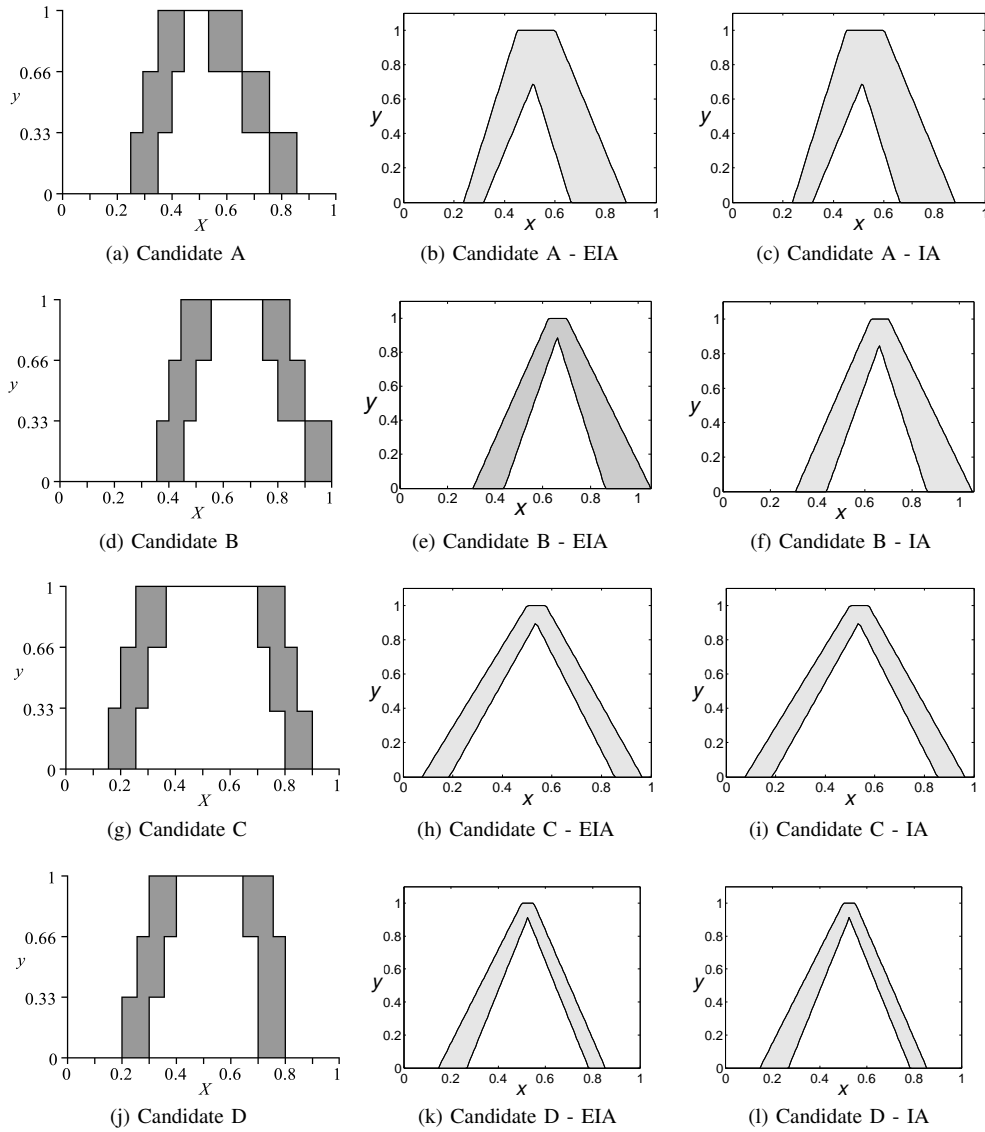


Fig. 17. Example 3 - IT2 FSs produced with IAA from uncertain intervals for each candidate (left column), and EIA and IA using corresponding crisp intervals for each candidate (middle and right column).

we are modelling the uncertainty about the endpoints of the uncertain intervals. Figure 18 shows 2D views of the zSlices at the respective secondary membership degrees (zLevels) of 0.25, 0.5, 0.75 and 1. Figures 19(a,b) show rear and front views of the complete zGT2 set  $\tilde{Z}$ , representing an aggregate of all of the experts' opinions over all surveys.

$$\begin{aligned}\tilde{Z}_1 &= 0.25 / (0.33 / [0.15, 1] + 0.66 / [0.20, 0.90] \\ &\quad + 1 / [0.25, 0.85]) \\ \tilde{Z}_1 &= 0.25 / (0.33 / [0.25, 0.90] + 0.66 / [0.30, 0.80] \\ &\quad + 1 / [0.35, 0.75])\end{aligned}\quad (21)$$

$$\begin{aligned}\tilde{Z}_2 &= 0.5 / (0.33 / [0.20, 0.90] + 0.66 / [0.25, 0.85] \\ &\quad + 1 / [0.30, 0.80]) \\ \tilde{Z}_2 &= 0.5 / (0.33 / [0.30, 0.80] + 0.66 / [0.35, 0.75] \\ &\quad + 1 / [0.40, 0.70])\end{aligned}\quad (22)$$

$$\begin{aligned}\tilde{Z}_3 &= 0.75 / (0.33 / [0.25, 0.85] + 0.66 / [0.30, 0.80] \\ &\quad + 1 / [0.35, 0.75]) \\ \tilde{Z}_3 &= 0.75 / (0.33 / [0.35, 0.75] + 0.66 / [0.40, 0.70] \\ &\quad + 1 / [0.45, 0.65])\end{aligned}\quad (23)$$

$$\begin{aligned}\tilde{Z}_4 &= 1 / (0.33 / [0.35, 0.80] + 0.66 / [0.40, 0.75] \\ &\quad + 1 / [0.45, 0.65]) \\ \tilde{Z}_4 &= 1 / (0.33 / [0.45, 0.70] + 0.66 / [0.50, 0.65] \\ &\quad + 1 / [0.55, 0.55])\end{aligned}\quad (24)$$

$$\tilde{Z} = \tilde{Z}_1 \cup \tilde{Z}_2 \cup \tilde{Z}_3 \cup \tilde{Z}_4 \quad (25)$$

It can be seen from the results that IAA allows the representation of multiple-expert, multiple-survey data with minimal loss or addition of information, i.e., the resulting representation uses all of the data available without, for example, employing a predefined type of membership function (e.g., triangular or Gaussian). Additionally, there is a clear separation between

TABLE X  
EXAMPLE 3 - zSLICE DETAILS WITH INTERVALS AND ASSOCIATED  
PRIMARY AND SECONDARY MEMBERSHIPS.

	$z_1 = 0.25$	$z_2 = 0.5$	$z_3 = 0.75$	$z_4 = 1$
<b>UMFs</b>	$\tilde{Z}_1$	$\tilde{Z}_2$	$\tilde{Z}_3$	$\tilde{Z}_4$
$y = 0.33$	[0.15,1]	[0.20,0.90]	[0.25,0.85]	[0.35,0.80]
$y = 0.66$	[0.20,0.90]	[0.25,0.85]	[0.30,0.80]	[0.40,0.75]
$y = 1$	[0.25,0.85]	[0.30,0.80]	[0.35,0.75]	[0.45,0.65]
<b>Centroid</b>	0.25/0.5542	0.5/0.55	0.75/0.55	1/0.5625
<b>LMFs</b>	$\tilde{Z}_1$	$\tilde{Z}_2$	$\tilde{Z}_3$	$\tilde{Z}_4$
$y = 0.33$	[0.25,0.90]	[0.30,0.80]	[0.35,0.75]	[0.45,0.70]
$y = 0.66$	[0.30,0.80]	[0.35,0.75]	[0.40,0.70]	[0.50,0.65]
$y = 1$	[0.35,0.75]	[0.40,0.70]	[0.45,0.65]	[0.55,0.55]
<b>Centroid</b>	0.25/0.5542	0.5/0.55	0.75/0.55	1/0.5625
<b><math>\tilde{Z}</math> defuzzified</b>	0.5554			

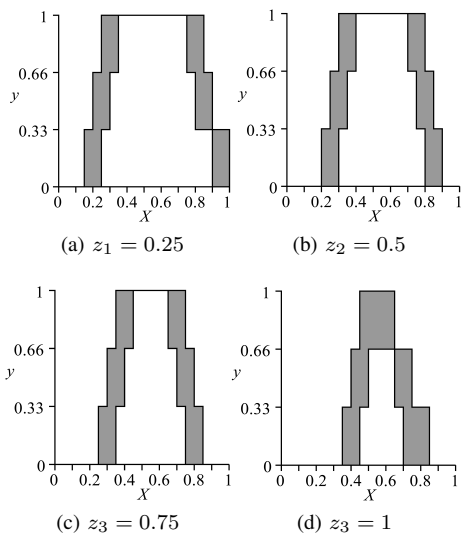


Fig. 18. Example 3 - 2D view of the zSlices produced with IAA.

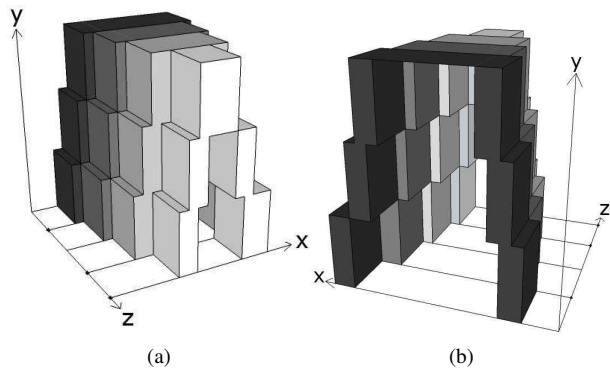


Fig. 19. Example 3 - GT2 FS  $\tilde{Z}$  produced using IAA for all candidates over all surveys, (a) is a rear view and (b) is a front view.

the representation of intra- and inter-participant uncertainty through the primary and secondary degrees of membership.

Finally, it is worthwhile repeating that the decision whether to model the intra-participant uncertainty in the primary membership domain and the inter-participant uncertainty in the secondary membership domain or vice-versa is application/design dependent. It is however essential that this choice is consciously made and taken into account at the time of the interpretation or use of the resulting sets.

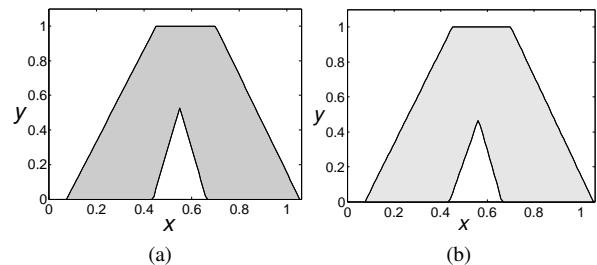


Fig. 20. Example 3 - IT2 FS produced for all candidates/surveys with (a) EIA and (b) IA.

1) *Comparison*: In order to compare IAA to IA/EIA, as in the previous examples we have generated FSs for each participant as well as an overall set capturing all participants. Figures 17 and 20 show the resulting sets for IA/EIA, and Tables XI and XII show the centroids and defuzzified centroids of IA/EIA FSs and those produced using the IAA respectively.

As in the previous examples, a direct comparison of the outputs of both approaches is not trivial, in particular based on the numeric results. However, comparing the candidate-specific sets (i.e., Fig. 17), the similarity in shape of the sets generated by all three approaches is apparent. While the shapes are similar, it can also be seen how the proposed IAA approach models the intra-participant uncertainty (across three survey iterations) in the primary membership, modelling the uncertainty about the interval endpoints in the FOU. The sets generated by the IA/EIA approaches model the intra-user uncertainty across the three surveys using the FOU. The uncertainty about the interval endpoints has not been captured as the IA/EIA approaches currently do not support the modelling of uncertain intervals. Finally, a comparison of the overall sets (i.e., Figs. 19 and 20(b)) shows how IAA models intra-participant uncertainty in the primary membership and how the uncertainty about the interval endpoints is captured in the FOU. The secondary membership models the inter-user uncertainty across all four participants. The FSs generated by the IA/EIA approach combine both intra- and inter-participant uncertainties in the triangular IT2 FS model.

## V. REAL-WORLD APPLICATION

While the creation of models from data such as those produced by the IAA may provide direct utility in terms of the interpretability and conciseness of the models, it is their utility in applications such as decision support and automatic decision making systems that provides crucial value. As an illustration of such applications for the models arising from the IAA, we showcase their application in conjunction with similarity measures as initially reported in [4]. Specifically, we derive linguistic descriptions of public houses using IAA-generated models of concepts (e.g., *ambience* or *service*) for each public house, and models of words (i.e., adjectives like *excellent* or *neutral*) in a specific context, where the context arises from the concepts mentioned (e.g., *excellent Ambience* may be different from *excellent Service*). The concept models are then compared with the word models using Jaccard's similarity measure for type-1 and type-2 fuzzy sets (see [23])

TABLE XI  
EXAMPLE 3 - CENTROIDS OF FSS (SEE FIGS. 17 AND 20) PRODUCED WITH THE EIA [1] AND IA [5].

Expert	EIA					IA				
	A	B	C	D	All	A	B	C	D	All
Centroid	[0.45,0.60]	[0.60,0.73]	[0.48,0.58]	[0.48,0.56]	[0.37,0.76]	[0.45,0.60]	[0.60,0.73]	[0.48,0.58]	[0.48,0.56]	[0.36,0.77]
Defuzz.	0.526	0.664	0.526	0.517	0.562	0.526	0.664	0.526	0.517	0.563

TABLE XII  
EXAMPLE 3 - DEFUZZIFIED CENTROIDS OF FSS PRODUCED WITH THE IAA.

Expert	A	B	C	D	All
Centroid	[0.482,0.582]	[0.611,0.711]	[0.474,0.575]	[0.465,0.565]	[0.509,0.611]
Defuzz.	0.532	0.661	0.525	0.515	0.560

to gauge the level to which each venue can be described using each concept-describing word.

The example shown in this section has been produced using Juzzy, an object-oriented toolkit for the development of T1, IT2 and GT2 fuzzy systems [24], which is freely available online at <http://juzzy.wagnerweb.net/>.

#### A. Comparing Concepts and Words

In this example two sets of models are created using the IAA:

- 1) Concept models of six concepts (*Ambience, Service, Variety of Beer, Quality of Beer, Cost of Beer and Overall*) for a series of 13 venues (public houses) in the centre of Nottingham, UK.
- 2) Word models describing a set of three adjectives (e.g., *low, medium, high*) for each of the six concepts in the context of public houses in general.

Both models are created with the IAA based on data collected during a single iteration, small-scale survey exercise with 18 local residents that employed two separate questionnaires; one for the rating of the words and one for the rating of the concepts. Note that only a single survey iteration was conducted, i.e., no participant was asked more than once (no intra-participant variation information is available). The survey uses the ellipse method of eliciting opinions as described in Section I and participants were issued with the two questionnaires (in random order). The overall aim is to provide linguistic descriptions of the public houses without relying on any expert-designed FSSs, meaning all FSSs are generated directly from data by the IAA: both the sets describing each public house, and the sets that describe the actual *low, medium, high* valuations of all six concepts in the context of public houses. For space considerations, we have not included all details of this process. For full details, please see [4].

Once the data was collected, the IAA was used to create T1 FSSs representing each word and each concept. In this example, the primary degree of membership represents the inter-user variation, that is, it shows the agreement between users over where the word lies on the scale. In the examples described in Section IV we modelled intra-user variability using the primary domain, and inter-user variation in the secondary domain. We could continue this approach and model each individual as an interval T1 FS, and combine them to produce a GT2 FS. However, we feel that in this case, where we have

only one set of data for each individual, it is more intuitive and practical to use the primary domain as this produces a more accessible result.

Figure 21a provides an example of the T1 FS models of the *Ambience* concept for the public houses *Trip to Jerusalem* and *Joseph Else*. The figure shows that the concept has been rated quite differently in the context of the individual public houses with the rating for the *Trip to Jerusalem* being much superior to the rating for the *Joseph Else*. From a technical perspective, we can note that the FS models resulting from the IAA (as noted in [3]) are non-convex and non-normal. The non-convexity is a direct result of the IAA which does not remove outliers (by design). The non-normality is a result of the agreement operation (see [3]), which results in wider and less “high” FSSs for lower levels of agreement (overlap) of constituting intervals and conversely, in higher FSSs with a narrower support for strong overlap. Thus, the resulting FSSs directly reflect the agreement and discord on the given concept in the given context across participants. While it may be tempting at first glance to normalise the sets, the fact that there is no perfect agreement across all participants means that the model for the word/concept should *not* be completely true (i.e., 1) anywhere.

Figure 21b provides an example of the T1 FS word models for *Negative, Neutral* and *Positive* in the context of *public house ambience*. As with the concept models, the resulting FSSs clearly indicate agreement and discord amongst participants.

It is interesting to note that the data captured for *Neutral* contains three contributions which would generally be referred to as outliers (as they do not overlap with any other contributions). Namely, one participant rated neutral as [10, 26], while another rated it as [77, 100]. Furthermore, one participant rated *Positive* as [34, 69]. Fig. 21b shows clearly how these inputs are preserved by the IAA. If one opts for pre-processing and outlier removal, the resulting model (without said participants) would change as shown in Fig. 21c. Note that even if outliers are removed, the general shape of the model stays intact, i.e., the IAA captures *outliers* but produces robust models that capture their input without being drastically altered. The main change in the model is the increased level of agreement (e.g., now 1.0 for *Neutral* as the discord in relation to the outliers has been removed. Employing the centroid as a basic numeric summarisation of the sets, the outlier free models of *Neutral* and *Positive* have centroids of 51.435 and 81.696 respectively, while their original counterparts (with outliers) have centroids

of 52.143 and 76.271. The centroids indicate that the *outliers* are captured and influence the models. However, they do not drastically alter them. Thus, we believe that it is up to the experimenter to make the decision if outlier-removal is warranted and to apply it where necessary/appropriate.

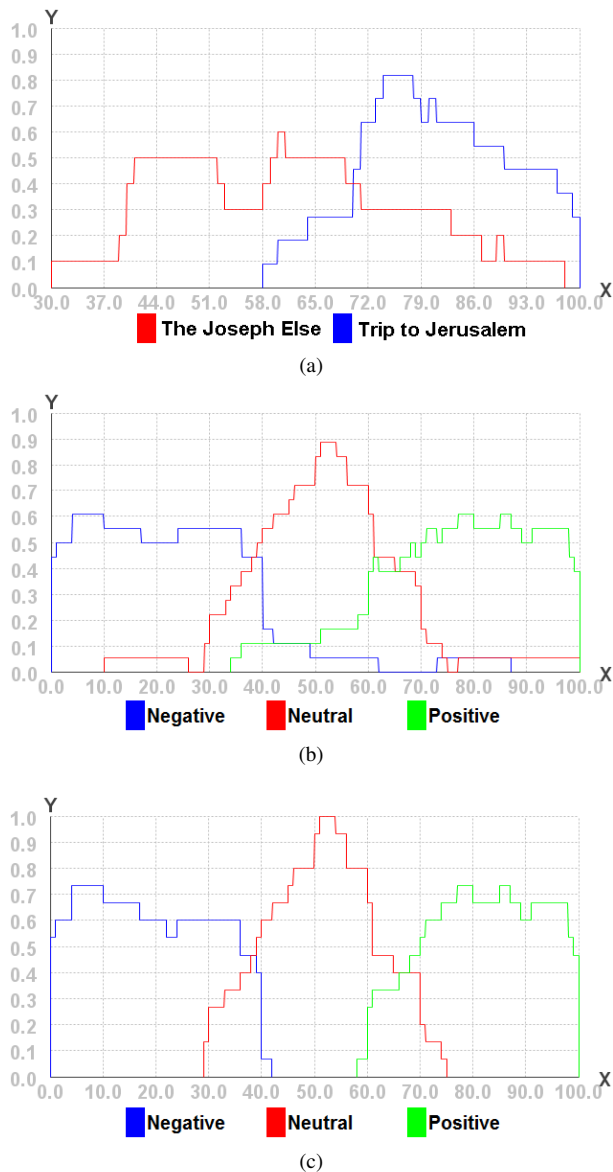


Fig. 21. T1 FSs produced with the IAA for (a) the concept of *Ambience* in *Trip to Jerusalem* and *Joseph Else*, (b) the word models associated with the concept *Ambience* and, (c) the same as in (b) but with outliers removed.

In the context of the application, it is noteworthy that the word models can easily change. For example, with a more demanding group different intervals may be elicited. Also, different words may provide (more) meaningful results. For example, one could build models for the concept of *music volume* with individual word models, e.g., *low*, *ok*, *loud*, *unbearable*. Again, different groups are likely to provide very different ratings for the words resulting in different user-centric mappings (similarities) between concepts (pubs/venues) and the given words. We provide an interpretation of the similarities computed from the FS models created for this example in subsection V-B below.

1) *Concept Models*: During the same survey exercise in which the data for the word models was collected, participants completed a second survey concerning a series of 13 pubs. For each pub (e.g., *Rose and Crown*), participants were asked a series of questions (e.g., ‘*How would you rate the service?*’) that required them to rate one of six concepts using the ellipse method. The resulting data was used to create a set of six (one for each concept) T1 FSs for each pub. Again, in this example the primary domain is used to represent the level of inter-user agreement.

## B. Relating Concepts to Words - Results

Once a set of word models and a set of concept models has been produced, the next task is to compare the concept models for each *specific* public house with each of the *generic* (in the context of public houses) word models for each concept. To recapitulate, the aim is to compare the ratings of the concepts (e.g., *service in pub x*) to the *x*-independent word models of the same concept (e.g., *poor*, *acceptable* or *excellent* service). This gives us an insight into which word model most closely matches the model of the pub-specific concept, in other words, is public house *x*'s *service* most similar to *poor*, *acceptable* or *excellent* service in the context of *service in pubs*? This similarity in turn can be used to generate, for example, a linguistic description of the public house which is specific to the individual/group of individuals surveyed. To compute similarity we use the Jaccard similarity measure as described in [4].

Figure 22 shows three example profiles detailing the similarity between the T1 fuzzy models of each concept for each venue and the T1 fuzzy word models representing the positive, neutral and negative labels associated with each question. In Fig. 22 we see that the concept models produced for *Trip to Jerusalem* have a high degree of similarity to the word models that represent positive labels for most questions. A more evenly balanced response is shown for the *Ropewalk* and the *Joseph Else*, where the concept models mostly match the neutral word models associated with each question.

Table XIII provides linguistic descriptions of each concept for each venue in Fig. 22. The words are selected by taking the word model (i.e., FS) that has the maximum similarity to the concept model for each venue. The linguistic descriptions are clearly representative of the profiles previously shown (Fig. 22). However, in some cases, taking the maximum may not provide as much information as one would like. For example, the concept *service* for *Trip to Jerusalem* produces close similarity values when compared with the word models *acceptable* and *excellent*. By taking the maximum (*excellent*) we discard the information that it has a (relatively) high similarity with *acceptable*. To address this, the method could be extended to contain qualifiers for each term (e.g., *quite* or *very*). Defining such qualifiers could potentially involve a similar methodology to that used to produce the word models.

The column charts in Fig. 23 illustrate two examples of the similarity between the word models associated with an individual concept (e.g., *positive*, *neutral*, *negative*) and the concept models (e.g., *ambience* of the pubs) generated from

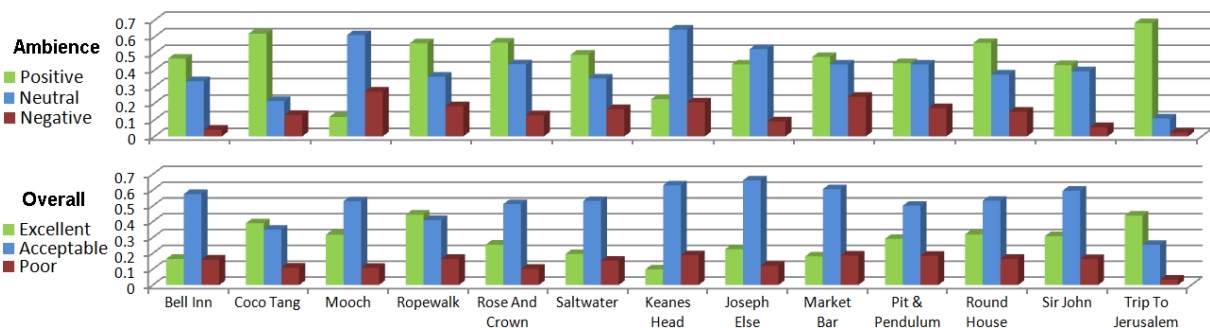


Fig. 23. Similarity profiles for *ambience* and *overall* rating of pubs.

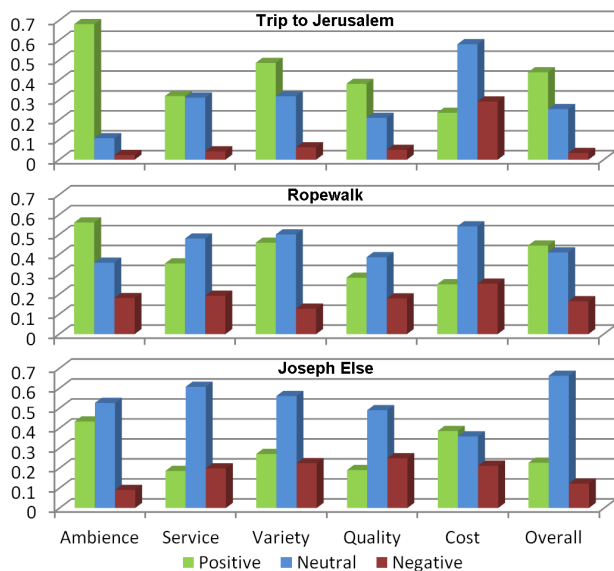


Fig. 22. Similarity profiles for Trip to Jerusalem, Ropewalk and Joseph Else.

TABLE XIII  
LINGUISTIC DESCRIPTIONS OF CONCEPTS FOR EACH PUB

Concept	Trip to Jerusalem	Ropewalk	Joseph Else
<b>Ambience</b>	Positive	Positive	Neutral
<b>Service</b>	Excellent	Acceptable	Acceptable
<b>Variety of Beer</b>	Wide	Reasonable	Reasonable
<b>Quality of Beer</b>	Delicious	Acceptable	Acceptable
<b>Cost of Beer</b>	Reasonable	Reasonable	Cheap
<b>Overall</b>	Excellent	Excellent	Acceptable

the survey. The Ambience chart shows that for most venues the concept model produced for the *ambience* question is most similar to the fuzzy word model for *positive* ambience. The results are slightly more varied for the *overall* concept which shows a stronger bias toward the *acceptable* label. Figure 24 contains three column charts, each showing the total similarity of the concept models for each venue to *groupings* of the *positive*, *neutral* and *negative* word models associated with each concept. As all concepts were rated with three words, we grouped the words describing *positive* ratings, the words describing *neutral* ratings and the words describing *negative* ratings in the top, middle and bottom rows respectively. In each chart the columns are divided, showing the contribution each concept has to the overall total similarity for each pub/word group. It is immediately obvious that very few of the venues had greatest similarity with the *negative* word

group models associated with each concept. For most venues the strongest similarities are with the *neutral* labels. Overall, this representation quickly summarises the desirability of the different public houses (e.g., height of column in positive section) and visualises which aspects (*ambience*, *service*, etc.) have contributed to this height/rating and how much.

All the results in this section are based directly on data-driven FS models generated by the IAA. While the models shown are still relatively simple, in the future we are looking to demonstrate the IAA in more complex scenarios with richer datasets. For now, we proceed to a discussion on the role of the IAA, in particular in the context of the existing techniques such as the IA and EIA.

## VI. DISCUSSION

The IA/EIA methods were developed to take series of interval definitions (from multiple sources) of linguistic terms (generally adjectives) and to create IT2 FS based models of an overall consensus opinion. In IA/EIA, data statistics of the source intervals are mapped to predefined FS types (e.g., triangular FSs) in order to generate the output FS model. The IAA method follows a similar aim of capturing an overall model of multiple intervals defining a specific linguistic term. However, it differs from the IA/EIA approaches in the use of GT2 rather than IT2 FSs and in the basic approach to generating the FS, i.e., it does not employ predefined FS types, aiming to avoid assumptions. Fundamentally, the IAA approach is designed to follow the principle of least commitment [21], [22], i.e. it avoids making any assumptions about what the “true” distribution of the uncertainty over the given intervals may look like. Further, IAA differentiates between different types of uncertainty—specifically intra- and inter-source uncertainty—which are mapped to the primary and secondary membership levels of the zGT2 FSs. We feel that allowing differentiation between the different types of uncertainty is an essential feature of the proposed approach. At a fundamental level, it is worth noting the difference in the resulting set model types generated by IAA and IA/EIA for different types/numbers of input intervals:

- when modelling crisp intervals from a single source (e.g., a single expert’s opinion) over one or more surveys, IAA produces a T1 FS capturing the intra-source uncertainty, whereas IA produces an IT2 FS;

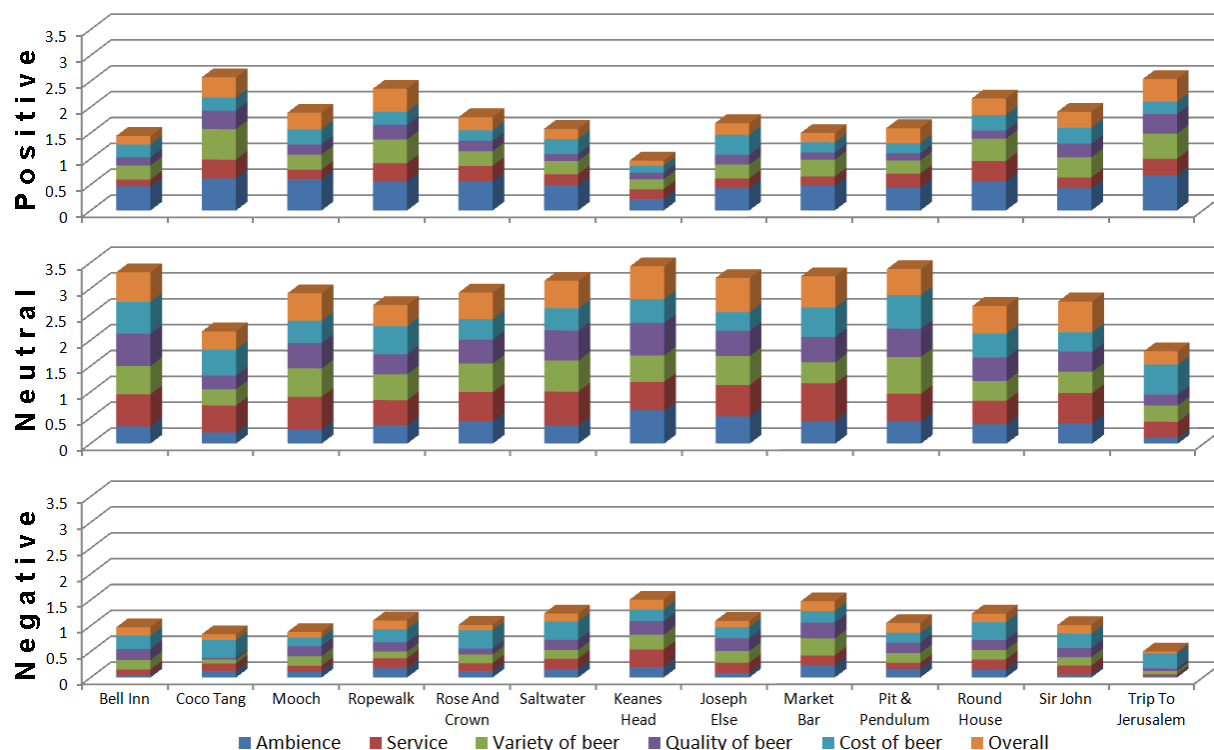


Fig. 24. Total similarity of each public house with *Positive*, *Neutral* and *Negative* word groups.

- when modelling crisp intervals from multiple sources over one or more surveys, IAA produces a zGT2 FS capturing both intra- and inter-source uncertainty in the primary and secondary membership levels of the set respectively, whereas IA/EIA produce an IT2 FS combining both types of uncertainty;
- the IAA approach enables the capturing and modelling of uncertain intervals which is currently not directly possible with the IA/EIA approaches.

While we fully recognise that pre-processing of data is often desirable, IAA neither *includes* nor *requires* any pre-processing stage. We view pre-processing (informally) as an orthogonal process to model creation. Of course, pre-processing and model creation are inextricably linked in that the choice of the pre-processing method may depend on the model creation method (and vice versa) and the effects of pre-processing will affect the resultant model. However, by de-coupling the two processes, we leave the choice of the pre-processing method—including none, one or a combination of several methods—to the model creator. In our experience, the selection of a pre-processing method is heavily context (and data) dependent. For example, if one is modelling expert opinion, then an ‘outlier’ (an observation far from the remainder of the sample) may be a perfectly valid difference of opinion, such as arising from new or ground-breaking knowledge, just as much as it may be an ‘error’ to be ignored. By separating pre-processing from model creation, pre-processing may be carried out before or alongside IAA, as deemed appropriate by the model creator.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced IAA, a novel approach to modelling interval-based data using non-parametric FS models. We detailed how IAA can capture intra- and inter-source uncertainty, which allows the capturing of information gathered over one or multiple surveys from one or more survey participants. Moreover, we showed how IAA enables the capturing of both crisp and uncertain intervals (where the endpoints themselves are uncertain) and how the resulting FSs capture uncertainty using the primary and secondary degrees of membership of zGT2 FSs, separating the modelling of intra- and inter-source uncertainty. We have highlighted the motivations for the creation of IAA, in particular the perceived need to design a method which generates interpretable FS models from interval-based data while minimising any loss of information and any assumptions (such as the application of predefined FS types such as triangular or Gaussian FSs). In particular, IAA makes minimal assumptions about the distributions within the collected interval-based data and further, it does not *rely on* data pre-processing or outlier removal. We feel the latter is important, as in our experience, outliers may still be correct or contribute essential information. The IAA method does not exclude information contributed by outliers (unless we *know* it is bad data, in which case we would remove it) but includes it in the final model with a low “weight,” whereas areas of high agreement (among sources) are weighted more strongly.

The paper includes a series of numerical examples based on both real-world as well as synthetic data which provide complete detail of the creation of FS models from raw data. The numerical examples include a comparison of the resulting models to those generated through the Interval Approach (IA)



[5] introduced by Liu and Mendel and the Enhanced Interval Approach (EIA) [1] by Coupland et al., which also enable the capturing of interval-based data to generate FSs but rely on data pre-processing, use predefined types of FSs and do not support the processing of uncertain intervals.

While the scope of the current paper does not allow for the inclusion of real-world application examples beyond the numeric examples provided, we refer the interested reader to [4] where we describe a selection of applications based on IAA and similarity measures applied to the resulting type-2 FS models.

We feel IAA contributes a highly useful method for capturing interval-based (survey) data and associated uncertainty information in FS models. It minimises any assumptions, any loss of information and supports crisp or uncertain intervals originating from one or multiple sources captured over one or many survey iterations.

In the future, we aim to both drive the theoretical development of IAA as well as the exploration of practical applications with a particular focus on conducting a web- or mobile-app based data collection exercise, which will enable us to access more representative data and better evaluate the proposed approach in real-world contexts.

#### ACKNOWLEDGMENT

We gratefully acknowledge the role of CESG in supporting this work. We would also like to thank all those from IMA who took part in the survey used in this paper.

#### REFERENCES

- [1] S. Coupland, J. Mendel, and D. Wu, "Enhanced Interval Approach for Encoding Words into Interval Type-2 Fuzzy Sets and Convergence of the Word FOU's," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–8.
- [2] D. Wu, J. Mendel, and S. Coupland, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis," *Fuzzy Systems, IEEE Transactions on*, vol. 20, no. 3, pp. 499–513, 2012.
- [3] S. Miller, C. Wagner, and J. Garibaldi, "Constructing general type-2 fuzzy sets from interval-valued data," in *Proceedings of the 2012 IEEE International Conference on Fuzzy Systems*, Brisbane, Australia., June 2012.
- [4] C. Wagner, S. Miller, and J. Garibaldi, "Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets," in *Fuzzy Systems (FUZZ-IEEE), 2013 IEEE International Conference on*, 2013, pp. 1–8.
- [5] F. Liu and J. Mendel, "Encoding Words Into Interval Type-2 Fuzzy Sets Using An Interval Approach," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 1503–1521, 2008.
- [6] F. Herrera, S. Alonso, F. Chiclana, and E. Herrera-Viedma, "Computing with words in decision making: foundations, trends and prospects," *Fuzzy Optimization and Decision Making*, vol. 8, pp. 337–364, 2009, 10.1007/s10700-009-9065-2. [Online]. Available: <http://dx.doi.org/10.1007/s10700-009-9065-2>
- [7] F. Liu and J. Mendel, "An interval approach to fuzzistics for interval type-2 fuzzy sets," in *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, July 2007, pp. 1–6.
- [8] C. Wagner and H. Hagrais, "Toward general type-2 fuzzy logic systems based on zSlices," *Fuzzy Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 637–660, 2010.
- [9] J. Mendel and R. John, "Type-2 fuzzy sets made simple," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, April 2002.
- [10] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I\* 1," *Information sciences*, vol. 8, no. 3, pp. 199–249, 1975.
- [11] S. Coupland and R. I. John, "Geometric type-1 and type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 3–15, February 2007.
- [12] J. Mendel, F. Liu, and D. Zhai, " $\alpha$ -plane representation for type-2 fuzzy sets: Theory and applications," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 5, pp. 1189–1207, oct. 2009.
- [13] F. Liu, "An efficient centroid type-reduction strategy for general type-2 fuzzy logic system," *Information Sciences*, vol. 178, pp. 2224–2236, 2008.
- [14] J. Mendel, R. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 6, pp. 808–821, December 2006.
- [15] J. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems - Introduction and New Directions*. Upper Saddle River, NJ 07458. USA.: Prentice Hall, 2001.
- [16] L. Zadeh, "Fuzzy logic = computing with words," *Fuzzy Systems, IEEE Transactions on*, vol. 4, no. 2, pp. 103–111, may 1996.
- [17] J. Mendel, "Computing with words and its relationships with fuzzistics," *Information Sciences*, vol. 177, pp. 988–1006, 2007.
- [18] C. Wagner and H. Hagrais, "Employing zSlices based general type-2 fuzzy sets to model multi level agreement," in *Advances in Type-2 Fuzzy Logic Systems (T2FUZZ), 2011 IEEE Symposium on*. IEEE, 2011, pp. 50–57.
- [19] T. C. Havens, D. T. Anderson, C. Wagner, H. Deilamsalehy, and D. Wonnacott, "Fuzzy integrals of crowd-sourced intervals using a measure of generalized accord," in *Fuzzy Systems (FUZZ-IEEE), 2013 IEEE International Conference on*, 2013, pp. 1–8.
- [20] C. Wagner and H. Hagrais, "zSlices - towards bridging the gap between interval and general type-2 fuzzy logic," in *Proceedings of IEEE International Fuzzy Systems Conference (FUZZ-IEEE 2008). IEEE World Congress on Computational Intelligence (WCCI 2008)*. Hong Kong: IEEE, 1-6 June 2008, pp. 489–497.
- [21] M. J. Stefik, "Planning with constraints," Ph.D. dissertation, Stanford, CA, USA, 1980, aAI8016868.
- [22] J. Allen, J. Hendler, and A. Tate, *Readings in planning*, ser. The Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann Publishers, 1990. [Online]. Available: [http://books.google.co.uk/books?id=OI\\_YDs\\_DIgC](http://books.google.co.uk/books?id=OI_YDs_DIgC)
- [23] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity measures of interval type-2 fuzzy sets to general type-2 fuzzy sets," in *Submitted to the 2013 IEEE International Conference on Fuzzy Systems*, Hyderabad, India., July 2013.
- [24] C. Wagner, "Juzzy - a java based toolkit for type-2 fuzzy logic," in *In Proceedings of the 2013 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, 16 - 19 April 2013, p. Accepted for Publication.

## APPENDIX A

## EXAMPLE 3: DETAILED COMPUTATION

This appendix contains a detailed view of the calculations required for Example 3 (Section IV-C), where  $y_1 \approx 0.33$ ,  $y_2 \approx 0.66$  and  $y_3 = 1$ .

## A. Step 1: Uncertain Intervals to Interval Type-2 FSs

$$\begin{aligned} \bar{\mu}(\tilde{A}) &= (y_1 / ([0.35, 0.85] \cup [0.25, 0.65] \cup [0.30, 0.75]) \\ &\quad + y_2 / (([0.35, 0.85] \cap [0.25, 0.65]) \cup ([0.35, 0.85] \\ &\quad \cap [0.30, 0.75]) \cup ([0.25, 0.65] \cap [0.30, 0.75])) \\ &\quad + y_3 / ([0.35, 0.85] \cap [0.25, 0.65] \cap [0.30, 0.75])) \\ &= (0.33 / [0.25, 0.85] + 0.5 / [0.30, 0.75] \\ &\quad + 1 / [0.35, 0.65]) \end{aligned} \quad (26)$$

$$\begin{aligned} \underline{\mu}(\tilde{A}) &= (y_1 / ([0.45, 0.75] \cup [0.35, 0.55] \cup [0.40, 0.65]) \\ &\quad + y_2 / (([0.45, 0.75] \cap [0.35, 0.55]) \cup ([0.45, 0.75] \\ &\quad \cap [0.40, 0.65]) \cup ([0.35, 0.55] \cap [0.40, 0.65])) \\ &\quad + y_3 / ([0.45, 0.75] \cap [0.35, 0.55] \cap [0.40, 0.65])) \\ &= (0.33 / [0.35, 0.75] + 0.5 / [0.40, 0.65] \\ &\quad + 1 / [0.45, 0.55]) \end{aligned} \quad (27)$$

## B. Step 2: Interval Type-2 FSs to General Type-2 FSs

$$\begin{aligned} \tilde{Z}_1 &= 0.25 / (y_1 / ([0.25, 0.85] \cup [0.35, 1] \cup [0.15, 0.90] \\ &\quad \cup [0.20, 0.80]) \\ &\quad + y_2 / ([0.30, 0.75] \cup [0.40, 0.90] \cup [0.20, 0.85] \\ &\quad \cup [0.25, 0.80]) \\ &\quad + y_3 / ([0.35, 0.65] \cup [0.45, 0.85] \cup [0.25, 0.80] \\ &\quad \cup [0.30, 0.75])) \\ &= 0.25 / (0.33 / [0.15, 1] + 0.66 / [0.20, 0.90] \\ &\quad + 1 / [0.25, 0.85]) \end{aligned}$$

$$\begin{aligned} \tilde{Z}_1 &= 0.25 / (y_1 / ([0.35, 0.75] \cup [0.45, 0.90] \cup [0.25, 0.80] \\ &\quad \cup [0.30, 0.70]) \\ &\quad + y_2 / ([0.40, 0.65] \cup [0.50, 0.80] \cup [0.30, 0.75] \\ &\quad \cup [0.35, 0.70]) \\ &\quad + y_3 / ([0.45, 0.55] \cup [0.55, 0.75] \cup [0.35, 0.70] \\ &\quad \cup [0.40, 0.65])) \\ &= 0.25 / (0.33 / [0.25, 0.90] + 0.66 / [0.30, 0.80] \\ &\quad + 1 / [0.35, 0.75]) \end{aligned} \quad (28)$$

$$\begin{aligned} \tilde{Z}_2 &= 0.5 / (y_1 / (([0.25, 0.85] \cap [0.35, 1]) \cup ([0.25, 0.85] \\ &\quad \cap [0.15, 0.90]) \cup ([0.25, 0.85] \cap [0.20, 0.80]) \cup ([0.35, 1] \\ &\quad \cap [0.15, 0.90]) \cup ([0.35, 1] \cap [0.20, 0.80]) \cup ([0.15, 0.90] \\ &\quad \cap [0.20, 0.80])) \\ &\quad + y_2 / (([0.30, 0.75] \cap [0.40, 0.90]) \cup ([0.30, 0.75] \\ &\quad \cap [0.20, 0.85]) \cup ([0.30, 0.75] \cap [0.25, 0.80]) \cup ([0.40, 0.90] \\ &\quad \cap [0.20, 0.85]) \cup ([0.40, 0.90] \cap [0.25, 0.80]) \cup ([0.20, 0.85] \\ &\quad \cap [0.25, 0.80])) \\ &\quad + y_3 / (([0.35, 0.65] \cap [0.45, 0.85]) \cup ([0.35, 0.65] \\ &\quad \cap [0.25, 0.80]) \cup ([0.35, 0.65] \cap [0.30, 0.75]) \cup ([0.45, 0.85] \\ &\quad \cap [0.25, 0.80]) \cup ([0.45, 0.85] \cap [0.30, 0.75]) \cup ([0.25, 0.80] \\ &\quad \cap [0.30, 0.75])) \\ &= 0.5 / (0.33 / [0.20, 0.90] + 0.66 / [0.25, 0.85] \\ &\quad + 1 / [0.30, 0.80]) \\ \tilde{Z}_2 &= 0.5 / (y_1 / (([0.35, 0.75] \cap [0.45, 0.90]) \cup ([0.35, 0.75] \\ &\quad \cap [0.25, 0.80]) \cup ([0.35, 0.75] \cap [0.30, 0.70]) \cup ([0.45, 0.90] \\ &\quad \cap [0.25, 0.80]) \cup ([0.45, 0.90] \cap [0.30, 0.70]) \cup ([0.25, 0.80] \\ &\quad + y_2 / (([0.40, 0.65] \cap [0.50, 0.80]) \cup ([0.40, 0.65] \\ &\quad \cap [0.30, 0.75]) \cup ([0.40, 0.65] \cap [0.35, 0.70]) \cup ([0.50, 0.80] \\ &\quad \cap [0.30, 0.75]) \cup ([0.50, 0.80] \cap [0.35, 0.70]) \cup ([0.30, 0.75] \\ &\quad \cap [0.35, 0.70])) \\ &\quad + y_3 / (([0.45, 0.55] \cap [0.55, 0.75]) \cup ([0.45, 0.55] \\ &\quad \cap [0.35, 0.70]) \cup ([0.45, 0.55] \cap [0.40, 0.65]) \cup ([0.55, 0.75] \\ &\quad \cap [0.35, 0.70]) \cup ([0.55, 0.75] \cap [0.40, 0.65]) \cup ([0.35, 0.70] \\ &\quad \cap [0.40, 0.65])) \\ &= 0.5 / (0.33 / [0.30, 0.80] + 0.66 / [0.35, 0.75] \\ &\quad + 1 / [0.40, 0.70]) \end{aligned} \quad (29)$$

$$\begin{aligned}
\tilde{Z}_3 &= 0.75 / (y1 / (([0.25, 0.85] \cap [0.35, 1] \cap [0.15, 0.90]) \\
&\cup ([0.25, 0.85] \cap [0.35, 1] \cap [0.20, 0.80]) \\
&\cup ([0.25, 0.85] \cap [0.15, 0.90] \cap [0.20, 0.80]) \\
&\cup ([0.35, 1] \cap [0.15, 0.90] \cap [0.20, 0.80])) \\
&+ y2 / (([0.30, 0.75] \cap [0.40, 0.90] \cap [0.20, 0.85]) \\
&\cup ([0.30, 0.75] \cap [0.40, 0.90] \cap [0.25, 0.80]) \\
&\cup ([0.30, 0.75] \cap [0.20, 0.85] \cap [0.25, 0.80]) \\
&\cup ([0.40, 0.90] \cap [0.20, 0.85] \cap [0.25, 0.80])) \\
&+ y3 / (([0.35, 0.65] \cap [0.45, 0.85] \cap [0.25, 0.80]) \\
&\cup ([0.35, 0.65] \cap [0.45, 0.85] \cap [0.30, 0.75]) \\
&\cup ([0.35, 0.65] \cap [0.25, 0.80] \cap [0.30, 0.75]) \\
&\cup ([0.45, 0.85] \cap [0.25, 0.80] \cap [0.30, 0.75])) \\
&= 0.75 / (0.33 / [0.25, 0.85] + 0.25 / [0.30, 0.80] \\
&+ 1 / [0.35, 0.75]) \\
\tilde{Z}_3 &= 0.75 / (y1 / (([0.35, 0.75] \cap [0.45, 0.90] \cap [0.25, 0.80]) \\
&\cup ([0.35, 0.75] \cap [0.45, 0.90] \cap [0.30, 0.70]) \\
&\cup ([0.35, 0.75] \cap [0.25, 0.80] \cap [0.30, 0.70]) \\
&\cup ([0.45, 0.90] \cap [0.25, 0.80] \cap [0.30, 0.70])) \\
&+ y2 / (([0.40, 0.65] \cap [0.50, 0.80] \cap [0.30, 0.75]) \\
&\cup ([0.40, 0.65] \cap [0.50, 0.80] \cap [0.35, 0.70]) \\
&\cup ([0.40, 0.65] \cap [0.30, 0.75] \cap [0.35, 0.70]) \\
&\cup ([0.50, 0.80] \cap [0.30, 0.75] \cap [0.35, 0.70])) \\
&+ y3 / (([0.45, 0.55] \cap [0.55, 0.75] \cap [0.35, 0.70]) \\
&\cup ([0.45, 0.55] \cap [0.55, 0.75] \cap [0.40, 0.65]) \\
&\cup ([0.45, 0.55] \cap [0.35, 0.70] \cap [0.40, 0.65]) \\
&\cup ([0.55, 0.75] \cap [0.35, 0.70] \cap [0.40, 0.65])) \\
&= 0.75 / (0.33 / [0.35, 0.75] + 0.66 / [0.40, 0.70] \\
&+ 1 / [0.45, 0.65]) \\
\tilde{Z}_4 &= 1 / (y1 / ([0.25, 0.85] \cap [0.35, 1] \cap [0.15, 0.90] \\
&\cap [0.20, 0.80]) \\
&y2 / ([0.30, 0.75] \cap [0.40, 0.90] \cap [0.20, 0.85] \cap [0.25, 0.80]) \\
&y3 / ([0.35, 0.65] \cap [0.45, 0.85] \cap [0.25, 0.80] \cap [0.30, 0.75])) \\
&= 1 / (0.33 / [0.35, 0.80] + 0.66 / [0.40, 0.75] \\
&+ 1 / [0.45, 0.65]) \\
\tilde{Z}_4 &= 1 / (y1 / ([0.35, 0.75] \cap [0.45, 0.90] \cap [0.25, 0.80] \\
&\cap [0.30, 0.70]) \\
&y2 / ([0.40, 0.65] \cap [0.50, 0.80] \cap [0.30, 0.75] \cap [0.35, 0.70]) \\
&y3 / ([0.45, 0.55] \cap [0.55, 0.75] \cap [0.35, 0.70] \cap [0.40, 0.65])) \\
&= 1 / (0.33 / [0.45, 0.70] + 0.66 / [0.50, 0.65] \\
&+ 1 / [0.55, 0.55]) \\
\tilde{Z} &= \tilde{Z}_1 \cup \tilde{Z}_2 \cup \tilde{Z}_3 \cup \tilde{Z}_4
\end{aligned} \tag{30}$$

$$\begin{aligned}
\tilde{Z}_4 &= 1 / (y1 / ([0.25, 0.85] \cap [0.35, 1] \cap [0.15, 0.90] \\
&\cap [0.20, 0.80]) \\
&y2 / ([0.30, 0.75] \cap [0.40, 0.90] \cap [0.20, 0.85] \cap [0.25, 0.80]) \\
&y3 / ([0.35, 0.65] \cap [0.45, 0.85] \cap [0.25, 0.80] \cap [0.30, 0.75])) \\
&= 1 / (0.33 / [0.35, 0.80] + 0.66 / [0.40, 0.75] \\
&+ 1 / [0.45, 0.65]) \\
\tilde{Z}_4 &= 1 / (y1 / ([0.35, 0.75] \cap [0.45, 0.90] \cap [0.25, 0.80] \\
&\cap [0.30, 0.70]) \\
&y2 / ([0.40, 0.65] \cap [0.50, 0.80] \cap [0.30, 0.75] \cap [0.35, 0.70]) \\
&y3 / ([0.45, 0.55] \cap [0.55, 0.75] \cap [0.35, 0.70] \cap [0.40, 0.65])) \\
&= 1 / (0.33 / [0.45, 0.70] + 0.66 / [0.50, 0.65] \\
&+ 1 / [0.55, 0.55]) \\
\tilde{Z} &= \tilde{Z}_1 \cup \tilde{Z}_2 \cup \tilde{Z}_3 \cup \tilde{Z}_4
\end{aligned} \tag{31}$$

## APPENDIX B

### EIA/IA EXISTING CODE AND DATA

In order to carry out a comparison with IA/EIA we used the code made available by the authors of the papers ([1], [5]). We

believe by doing this we are able to apply the methods as the authors intended, and prevent the possibility of inadvertently misrepresenting them. However, there are some limitations in the code provided which needed to be overcome in order to apply the software supplied to our data set. To maintain the integrity of the code, we did not alter it, instead we modified our dataset to work with the existing code. The following changes were made:

#### A. Example 1

The existing IA code requires a minimum of 3 left and 3 right values to be entered in order to create a FS. To achieve this with a single survey response we repeated the left and right hand values 3 times, the equivalent of entering 3 identical intervals. The EIA code requires 4 left and 4 right values to be entered, and so we repeated the left and right hand values 4 times.

#### B. Example 2

For both the IA and EIA code we repeated the left and right hand values twice, giving 4 values for each. Though only 3 are required for the IA code, it is necessary to repeat both intervals so that they carry equal weighting in the result.



**Christian Wagner** received his PhD in Computer Science in 2009 from the University of Essex, UK. He is currently a Lecturer in Computer Science at the University of Nottingham, UK. His research interests include the capture, modelling, interpretation and processing/aggregation of uncertain data. He focuses on the development of novel techniques in Computational Intelligence and their application in often multi-disciplinary contexts. He has published over 40 articles and conference papers and received the Best Paper Award from the IEEE International

Conference on Fuzzy Systems (FUZZ-IEEE) 2012 and the IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2013 (for the best paper published in 2010). He is chair of the IEEE CIS Task Force on Affective Computing and an Associate Editor for the IEEE Transactions on Fuzzy Systems.



**Simon Miller** is a Research Fellow within the Intelligent Modelling and Analysis research group at the University of Nottingham in the UK. The focus of his research is the development and application of Computational Intelligence (CI) techniques in real-world uncertain environments. He was awarded a PhD in Computer Science in 2011 from De Montfort University, Leicester, UK. Recently, he worked in collaboration with GCHQ on research that measured and modelled the variation that exists in the decision making of cyber-security experts. Dr Miller has published articles in leading journals such as Knowledge Based Systems and Soft Computing, and international conferences including the IEEE International

Conference on Fuzzy Systems (FUZZ-IEEE), is a reviewer for a number of international journals including IEEE Transactions on Fuzzy Systems and Soft Computing, and is reviewer for international conferences including FuzzIEEE 2014 and IEEE SSCI 2013.



**Jonathan M. Garibaldi** received the B.Sc (Hons) degree in Physics from Bristol University, UK in 1984, and the M.Sc. degree in Intelligent Systems and the Ph.D. degree in Uncertainty Handling in Immediate Neonatal Assessment from the University of Plymouth, UK in 1990 and 1997, respectively. He is Professor of Computer Science, and leads the Intelligent Modelling and Analysis (IMA) Research Group, in the School of Computer Science at the University of Nottingham, UK. The IMA research group undertakes research into intelligent modelling,

utilising data analysis and transformation techniques to enable deeper and clearer understanding of complex problems. His main research interests are modelling uncertainty and variation in human reasoning, and in modelling and interpreting complex data to enable better decision making, particularly in medical domains. He has made many theoretical and practical contributions in fuzzy sets and systems, and in a wide range of generic machine learning techniques in real-world applications. Prof. Garibaldi has published over 200 papers on fuzzy systems and intelligent data analysis, is an Associate Editor of *Soft Computing*, and three other journals. He has served regularly in the organising committees and programme committees of a range of leading international conferences and workshops, such as FUZZ-IEEE, WCCI, EURO and PPSN. He is a member of the IEEE.



**Derek T. Anderson** received the Ph.D. in electrical and computer engineering (ECE) in 2010 from the University of Missouri, Columbia, MO, USA. He is currently an Assistant Professor in ECE at Mississippi State University, MS, USA. His research interests include pattern recognition, data fusion, image and signal processing, clustering, computer vision and linguistic summarization of human activity from video. Prof. Anderson has published more than 60 journal articles and conference papers. He co-authored the 2013 best student paper in Automatic

Target Recognition at SPIE, he received the best overall paper award at the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) 2012, he also received the 2008 FUZZ-IEEE best student paper award, and he is an associate editor for the IEEE Transactions on Fuzzy Systems.



**Timothy C. Havens** received the M.S. degree in electrical engineering from Michigan Technological University, Houghton, in 2000 and the Ph.D. degree in electrical and computer engineering from the University of Missouri in 2010.

He is currently the William and Gloria Jackson Assistant Professor in the Departments of Electrical and Computer Engineering and Computer Science at Michigan Technological University. Prior to his current position, he was an NSF/CRA Computing Innovation Fellow at Michigan State University, East

Lansing. He has published more than 60 journal articles, conference papers, and book chapters.

Dr. Havens received the Best Overall Paper Award at the IEEE International Conference on Fuzzy Systems in 2012, the IEEE Franklin V. Taylor Award for Best Paper at the 2011 IEEE Conference on Systems, Man, and Cybernetics, and the Best Paper Award from the Midwest Nursing Research Society in 2009. He is an Associate Editor for the IEEE Transactions on Fuzzy Systems.