Information Fusion-2-Text: Explainable Aggregation via Linguistic Protoforms

Bryce J. Murray[1(✉)], Derek T. Anderson[1], Timothy C. Havens[2], Tim Wilkin[3], and Anna Wilbik[4]

[1] Electrical Engineering and Computer Science Department, University of Missouri, Columbia, MO, USA
bmndc@mail.missouri.edu
[2] College of Computing and Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI, USA
[3] School of Information Technology, Deakin University, Geelong, VIC, Australia
[4] Information Systems Group of the Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract. Recent advancements and applications in artificial intelligence (AI) and machine learning (ML) have highlighted the need for explainable, interpretable, and actionable AI-ML. Most work is focused on explaining deep artificial neural networks, e.g., visual and image captioning. In recent work, we established a set of indices and processes for explainable AI (XAI) relative to information fusion. While informative, the result is information overload and domain expertise is required to understand the results. Herein, we explore the extraction of a reduced set of higher-level linguistic summaries to inform and improve communication with non-fusion experts. Our contribution is a proposed structure of a fusion summary and method to extract this information from a given set of indices. In order to demonstrate the usefulness of the proposed methodology, we provide a case study for using the fuzzy integral to combine a heterogeneous set of deep learners in remote sensing for object detection and land cover classification. This case study shows the potential of our approach to inform users about important trends and anomalies in the models, data and fusion results. This information is critical with respect to transparency, trustworthiness, and identifying limitations of fusion techniques, which may motivate future research and innovation.

Keywords: Deep learning · Machine learning · Information fusion · Information aggregation · Fuzzy integral · Explainable artificial intelligence · XAI · Protoform · Linguistic summary

1 Introduction

We live in a world that is recognizing the potential of *artificial intelligence* (AI) and *machine learning* (ML) in everyday settings. These tools have been integrated into many aspects of our daily lives—whether we realize it or not. These
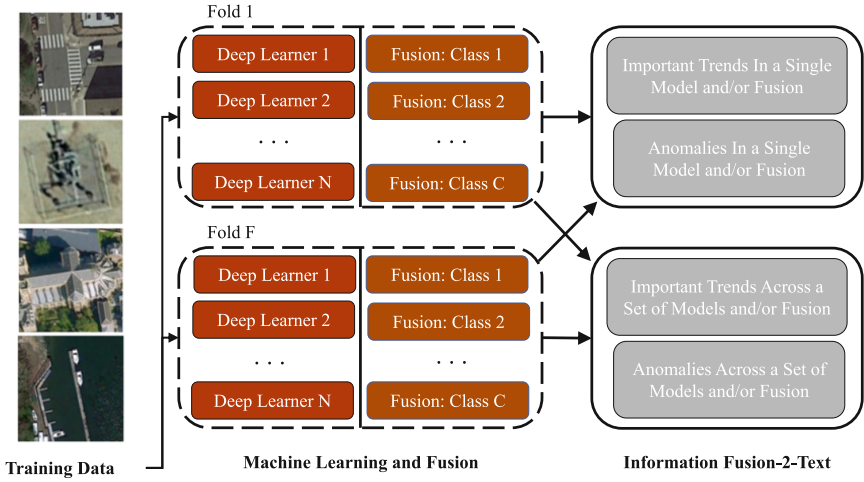
Fig. 1. Graphical illustration of fusion-2-text for the case study explored herein involving object detection and land cover classification in remote sensing. First, multiple machine learning (ML) models are trained in a cross validation context. Next, fusion is used to combat the fact that no single ML architecture is best across all data and classes. However, what have we learned? Fusion-2-text is used to discover a succinct set of important summaries or anomalies for a user in or across models.

tools, which were birthed from academic exercise, are no longer just in academia; they have found home in many different applications. Various AIs are being used to solve real-world problems, or they *simply* make our lives more convenient. Many of these algorithms are built on data-driven methods which scientists, researchers, and engineers have creatively developed and applied mathematics to build. Despite the mathematical foundations, it has become common for these tools to produce solutions that are not understandable. However, many applications require an explanation as to why a machine made a particular decision.

One task in the AI community is data or information fusion. One type of fusion revolves around the *Choquet Integral* (ChI), which can be learned from data. There are many ways to learn the ChI, the reviewer can refer to [18] for a recent review. However, once these parameters are learned, explanations can be derived from the data and learned model. In [9], we exploited properties of the ChI to understand which parameters in a learned model are supported by data. In [17], we exploited this knowledge to produce indices that describe different properties of a learned ChI. Unfortunately, the large quantity of values that our indices produce can be daunting and can lead to information overload. For example, when fusing multiple deep convolutional neural networks (DCNN) for classification, there will be one ChI learned per class—a set of XAI indices for each learned ChI. As a result, there is a need to summarize these results to explain the model at a higher-level as well as reduce the amount of information that they produce.

This begs the question of how do we reduce the number of outputs while maintaining the integrity of our XAI descriptions. Herein, we explore the production of linguistic summaries that concisely describe the relevant information coming from the XAI indices. Linguistic protoform summaries (LPS) are an efficacious mechanism of describing data because natural language is easier for humans to interpret. LPSs are statements with a specific format that are used to describe data in natural language. LPSs have been shown to be an effective means to more easily comprehend a set of data. For example, in [14] LPSs were applied to time-series data, the authors of [26] utilized LPSs to describe the restlessness of eldercare residents for healthcare, and LPSs have also been used for data mining for knowledge discovery [15].

An LPS generally follows one of several templates. These templates fall into the category of *simple protoforms* or *extended protoforms*. The simple protoform, first introduced in [29], is constructed from three different concepts: the *quantifier*, *summarizer*, and *truth value*. An example of a simple protoform is "most papers are readable". In this example, the quantifier is "most," the summarizer is "readable," and the truth value would be computed to determine the degree to which that statement is valid. As time passed, others have extended the simple protoform's template [14,28]. While there are several extensions, one example of an extended protoform (that includes an additional summarizer) is "few papers are readable and noteworthy". Not only has the LPS template been modified, but Yager's original computation of truth values has been scrutinized. In [25], it was shown that Yager's original equations to compute truth may not suited for all membership functions that model the protoforms because they may produce non intuitive summaries. As such, [25] and [12] used the Sugeno integral to overcome shortcomings in Yager's equations. Moreover, the authors of [3] present a holistic view of the development of quantifying sentences and the equations that drive this process.

The main contributions of this paper are as follows. First, we explore the potential for LPSs to reduce the complexity and amount of XAI information for the ChI. To the best of our knowledge, this has not been explored to date. Second, at a high-level, we explore what type of summaries are useful and relevant and should be reported. Third, we propose a way to derive LPSs from two of our data-centric and model-centric indices. While this is only performed on two indices herein, due to space, we discuss how our procedures generalize to other XAI indices. Last, we give a case study for aggregating a set of heterogeneous architecture DCNNs for object detection and land cover classification on a benchmark remote sensing dataset. The benefit of this study is to show actual summaries and assess if they are useful.

The breakdown of this paper is as follows. In Sect. 2, we give a brief overview of the ChI and its optimization, and we identify its data supported parameters. In Sect. 3, we present the XAI indices, Sect. 4 describes how to construct LPSs, and Sect. 5 shows how to construct fuzzy sets with respect to our indices. Last, we present our case study in Sect. 6 and insights are drawn from our data and LPSs. Figure 1 shows the technical breakdown of our fusion-2-text and Fig. 1 shows our remote sensing case study.

2   Choquet Integral

The utility of the ChI has been demonstrated in numerous applications, e.g., [4,6,21,24]. The ChI is a powerful, nonlinear aggregation operator that is parameterized by a fuzzy measure (FM). Let $X = \{x_1, ..., x_N\}$ be a set of N information sources. With respect to a finite domain, the FM, $\mu : 2^X \to \mathbb{R}$, is a function that satisfies: (i) (boundary condition) $\mu(\emptyset) = 0$, and (ii) (monotonicity) if $A, B \subseteq X$, and $A \subseteq B$, $\mu(A) \leq \mu(B)$[1]. It is often convenient to think about the FM as not just free parameters but as a modeling of *interactions* (e.g., possibly correlations) between subsets. The ChI[2] is

$$\int \mathbf{h} \circ \mu = \sum_{j=1}^{N} h_{\pi(j)} \Big(\mu(A_j) - \mu(A_{j-1}) \Big), \tag{2}$$

where π is an ordering of $\mathbf{h} = (h(x_1), ..., h(x_N))$[3]. Furthermore, $h(x_i) = h_i \in \mathbb{R}$ is the input from source i, such that $h_{\pi(1)} \geq h_{\pi(2)} \geq \ldots \geq h_{\pi(N)}$. Last, A_j corresponds to the subset $\{x_{\pi(1)}, ..., x_{\pi(j)}\}$.

The ChI can alternatively be thought about as $N!$ linear convex sums (LCS)[4,5], as each sort of the data yields an LCS. Herein, we follow the nomenclature defined in [17], and we call each sort of the data a *walk* (in the Hasse diagram).

2.1   Optimization

Defining the FM variables in the ChI is not a trivial task and there are many ways to identify them, e.g., [5,10,11,16]. However, in our current data-driven era, it is common place to learn the FM variables. Herein, we use our learning algorithm put forth in [9]. We do not describe the algorithm due to limited page length. The techniques proposed herein extend beyond a specific learner, they are applicable to any ChI derived from data.

2.2   Data Supported Variables

In [9], we established that data supported variables can be identified for the ChI. A variable is called supported if any walk of the data includes it. For example, let $N = 3$ and $h_3 > h_1 > h_2$. The FM variables that are encountered are $\mu(\{h_3\})$, $\mu(\{h_1, h_3\})$ and $\mu(X)$. By considering all the given inputs in the training data, we can easily determine all data supported variables. This fact is important to many of the upcoming indices (Fig. 2).

---

[1] While not required, it is common in practice to impose $\mu(X) = 1$.

[2] It is important to note that when a FM is set (values are specified), the ChI becomes a specific aggregation operator. For example, consider $\mu(A) = 1, \forall A \in X$, except $\mu(\emptyset)$. As such, the ChI reduces to the max operator.

[3] Hereafter, $h(x_i)$ will be shortened to h_i for simplicity.

[4] When $\mu(X) = 1$.

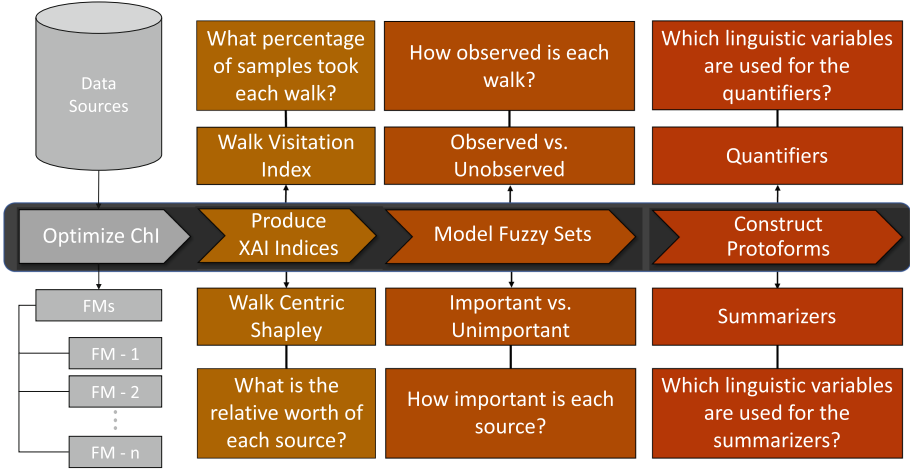[5] Who share 2^N weights.

Fig. 2. Illustration of computational stages in our fusion-2-text.

3   Existing Low-Level XAI Indices

In [17], we proposed a set of XAI indices—measures/functions that summarize and highlight important properties about the FM and ChI—in the context of data-driven learning. In [18], we expanded our initial set of indices, including the Shapley and interaction index, to operate more accurately on partially observed domains. In general, our indices can be partitioned into three sets: information about the FM (the sources and their interactions), inquiries about the FM-ChI (e.g., what is the specific aggregation), and inquiries about the data relative to the ChI (e.g., what parts of our model are missing). While each of our indices provide valuable and different insights, we limit the scope herein to one data specific index, walk visitation, and one model specific index, the walk centric Shapley index.

3.1   Walk Centric Shapley

The Walk Centric Shapley (WCS) index is an extension of the Shapley index [20]. The WCS defines the relative worth of each source with respect to its data supported variables. This extension is valuable because the traditional Shapley may be drastically over- or under-estimate the worth of sources as it assumes that the FM is fully observable. The WC Shapley is

$$\bar{\Phi}_\mu(i) = \sum_{K \subseteq X \backslash \{i\}} \zeta_{X,2}(K) \left(\mu(K \cup \{i\}) - \mu(K) \right), \tag{3a}$$

$$\zeta_{X,2}(K) = \frac{(|X| - |K| - 1)! |K|!}{|X|!} \mathbb{1}_{(K \cup \{i\})} \mathbb{1}_{(K)}, \tag{3b}$$

$$\widetilde{\Phi}_\mu(i) = \frac{\bar{\Phi}_\mu(i)}{\sum_{j=1}^i \bar{\Phi}_\mu(j)}, \tag{3c}$$

where $K \subseteq X \backslash \{i\}$ denotes all proper subsets from X that do not include source i and $\mathbb{1}$ is an indicator function that is 1 if the FM value is data-supported and 0 otherwise. The Shapley values of μ is the vector $\widetilde{\boldsymbol{\Phi}}_\mu = [\widetilde{\Phi}_\mu(1), ..., \widetilde{\Phi}_\mu(N)]^t$ where $\sum_{i=1}^N \widetilde{\Phi}_\mu(i) = 1$. The WCS values are important because they inform us about the relative worth of each information source.

3.2   Walk Visitation

Understanding the quality of the information sources is merely one aspect of the big XAI picture. It is also important to understand the quality (e.g., completeness) of a learned ChI. Herein, we use the walk visitation metric [17], which describes how many unique walks were taken within the training data. We quickly summarize the index due to limited page count. The index works by sorting all samples (according to their input values), finding which walks were encountered, and dividing the number of times that they were observed by the total number of samples. The goal of this metric is to determine the degree to which each walk was observed. If a probability is zero, then a walk was never seen. Furthermore, if we get a new input for the ChI and its walk was not encountered in training, then one should question the ChI output. In [17], this index was used to derive additional indices, like to what degree should we trust an output of the ChI.

4   Protoforms

Protoform-based linguistic summaries are often an effective liaison between data the data interpreter. As such, deriving linguistic summaries with respect to the XAI indices has the potential to effectively reduce the amount of information by producing concise summaries. Furthermore, as less is often more, there is also the potential to remediate confusion due to complexity, which can improve decision making. While there are multiple LPS templates to follow, the *simple* protoform will suffice for the insights that we are drawing herein. The simple protoform takes the following format,

$$Qy\text{'s are } P. \tag{4}$$

Within the protoform, Q is a linguistic *quantifier*, y is a set of *objects*, and P is the *summarizer*. Both Q and P are modelled by fuzzy sets over the desired domains.

Examples of a quantifier that are suitable for our problem are words like "few", "many", or "most"; whereas examples of a summarizer may be "important" or "observed" when referencing the sources or the walks, respectively. Moreover, an example of a protoform with respect to the XAI indices may be "few walks are observed". With respect to the XAI indices, we produce summaries that describe the importance of each of the sources across all models and how many walks are observed in each model. To do this, we use the vocabulary in Fig. 3.
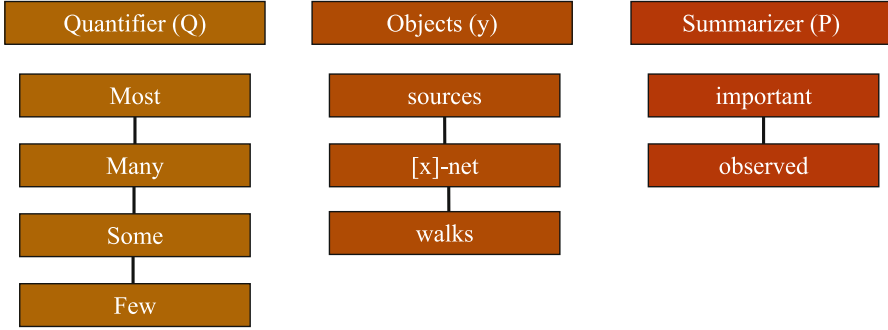

Fig. 3. Vocabulary used herein to produce fusion LPSs.

An LPS has a value of *truth* associated with it. This concept, first introduced by Yager [29], utilizes Zadeh's calculus to use Eq. 5 to determine the truth value, T, associated with the linguistic summary. This equation is as follows,

$$T(Ay's\,are\,P) = A(\frac{1}{N}\Sigma_{i=1}^{N}P(y_i)).$$ (5)

However, Eq. 5 may produce non-intuitive results, as noted in [25] and then in [12]. As a result, the fuzzy integral can also be employed to determine the truth value; however, for the scope of our case study, Eq. 5 will suffice.

5   Fuzzy Sets

As mentioned in Sect. 4, $Q(x)$ and $P(x)$ are fuzzy sets. In each case, we use the trapezoidal membership function for the fuzzy sets, and we have empirically determined their parameters. This is acceptable for our initial work, as we have spent a good amount of time working with the ChI and the benchmark data set. However, in future work these sets and their values clearly should be learned from data or a set of experts.

5.1   Walk Centric Shapley

To define the fuzzy sets with respect to the WCS, a fuzzy set must be defined for $P(x)$ (i.e. *important*) and $Q(x)$ (*few, some, many, most*). The fuzzy set is defined

by the trapezoidal membership values from the WCS values. In an ideal fusion solution, all Shapley values would be equal to $\frac{1}{N}$, meaning equal contribution. As a result, we consider the source to be *important* if it has a value greater than or equal to $\frac{1}{N}$. Once $P(x)$ is defined, $Q(x)$ can be modelled. The output of averaging $P(x)$ is passed to $Q(x)$, and it is between $[\frac{1}{N}, 1]$ because at least 1 source will be important. As such, the domain of $Q(x)$ is $[\frac{1}{7}, 1]$. Using Eq. 5, we can compute the truth value of each of the statements allowing us to isolate the most relevant summaries that are produced.

5.2   Walk Visitation

Similar to the fuzzy sets that govern the WCS values, fuzzy sets are used to model the walk visitation index. For $P(x)$, the fuzzy sets model the concept of *observed* and *unobserved*. In an ideal case, we desire each walk to have an equal walk visitation index, so with this in mind, we consider the walk to be observed if $z \geq \frac{1}{N!}$. Next, the fuzzy set for $Q(x)$ must be modelled. At least 1 walk will be observed, and it is possible to observe up to however many training samples exist. As such, the domain of $Q(x)$ is $[\frac{1}{M}, 1]$, which is approximately $[0, 1]$. The values of each of the fuzzy sets can be found in Table 1.

Table 1. Trapezoidal membership function parameters.

Walk Visitation Quantifier	a	b	c	d	WC Shapley Quantifier	a	b	c	d
Few	0	0	$\frac{1}{7}$	$\frac{2}{7}$	Few	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{3}{7}$
Some	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{4}{7}$	Some	$\frac{2}{7}$	$\frac{2.5}{7}$	$\frac{3.5}{7}$	$\frac{5}{7}$
Many	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{5}{7}$	$\frac{6}{7}$	Many	$\frac{3}{7}$	$\frac{4.5}{7}$	$\frac{5.5}{7}$	$\frac{6}{7}$
Most	$\frac{5}{7}$	$\frac{6}{7}$	1	1	Most	$\frac{5}{7}$	$\frac{6}{7}$	1	1
Summarizers	a	b	c	d					
Important	0	$\frac{1}{7}$	1	1					
Observed	0	$\frac{1}{7!}$	1	1					

6   Case Study

To show how these indices work in a real-world application, we consider the fusion of a set of 7 different DCNNs for object detection and land classification of remote sensing data. The DCNNs that we fuse are CaffeNet [13], GoogleNet [23], ResNet 50 [7], ResNet 101, DenseNet [8], InceptionResNetV2 [22], and Xception [2]. The dataset is the AID remote sensing data set [27]. This dataset is composed of 10,000 images over 30 different aerial scene types.

The complete training procedure of how these networks were trained can be found in [19]. Furthermore, the complete description of how these DCNNs

are fused can be found in [1]. Due to the nature of the multistep classification problem (DCNNs and ChI), it is an important step to determine how to split the data into training and testing. For the DCNNs, five-fold cross validation was used. This means that four folds are used for training, and one fold is used for evaluation. From the evaluation fold, two-fold cross validation is used, due to the limited number of samples in AID. To ensure that each class is approximately balanced in each of the folds, an approximately equal number of samples were chosen from each class. There are multiple ways to perform fusion across the DCNNs. Herein, we train a ChI for each of the 30 classes. As a result, there are $30 \times 7!$ walk visitation values produced, and 30×7 Shapley values produced (a total of $151,410$ values). Using the proposed LPS configuration, we reduce the XAI indices to a few sets of LPSs that are more easily comprehended.

6.1   Source Summaries

XAI Question: How Many Sources are Important?

With respect to each ChI, we can determine how many important sources there are for each class. To produce this set of summaries, we treat each of the 30 ChIs (one per class) as our objects, y. The linguistic summarizer is "important", and our quantifiers are "few", "several", "many", and "most". Figure 4 illustrates the quantity of summaries that are assigned each of the quantifiers.
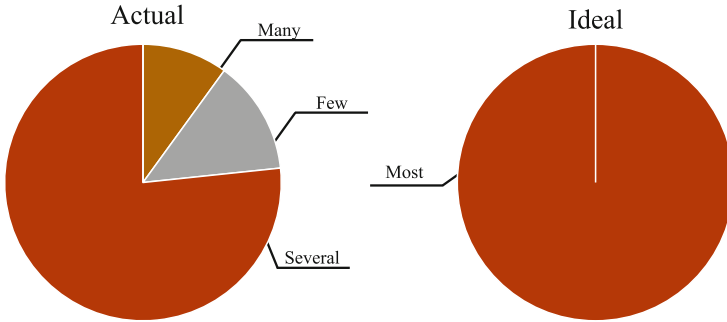

Fig. 4. Actual versus the ideal distribution of quantifier.

In an ideal case, "most" of the sources would be important, meaning all sources are contributing to the fusion solution. However, our experiment produced no summaries of "most sources are important". The majority of the ChIs are summarized by "several sources are important." This begs the question of which sources are important because we now know that not all are important.

XAI Question: How Important is Each Source?

Taking our knowledge from the last set of summaries, it is logical to now isolate the important sources. To do this, we produce a set of summaries specifically for each of the sources. We treat each DCNN across all models as

our objects, y. For example, one set of objects will be *GoogleNets*, such that
GoogleNets = {*GoogleNets*$_1$, *GoogleNets*$_2$, ..., *GoogleNets*$_{30}$}. In Fig. 4, we show
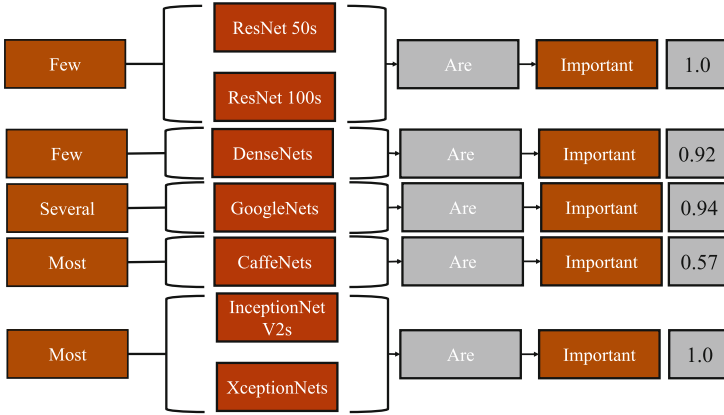the resulting summaries (Fig. 5).

Fig. 5. LPSs describing the importance of each DCNN with truth degree (gray).

From this set of summaries (with strong truth degrees), we can conclude that
ResNet 50 and ResNet 100 are not contributing to the fusion solution; however,
InceptionNet and XceptionNet are important in *most* of the ChIs, meaning they
are strong contributors. This leads us to conclude that ResNet 50 and ResNet
100 can likely be removed, speeding up inference by reducing DCNNs.

XAI Question: How Many Walks are Observed Per Model?
In this case, the summarizer is "observed," the objects are each ChI, and the
quantifier is again, "few", "several", "many", and "most". However, for this
set, there was only one summary ever found, "few walks are observed." This
quickly magnifies the flaw with these models because many of the possible walks
have not been observed—meaning the FM values of many of the walks are not
actually learned from data. In order to fully learn the ChI, "most" walks must
be observed. Observing few walks means that there is not much diversity in the
data. This highlights that we may have a dominant walk, or that we only ever
observe a relatively low number of walks.

XAI Question: How Observed is Each Walk Across the Data?
Whereas the last summary encapsulated information pertaining to how many
walks are observed per model, this set of summaries answers the question of
how observed is each walk across the entire data set. We consider each specific
walk as an object; for example the walk $[1, 2, 3, 4, 5, 6, 7] = \{[1, 2, 3, 4, 5, 6, 7]_1,$
$[1, 2, 3, 4, 5, 6, 7]_2, ..., [1, 2, 3, 4, 5, 6, 7]_{30}\}$. When producing these summaries,

there would be one summary per walk, so in our example, this would generate 7! summaries, which is far too many for anyone to digest. However, by only evaluating the walks that are observed in "most" ChIs with a high truth degree, we only consider 5 of the walks. The summaries shown in Fig. 6 are produced.
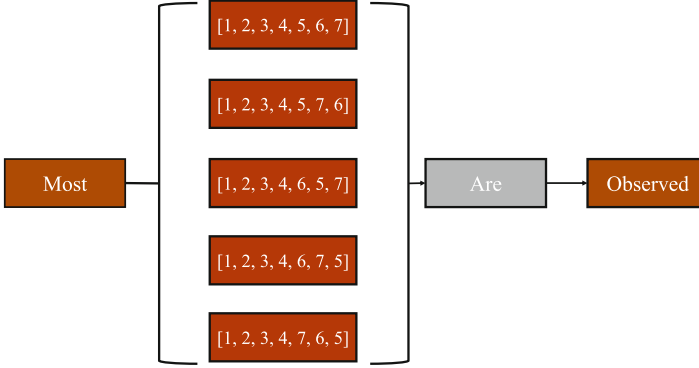

Fig. 6. Most *specific walk*s are observed.

While each of these 5 walks have a truth value of 1, they are also the only walks to have the quantifier "most". There are 4,978 walks that are observed a "few" times with a truth degree of 1. This leaves 67 walks that are observed "few", "several", or "most" with a truth values less than 1 for each of them. These summaries clearly highlight that there may be some bias in the data. Specifically, the first 4 sources are typically encountered in the same order, which shows that something is not quite right. This allows us to dig into the data to figure out what might be going on.

6.2   Code

The ability to reproduce an experiment is a cornerstone in the scientific community. As such, we provide the code that produced these summaries at the following repository, https://github.com/B-Mur/ChoquetIntegral.git. Moreover, the data set that was used can be found at the following repository, https://github.com/aminb99/remote-sensing-nn-datasets.

7   Summary and Future Work

Herein, we have proposed and implemented the use of LPSs to reduce a high number of metrics to a short number of concise and more useful summaries. To our knowledge, this is the first work that produces linguistic summaries to explain fusion, without a doubt relative to data-driven learning. Before producing the summaries, the indices produce a large quantity of metrics that are complex to

interpret. By producing the summaries, the indices effectively reduce the information that must be digested, while maintaining the integrity of the indices. By first determining only few sources are important, it is a logical step to determine which sources are important. If all sources were important, it would be unnecessary to determine the important sources as they are all important.

The walk visitation summaries tell a similar story. Only few walks are ever observed; this leads us to produce summaries determining the walks are observed across all data (only 5). Before we produce these summaries, these metrics are a raw stream of data that are not intuitive, and interpretable only by those with significant domain experience. However, the summaries allow someone unfamiliar with the indices (and the values they produce) to be practitioners of XAI with their fusion. To our knowledge, has never been done before.

In the future, we hope to generate summaries from the remaining XAI indices to provide more complete and comprehensive insights. By doing this, we will likely produce extended LPSs such that a single, extended LPS contains more information than a simple protoform can provide. We will also explore how to present, in a textual or visual fashion, this information to a human. This foundation also excites us because it is a structured format or language in which information can be extracted and then subsequently computed with. Possibilities including deriving higher-level conclusions about the data, models, and systems, or perhaps using the information to improve the training and/or use of fusion.

References

1. Anderson, D., Scott, G., Islam, M., Murray, B., Marcum, R.: Fuzzy Choquet integration of deep convolutional neural networks for remote sensing. In: Pedrycz, W., Chen, S.M. (eds.) Computational Intelligence for Pattern Recognition. Studies in Computational Intelligence, pp. 1–28. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-89629-8_1

2. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807, July 2017. https://doi.org/10.1109/CVPR.2017.195

3. Delgado, M., Ruiz, M.D., Sánchez, D., Vila, M.A.: Fuzzy quantification: a state of the art. Fuzzy Sets Syst. **242**, 1–30 (2014). https://doi.org/10.1016/j.fss.2013.10.012. http://www.sciencedirect.com/science/article/pii/S0165011413004247, theme: Quantifiers and Logic

4. Du, X., Zare, A.: Multiple instance Choquet integral classifier fusion and regression for remote sensing applications. IEEE Trans. Geosci. Remote Sens. 1–13 (2018). https://doi.org/10.1109/TGRS.2018.2876687

5. Du, X., Zare, A., Keller, J.M., Anderson, D.T.: Multiple instance Choquet integral for classifier fusion. In: 2016 IEEE Congress on Evolutionary Computation (CEC), pp. 1054–1061, July 2016. https://doi.org/10.1109/CEC.2016.7743905

6. Grabisch, M.: The application of fuzzy integrals in multicriteria decision making. Eur. J. Oper. Res. **89**(3), 445–456 (1996)

7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)

8. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, July 2017. https://doi.org/10.1109/CVPR.2017.243

9. Islam, M.A., Anderson, D.T., Pinar, A.J., Havens, T.C.: Data-driven compression and efficient learning of the Choquet Integral. IEEE Trans. Fuzzy Syst. **PP**(99), 1 (2017). https://doi.org/10.1109/TFUZZ.2017.2755002

10. Islam, M.A., Anderson, D., Petry, F., Elmore, P.: An efficient evolutionary algorithm to optimize the Choquet Integral. Int. J. Intell. Syst. **34**, 366–385 (2018). https://doi.org/10.1002/int.22056

11. Islam, M.A., Anderson, D.T., Pinar, A., Havens, T.C., Scott, G., Keller, J.M.: Enabling explainable fusion in deep learning with fuzzy integral neural networks. IEEE Trans. Fuzzy Syst. 1 (2019). https://doi.org/10.1109/tfuzz.2019.2917124

12. Jain, A., Keller, J.M.: On the computation of semantically ordered truth values of linguistic protoform summaries. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8, August 2015. https://doi.org/10.1109/FUZZ-IEEE.2015.7337822

13. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, MM 2014, pp. 675–678. ACM, New York (2014). https://doi.org/10.1145/2647868.2654889

14. Kacprzyk, J., Wilbik, A., Zadrozny, S.: Mining time series data via linguistic summaries of trends by using a modified Sugeno integral based aggregation. In: 2007 IEEE Symposium on Computational Intelligence and Data Mining, pp. 742–749, March 2007. https://doi.org/10.1109/CIDM.2007.368950

15. Kacprzyk, J., Zadrozny, S.: Data mining via protoform based linguistic summaries: some possible relations to natural language generation. In: 2009 IEEE Symposium on Computational Intelligence and Data Mining, pp. 217–224, March 2009. https://doi.org/10.1109/CIDM.2009.4938652

16. Keller, J.M., Osborn, J.: A reward/punishment scheme to learn fuzzy densities for the fuzzy integral. In: Proceedings of International Fuzzy Systems Association World Congress, pp. 97–100 (1995)

17. Murray, B., Anderson, D., Islam, M.A., Pinar, A., Scott, G., Havens, T.: Explainable ai for understanding decisions and data-driven optimization of the Choquet integral. In: World Congress on Computational Intelligence (WCCI), July 2018

18. Murray, B., et al.: Explainable AI for the Choquet integral (accepted). IEEE Trans. Emerg. Top. Comput. Intell.

19. Scott, G.J., England, M.R., Starms, W.A., Marcum, R.A., Davis, C.H.: Training deep convolutional neural networks for land-cover classification of high-resolution imagery. IEEE Geosci. Remote Sens. Lett. **14**(4), 549–553 (2017)

20. Shapley, L.S.: A value for n-person games. Contrib. Theory Games **2**, 307–317 (1953)

21. Smith, R.E., et al.: Genetic programming based Choquet integral for multi-source fusion. In: IEEE International Conference on Fuzzy Systems, July 2017

22. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: AAAI (2016)

23. Szegedy, C., et al.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition (CVPR) (2015)

24. Tahani, H., Keller, J.: Information fusion in computer vision using the fuzzy integral. IEEE Trans. Syst. Man Cybern. **20**, 733–741 (1990)

25. Wilbik, A., et al.: Evaluation of the truth value of linguistic summaries - case with nonmonotonic quantifiers. In: Angelov, P., et al. (eds.) Intelligent Systems. Advances in Intelligent Systems and Computing, vol. 322, pp. 69–79. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11313-5_7
26. Wilbik, A., Keller, J.M., Bezdek, J.C.: Linguistic prototypes for data from eldercare residents. IEEE Trans. Fuzzy Syst. **22**(1), 110–123 (2014). https://doi.org/10.1109/TFUZZ.2013.2249517
27. Xia, G., et al.: AID: a benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. **55**(7), 3965–3981 (2017). https://doi.org/10.1109/TGRS.2017.2685945
28. Yager, R.R.: Fuzzy summaries in database mining. In: Proceedings the 11th Conference on Artificial Intelligence for Applications, pp. 265–269, February 1995. https://doi.org/10.1109/CAIA.1995.378813
29. Yager, R.R.: A new approach to the summarization of data. Inf. Sci. **28**(1), 69–86 (1982). https://doi.org/10.1016/0020-0255(82)90033-0