Notes on Intuitionistic Fuzzy Sets
Print ISSN 1310–4926, Online ISSN 2367–8283
2024, Volume 30, Number 2, 113–120
DOI: 10.7546/nifs.2024.30.2.113-120

On OWA, Machine Learning and Big Data: The case for IFS over universes

Panagiotis Chountas [1], Mustafa Hajmohammed [2] and Ismael Rhemat [3]

[1] School of Computer Science & Engineering, University of Westminster
15 New Cavendish Street, London W1W 6UW, London, United Kingdom
e-mail: chountp@westminster.ac.uk

[2] School of Computer Science & Engineering, University of Westminster
15 New Cavendish Street, London W1W 6UW, London, United Kingdom
e-mail: m.hajmohammed@westminster.ac.uk

[3] School of Computer Science & Engineering, University of Westminster
15 New Cavendish Street, London W1W 6UW, London, United Kingdom
e-mail: i.rhemat@westminster.ac.uk

Received: 24 April 2024 **Revised:** 13 May 2024
Accepted: 18 May 2024 **Online First:** 1 July 2024

Abstract: This paper provides a holistic view of open-world machine learning by investigating class discovery, and class incremental learning under OWA. The challenges, principles, and limitations of current methodologies are discussed in detail. Finally, we position IFS over multiple universes as a formalism to capture the evolution in Big Data as part of incremental learning.

Keywords: Intuitionistic fuzzy sets, Big data, Incremental learning, Machine learning.

2020 Mathematics Subject Classification: 03E72, 68T09.

1 Big Data and OWA

Research in Semantic Web, Data Streams, Big Data assume that the Semantic Web assumes an Open World known commonly as the *Open World Assumption* (*OWA*). What is the distinction between the Open World Assumption and the Closed World Assumption?

The Closed World Assumption (CWA) is the assumption that what is not known to be true must be false. The Open World Assumption (OWA) is the opposite. In other words, it is the assumption that what is not known to be true is simply unknown. Consider the following statement: *"Christo is a footballer."* Now, what if we were to ask *"If Christo is a manager?"* Under a CWA, the answer is *"no"*. Under the OWA, it is *"I don't know"*.

1.1 CWA and OWA in databases and ontologies

The CWA applies when a system has complete information. This is the case for many database applications. For example, consider a database application for airline reservations. If you are looking for a direct flight between Sofia and Tripoli, and it does not exist in the database, then the result is *"There is no direct flight between* Sofia and Tripoli*."* For this type of application, this is the expected and correct answer.

On the other hand, OWA applies when a system has incomplete information. This is the case when we want to represent knowledge (ontologies) and want to discover new information. For example, consider a patient's clinical history. If the patient's clinical history does not contain a particular allergy, it would be incorrect to state that the patient is not likely to suffer from that allergy. It is unknown if the patient is exposed to that allergy, unless more information is given to disprove the assumption. How to decide what is accurate? Rules that define what is syntactically correct, e.g.,

- regular expressions;
- constraints to define what values are semantically acceptable;
- validity interval with reference to valid time.

1.2 OWA and Semantic web

CWA is not only about returning *"no"* and OWA is not only about returning *"I don't know."* Consider the following example:

Note that in the CWA case, we assumed that footballer and manager are different professions. With OWA, this is not assumed. This is what is called Unique Named Assumption (UNA). CWA systems have UNA. OWA systems do not have UNA. However, one could manually add the UNA. In our example, if we add the following statement: *"Footballer is different from Manager,"* the OWA would now generate an inconsistency. The OWA logic is the following: *"If a person can only have a single occupation, and if Christo is a Footballer and Manager, then Footballer and Manager must be the same thing; but hold on, Footballer and Manager are different, so they can't be the same! Something is wrong."*

Recall that OWA is applied in a system that has incomplete information. Guess what the Web is? The Web is a system with incomplete information. Absence of information on the web means that the information has not been made explicit. That is why the Semantic Web uses the OWA. The essence of the Semantic Web is the possibility to infer new information.

2 Machine learning on Big Data

In machine learning techniques, supervised machine learning involves isolated classification or regression task, which learns a function (model)

$$f : X \rightarrow Y \text{ from a training dataset } D = \{(\boldsymbol{x}_i, y_i)\},\ i = 1, \ldots, N. \tag{1}$$

In the classification task, all the classes y that the model will encounter during deployment must have been seen in training, i.e., $y \in Y$, then the model f can be deployed to predict future encountered inputs.

However, the current building of machine learning models is largely based on the closed-world assumption where the important factors of learning are limited to what has been observed during training. This assumption is correct in restricted domains where, possible classes are well-defined and unlikely to change over time. For example, in image recognition task, the closed-world assumption holds because the set of digits (0 – 9) is fixed and known in advance. Besides, this assumption also makes the data preprocessing easier and straightforward.

For example, in a sensor driven system, a sensor may encounter novel objects that have been never learned before; the closed-world assumption can be problematic in such situations. In this paradigm, a machine learning model must be able to identify and separate unknown inputs that deviate from training classes to keep safe, and then discover new classes from unknown instances with incremental learning and accumulate knowledge without re-training the whole model from scratch.

In the canonical closed world setting, the test dataset shares categories with training data with no additional classes, i.e., $Y_{\text{test}} = Y_{\text{train}}$.

By contrast, in the open-world, samples from new classes emerge, i.e., $Y_{\text{train}} \neq Y_{\text{test}}$, we further express it as $Y_{\text{train}} \subset Y_{\text{test}}$.

Unknown treatment is the first step towards open-world machine learning, which is also a fundamental ability of the classifier in the open-world. Considering that a training set comprises of K classes, i.e.,

$$D_{\text{train}} = D_{\text{in}} = \left\{\left(\mathbf{x}_i, y_i\right)\right\}_{i=1}^{N} \subset X \times Y_{\text{in}}, \text{ where } Y_{\text{in}} = \{1, 2, \ldots, K\}. \tag{2}$$

Once trained on D_{in}, models are supposed to move unknown samples from classes outside of Y_{in}, rather than classifying it irresponsibly into one of K categories. The idea behind unknown treatment is that when encountering unfamiliar knowledge, it is essential to acknowledge the limitations rather than pretend false understanding and give random answers.

Many efforts have been made to enhance the unknown treatment ability of machine learning systems. There are multiple research areas related to unknown rejection, such as anomaly detection [5, 12] out-of-distribution (OOD) detection [9, 13], and open-set recognition [8].

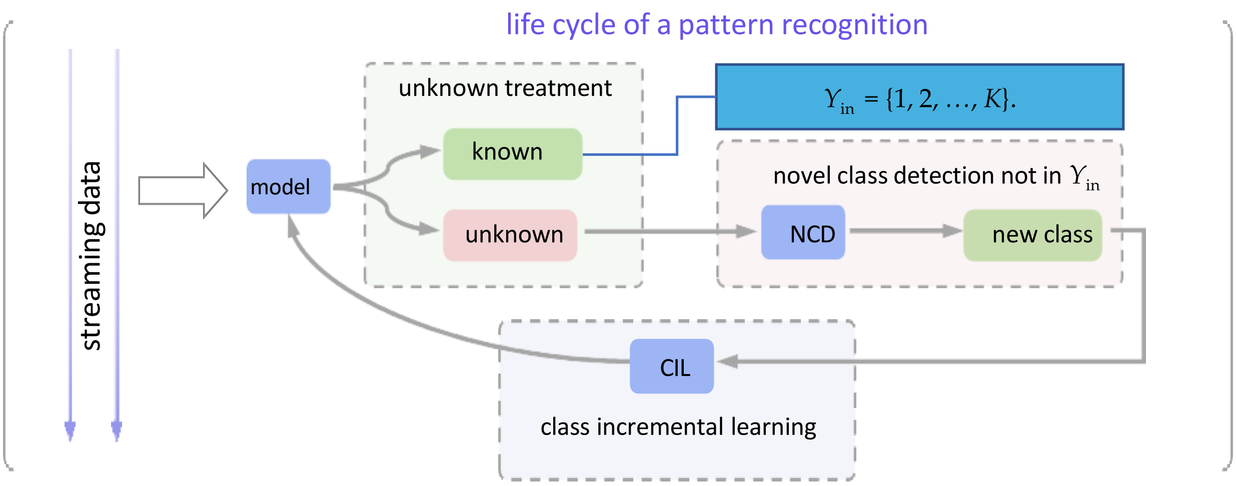

Figure 1. Life cycle of a pattern recognition system in data streaming under OWA

2.1 Novel class detection under OWA

Novel class detection [4, 14] is the next step in open-world machine learning, which aims to automatically discover original categories from unlabeled data based on the model's prior knowledge. Novel class detection is an extension of unknown treatment, requiring models to not only reject unknown samples, but also further extrapolate to classify the rejected samples.

Given a sample training dataset $D_{\text{train}} = D_\ell \cup D_u$ with two subsets, where

$$D_\ell = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_\ell} \subset X \times Y_\ell \tag{2}$$

is the labelled dataset, and

$$D_u = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_u} \subset X \times Y_u \tag{3}$$

is the unlabelled data.

It is worth noting under CWA that $Y_u \neq Y_\ell$, i.e., there are new classes in the unlabeled data outside the old classes in D_ℓ. Let C_{old}, C_{new} denote the old and new classes in the labeled and unlabeled dataset, with the number of classes $K_{\text{old}} = |C_{\text{old}}|$, $K_{\text{new}} = |C_{\text{new}}|$, respectively. The objective is to cluster and discover C_{new} in D_u leveraging the knowledge learned on D_ℓ.

Generalize Class Discovery (GCD) in the spirit of OWA is a more pragmatic and challenging task, relaxing the strong assumption in NCD that unlabeled data all come from novel classes, i.e.,

$$Y_u = C_{\text{old}} \cap C_{\text{new}} \text{ and } Y_\ell \subset Y_u . \tag{4}$$

In essence, novel class discovery is a clustering task on D_u. Regarding its differences from unsupervised clustering [4, 14] the latter aims to cluster unlabeled data in a purely unsupervised manner without any prior knowledge of classification criterion, as a result, unsupervised clustering is not a fully learnable task.

After detecting unknown instances, those should be labelled by humans or novel class discovery strategies. Then the system must continually extend the multi-class classifier to learn those new classes, which is referred to as class-incremental learning (CIL), being the third step in the open-world recognition process. Typically, an incremental learner learns several tasks sequentially, and only the data of the current task can be accessed by the learner.

In Generalize Class Discovery (GCD) we had $Y_\ell \subset Y_u$ because $D_\ell \subset D_u$. However L and U represent different universes at different timepoints as the number of features is different as $N_\ell \subset N_u$. What we are pointing that the Generalize category discovery (GCD) under OWA can be defined and represented as an IFS over Hierarchical Universes using the definitions and properties below.

3 IFS over hierarchical universes in big data and OWA

Let E and F be two different universes, [1, 2] and let A_E and B_F be IFSs over E and F, respectively, i.e.,

$$A_E = \{\langle x, \mu_A(x), \nu_A(x)\rangle \mid x \in E\},$$

$$B_F = \{\langle x, \mu_B(x), \nu_B(x)\rangle \mid x \in F\}.$$

The IFS A, which is defined over the universe E, we shall call an "E-IFS". The operations over A and B are defined by:

$$A_E = \{\langle x, \nu_A(x), \mu_A(x) \rangle \mid x \in E\},$$

$$A_E \cap B_F = \{\langle x, \min(\mu_A(x), \mu_B(x)), \max(\nu_A(x), \nu_B(x))\rangle \mid x \in E \cup F\},$$

$$A_E \cup B_F = \{\langle x, \max(\mu_A(x), \mu_B(x)), \min(\nu_A(x), \nu_B(x))\rangle \mid x \in E \cup F\},$$

$$A_E + B_F = \{\langle x, \mu_A(x) + \mu_B(x) - \nu_A(x)\nu_B(x), \nu_A(x)\nu_B(x)\rangle \mid x \in E \cup F\},$$

$$A_E \,.\, B_F = \{\langle x, \mu_A(x)\, \mu_B(x), \nu_A(x) + \nu_B(x) - \nu_A(x)\nu_B(x)\rangle \mid x \in E \cup F\},$$

Now let E be a fixed universe and let A be an Intuitionistic Fuzzy Set (IFS, see [1]) over E. Let F be another universe and let the set E be an IFS over F having the form:

$$E = \{(y, \mu_E(y), \nu_E(y)) \mid y \in F\}.$$

Therefore, the element $x \in E$ has the form (see [2]):

$$x = (y, \mu_E(y), \nu_E(y)), \text{ i.e., } x \in F \times [0, 1] \times [0, 1],$$

$$A = \{((y, \mu_E(y), \nu_E(y)), \mu_A((y, \mu_E(y), \nu_E(y))), \nu_A((y, \mu_E(y), \nu_E(y)))) \mid (y, \mu_E(y), \nu_E(y)) \in E\}$$

If the degrees of membership and non-membership of an element y to a set A in the frames of a universe E are $\mu_A(y)$ and $\nu_A(y)$ and the element $(y, \mu_A(y), \nu_A(y))$ has degrees of membership and non-membership to the set E within the universe F are $\mu_E(y)$ and $\nu_E(y)$, then we define:

$$A = \{(y, \mu_E(y).\mu_A(y), \nu_E(y).\nu_A(y)) \mid y \in F\}. \quad (5)$$

Using the extend concept of H-IFS transforming some ideas and results from [1–3]. First, we shall start with an example. Let E be a finite universe with the form

$$E = \{e_1, e_2, e_3, \{e_1, e_2\}, \{e_1, e_3\}, \{e_1, e_2, \{e_1, e_3\}\}\}.$$

Therefore, the IFS A over E will have the form

$$A = \{(e_1, \mu_A(e_1), \nu_A(e_1)), (e_2, \mu_A(e_2), \nu_A(e_2)), (e_3, \mu_A(e_3), \nu_A(e_3)),\\ (\{e_1, e_2\}, \mu_A(\{e_1, e_2\}), \nu_A(\{e_1, e_2\})), (\{e_1, e_3\}, \mu_A(\{e_1, e_3\}), \nu_A(\{e_1, e_3\})),\\ (\{e_1, e_2, \{e_1, e_3\}\}, \mu_A(\{e_1, e_2, \{e_1, e_3\}\}), \nu_A(\{e_1, e_2, \{e_1, e_3\}\})\}. \quad (6)$$

$$\mathrm{card}(E) = 6.$$

Let the set E_1 be defined as

$$E_1 = \{e_1, e_2, e_3, \{f_1, f_2\}, \{f_1\}, \{g_1, g_2, \{g_1, g_3\}\}\}.$$

We can tell that elements e_1, e_2, e_3 are *"elements from first level",* elements f_1, f_2, g_1, g_2 are *"elements from second level"* and elements g_1, g_3 are *"elements from third level".*

An element of E, E_1 can be an element from two or more different types. For example, for the set E objects e_1, e_2, e_3 are elements from each one of the three types.

If there is an order between some of the elements of E, e.g., if for $i = 1, 2, 3$: $e_i = i$, this order ($\leq$ or $<$) cannot be extended over the rest E-elements. This is the basis for OLAP and materialized view queries [3].

If the order is $\subset$, it will be valid for the fourth and sixth elements of E, but will not be possible for the rest E-elements. This is true un the case of Generalized Class Discovery (GCD) under Novel Class Detection see equation [6].

For a more detailed description of Intuitionistic Fuzzy Sets over Universes with Hierarchical Structures the reader may refer to [7]. In practice, an IFS over Hierarchical Universes represents the potential hierarchy of a structure of possible worlds with an accessibility (see Figure 2).

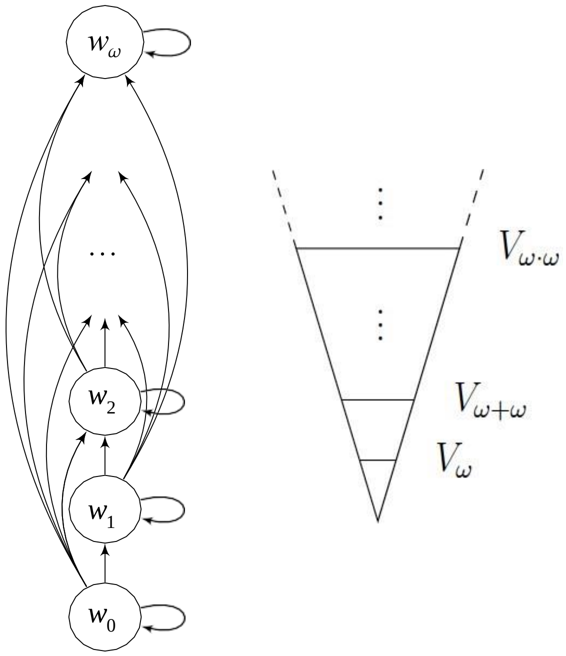

Figure 2. The potential hierarchy of sets vs. Cumulative hierarchy

Worlds are stages of the set forming process, the domain of each world consists of the sets that have been formed thus far, and the domain of each world accessible from a given world $w\alpha$ is a superset of the domain of $W\alpha$ [10, 11] From each world $W\alpha$, there are sets that do not exist at $W\alpha$ but that are possibly relative to $W\alpha$, i.e., sets that exist in the domain of a world $w\beta$ which is accessible from $w\alpha$.

The overall structure of the potential hierarchy is isomorphic to that of the cumulative $V\alpha$-hierarchy (Figure 2).

3 Conclusions

The definition of the Generalize Class Discovery (GCD) based on IFS over multiple universes on big data repositories with evolving data models due to changes in the data domain, paves the way to the next question "what are the implications for Big Data Quality Metrics? What their interpretation will be then?

Additional research will be required to identify effective mitigation measures for the training and testing of machine learning techniques for data streams and/or into improved sampling and resampling techniques for data streams that should, ideally, avoid the issue of sampling induced concept drift altogether or, at the very least, minimize it.

References

[1] Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1), 87–96.

[2] Atanassov, K. (1992). Remark on the intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 51(1), 117–118.

[3] Atanassov K. (1999). *Intuitionistic Fuzzy Sets: Theory and Applications.* Springer, Heidelberg, 1999.

[4] Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2018). Deep clustering for unsupervised learning of visual features. *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132–149, 2018.

[5] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1–58.

[6] Chountas, P. (2007). Representation of null values with the Aid H-IFS. *Notes on Intuitionistic Fuzzy Sets*, 13(1), 20–33.

[7] Chountas, P., & Atanassov, K. (2007). On intuitionistic fuzzy sets over universes with hierarchical structures. *Notes on Intuitionistic Fuzzy Sets*, 13(1), 52–56.

[8] Geng, C., Huang, S.-J., & Chen, S. (2020). Recent advances in open set recognition: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10), 3614–3631.

[9] Han, K., Rebuffi, S.-A., Ehrhardt, S., Vedaldi, A., & Zisserman, A. (2021). Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), 6767–6781.

[10] Linnebo, Ø. (2010). Pluralities and sets. *Journal of Philosophy*, 107(3), 144–164.

[11] Linnebo, Ø. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, 6(2), 205–228.

[12] Pang, G., Shen, C., Cao, L., & Van Den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2), 1–38.

[13] Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., & Boult, T. E. (2022). Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7), 1757–1772, 2012.

[14] Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. *Proceedings of International Conference on Machine Learning*, 478–487. PMLR.