A Binary Hyperbox Classifier Model for Hydrogen Storage in Metal Hydrides

K Anthea Rana, John Andrei Acantilado, Jared Ethan Santos, Raymond R. Tan, Kathleen B. Aviso*

Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines
 kathleen.aviso@dlsu.edu.ph

Despite being recognized as a key component towards reducing global GHG emissions, many challenges remain throughout the hydrogen value chain. Hydrogen storage is one critical aspect, since proper storage is essential for its wide scale deployment. One storage option is the chemisorption of hydrogen in metals and intermetallic alloys (i.e., metal hydrides). This approach can potentially address issues on safety, infrastructure, and cost which currently hinder the transition toward a hydrogen economy. The only suitable method for determining appropriate hydrogen storage materials is through experimentation which is resource intensive. There are many parameters that can affect the hydrogen storage capacity of a material; databases on previously investigated materials can be used to narrow down future investigation to the most promising candidates. Machine learning (ML) techniques can be employed to determine how different properties predict for hydrogen storage capacity. ML can use previously compiled data to generate a classification model that can be utilized for determining and predicting a material's viability for hydrogen storage. This paper uses the hyperbox ML technique to generate interpretable decision rules to predict if a metal hydride is a good candidate for hydrogen storage. A case study which specifically focuses on complex and magnesium hydrides is used to demonstrate this approach. The generated decision model had a false positive rate of 22.0 % and false negative rate of 36.8 %.

1. Introduction

Global warming and climate change is expected to have a drastic effect on Earth, as witnessed by the rate of increase in sea levels. Greenhouse gas (GHG) emissions from energy systems are major contributors to climate change. Furthermore, recent data has also confirmed that majority of the global anthropogenic GHG emissions are due to fossil fuel use (IPCC, 2019). There are different options available for reducing GHG emissions from energy systems. These can range from simple tasks such as energy conservation up to the difficult ones such as carbon capture; between these two options exists hydrogen (Mukhopadhyay, 2018). Although hydrogen is not an actual energy source, it is considered an energy carrier or vector. The energy is stored through the chemical bonding of hydrogen, which can be released via combustion or reaction with oxygen through the use of fuel cells. This concept serves as the basis of the hydrogen economy, shifting from the dependence on fossil-fuels towards alternative and cleaner sources (Falcone et al., 2021). However, there are challenges present that make a hydrogen fuel economy difficult to pursue. Hydrogen by nature is highly combustible in the presence of oxygen, which is one of the major safety issues which is currently associated with hydrogen applications. Another is the storage of hydrogen due to its low energy density and low temperature requirement for cryogenic systems (Dincer and Siddiqui, 2020). These challenges, however, may be solved with the application of metal hydride materials for hydrogen storage.

A metal hydride is a hydrogen compound that is formed with hydrogen and metals. These compounds have unique properties which make them attractive for storage of hydrogen within a crystal lattice (Young, 2018). The focal point of a research on metal hydride was on further improving the volumetric and gravimetric capacities, hydrogen absorption and desorption kinetics, and reaction thermodynamics of potential material candidates (von Colbe et al., 2019). The viability of these materials, however, is difficult to predict and are only obtained

through experimentation. Machine learning (ML) is a potential solution that can generate predictive models with the use of data collected from previous studies. There are several dimensions to determining the material's capabilities for hydrogen storage. With the utilization of ML algorithms, rapid data predictions are now made possible, this makes it easier to identify patterns which were previously difficult to recognize. (Rahnama et al., 2019). In ML, models can typically be described as 'black boxes', with models which are difficult to interpret. In high-risk applications, this lack of understanding may lead to serious issues which are not easily addressed (Rudin, 2019). Interpretable ML can be used instead to generate transparent models to predict the suitability of metal hydrides to hydrogen storage applications. An example of such is the use of hyperbox-based ML.

The ML algorithm used in this study was developed by Tan et al. (2020) and addresses the interpretability problem. It is based on the hyperbox framework developed by Xu and Papageorgiou (2009) and generates easily understandable set of rules calibrated using training data. The model used is considered as a rule-based classifier which results in a set of disjunctive if/then statements formed from the hyperboxes. The hyperbox-based model is a mixed integer linear programming (MILP) model which can also generate alternative near-optimal rule-based classifiers via integer cuts. A recent search on the Scopus database as of July 2021, using keywords 'hydrogen storage', 'metal hydrides', and 'machine learning' found only 9 studies on this topic. Although this does show there has been some research done, this is a very small number, and much more studies can be done in order to contribute to this body of knowledge. Previous works have not explored the use of hyperbox-based ML in predicting material performance for hydrogen storage. This study applies a novel hyperbox-based ML technique in the determination of the relationship between the hydrogen capacity of a hydride to its different properties, particularly: its heat of formation, temperature, and pressure. Rules formed from these properties are more easily understandable and can then be used to predict the viability of a material's hydrogen capacity. The remainder of this paper is arranged as follows: section 2 presents the formal problem statement that is addressed by this study, section 3 shows the MILP model, section 4 outlines the approach undertaken in the study and the result, and section 5 gives the conclusion and the study's potential for future research.

2. Problem Statement

The formal problem statement is as follows:
- Given a set of criteria I (i.e. the heat of formation, temperature, and pressure) which characterize metal hydrides;
- Given a binary decision set D (i.e. the hydrogen weight percentage);
- Given a set of sample J which have known performance in the defined set of criteria I and known classification in decision set D;
- Given a known target for predictions resulting in false negatives;
- Given the minimum distance between the performance of positive and negative samples in a given criterion;
- The objective is to determine the boundaries of the hyperbox which can adequately classify the performance of a metal hydride in terms of hydrogen storage capacity by minimizing the number of false positive results in the training data set.

This model assumes that the only parameters that affect the hydrogen capacity are the heat of formation, temperature, and pressure. These properties are limited to the thermodynamic data of each hydride and provide no information on the kinetics of the reaction. It was also assumed that the hydrogen capacity of the hydride is fully reversible and can be used to store and release energy consistently. The problem is to obtain a model with realistic parameters that assist future research on the viability of a metal hydride's hydrogen capacity. This study develops a rule-based classifier model for metal hydrides in hydrogen storage applications by finding a relationship between the hydrogen storage capacity to the different properties of the hydride. The main goal of the research is to enable future studies to apply the findings from this paper into the applications of hydrogen storage and hydrogen energy. With the discovery of a viable solution to hydrogen storage, it brings the world closer to more sustainable and cleaner energy.

3. MILP for Generating Hyperbox Decision Model

The model used in this study focused on optimizing for type I errors (false positives), as seen in Eq(1); while setting ε, to represent the threshold for Type II errors (false negatives), shown in Eq(2). The proportion for Type I, α, and Type II errors, β, are seen in Eq(3) and Eq(4) respectively, where c_j is the predicted classification of sample j, C_j^* is the true classification of the sample, and N^T and P^T are the total negative and positive samples.

$$\min \alpha \qquad\qquad (1)$$

$$\beta \leq \varepsilon \qquad\qquad (2)$$

$$\alpha = \frac{\sum_j (c_j - C_j^*)}{N^T} \qquad\qquad \forall\, j \in N^T \quad (3)$$

$$\beta = \frac{\sum_j (C_j^* - c_j)}{P^T} \qquad\qquad \forall\, j \in P^T \quad (4)$$

The algorithm uses a bi-objective MILP to determine the dimensions of each hyperbox while ensuring that it is correctly classifying the given samples. Each sample, j, has a performance variable in the dimension (parameter) i, X_{ji}. Eq(5) and Eq(6) create the outer boundaries of hyperbox k to avoid incorrect classification of samples at the border of the hyperboxes, while Eq(7) and Eq(8) create the inner boundaries of hyperbox k. The binary output variable, b_{jk}, indicates whether sample j is contained within hyperbox k by having a value of 1 and a value of 0 if not. Eq(9) and Eq(10) determine the lower bound, x_{ik}^L, and upper bound, x_{ik}^U, of hyperbox k in the dimension (parameter) i. In these equations, Z_{ik}^U and Z_{ik}^L represent the highest and lowest possible values of the dimension (parameter) i for hyperbox k, b_{ik}^U and b_{ik}^L indicate whether or not hyperbox k has an upper boundary and lower boundary respectively, and M is an arbitrary large number.

$$X_{ji} > x_{ik}^L - \Delta - M(1 - b_{jk}) \qquad\qquad \forall\, i, j \quad (5)$$

$$X_{ji} < x_{ik}^U + \Delta + M(1 - b_{jk}) \qquad\qquad \forall\, i, j \quad (6)$$

$$X_{ji} > x_{ik}^L - M(1 - b_{jk}) \qquad\qquad \forall\, i, j \quad (7)$$

$$X_{ji} < x_{ik}^U + M(1 - b_{jk}) \qquad\qquad \forall\, i, j \quad (8)$$

$$Z_{ik}^L - M(1 - b_{jk}^L) \leq x_{ik}^L \leq Z_{ik}^L + M b_{ik}^L \qquad\qquad \forall\, i, k \quad (9)$$

$$Z_{ik}^U - M b_{ik}^U \leq x_{ik}^U \leq Z_{ik}^U + M(1 - b_{jk}^U) \qquad\qquad \forall\, i, k \quad (10)$$

Eq(11) and Eq(12) determine whether the sample j lies within hyperbox k in the dimension (parameter) i. The variables q_{ik}^L and q_{ik}^U are binary variables that activate if the sample j in the dimension (parameter) i is below or above the lower and upper limits respectively. In essence, $q_{ik}^L = 1$ when it is less than x_{ik}^L, and $q_{ik}^U = 1$ when it is greater than x_{ik}^U. Therefore, if either q_{ik}^L or q_{ik}^U is equal to 1, then sample j does not lie within hyperbox k implying that $b_{jk} = 0$. This is shown in Eq(13) and Eq(14). A sample j is considered a positive sample ($c_j = 1$) if it belongs in at least one of the hyperboxes formed as described in equation Eq(15). Eq(16) is included in order to tighten the constraint. Finally, Eq(17) lists all binary variables used in the algorithm.

$$X_{ji} \leq x_{ik}^L - \Delta + M(1 - q_{ijk}^L) \qquad\qquad \forall\, i, j \quad (11)$$

$$X_{ji} \geq x_{ik}^U + \Delta - M(1 - q_{ijk}^U) \qquad\qquad \forall\, i, j \quad (12)$$

$$\sum_i q_{ijk}^L + q_{ijk}^U \leq M(1 - b_{jk}) \qquad\qquad \forall\, i, k \quad (13)$$

$$\sum_i q_{ijk}^L + q_{ijk}^U \geq (1 - b_{jk}) \qquad\qquad \forall\, i, k \quad (14)$$

$$\sum_k b_{jk} \leq M c_j \qquad\qquad \forall\, i, k \quad (15)$$

$$\sum_k b_{jk} \leq c_j \qquad\qquad \forall\, i, k \quad (16)$$

$$b_{jk}, b_{jk}^U, b_{jk}^L, q_{jk}^U, q_{ijk}^L, C_j \in 0,1 \qquad\qquad (17)$$

4. Case Study

The properties of hydrides were taken from an online open access database provided by the US Department of Energy (DOE, n.d.), which contains 2,722 entries. The input variables are the heat of formation, pressure (at 25 °C), temperature (at 101.3 kPa), and the output variable is the hydrogen weight percentage. The dataset was first narrowed down to 94 datapoints which consisted of complex and Mg hydride samples. Of these samples, 25 datapoints are taken as training data. The hydrogen weight percentage is assigned a binary target value using a cut-off value of 4.5 %. The input variables were also normalized within the interval 0 to 1. The determination of the optimal model is dependent on the combination, arrangements of the 25 datapoints from the database, the number of hyperboxes created, threshold for proportion of Type II errors, and the boundary distance between the inner and outer border of said hyperboxes. The threshold was limited to values below 0.4 and the boundary distances were kept constant to a value of 0.05. The classifier model was then validated by testing it against the 69 datapoints, determining which samples are positive or negative based on the generated rules formed. K-fold validation was implemented by running various combinations of the datasets. Once validated, the best model in terms of Type I and II error rates is selected. Lastly, the rules formed must be studied if they are consistent with the laws of chemistry and physics. At the end of the experimentation, a total of approximately 600 datasets were run.

As a result of the repeated testing, the best classifier model consisted of the two hyperboxes shown in Table 1. This model had a performance of 15.8 % false positive and 33.3 % false negative during training and a performance of 22.0 % false positive and 36.8 % false negative rate when checked against the validation data of 69 samples.

The results from Table 1 indicate the lower and upper limit of the hyperboxes which enclose samples with a hydrogen weight percentage $\geq$ 4.5 % as determined by the model in Eq(1) to Eq(17). These can be translated into rules as follows:

Rule A: IF (Heat of Formation ≥ 75.5 kJ/mol) AND (337 °C ≥ Temperature > 274 °C) AND (Pressure ≥ 0.14 MPa)) THEN (Hydrogen Weight Percentage ≥ 4.5 %)

Rule B: IF (Pressure ≥ 1.78 MPa) THEN (Hydrogen Weight Percentage ≥ 4.5 %)

Table 1: Lower (x_{ik}^L) and Upper (x_{ik}^U) Limits of Hyperboxes from the Best Performing Model

HYPERBOX	BOX A		BOX B	
Limit	Lower	Upper	Lower	Upper
Heat of Formation (kJ/mol)	75.5	-	-	-
Temperature (°C)	274	337	-	-
Pressure (MPa)	0.14	-	1.78	-

Figures 2 and 3 show the graphical representation of these hyperboxes. The red shaded region represents Box A, the blue shaded region represents Box B, and the green and red coloured points represent the positive and negative samples respectively.
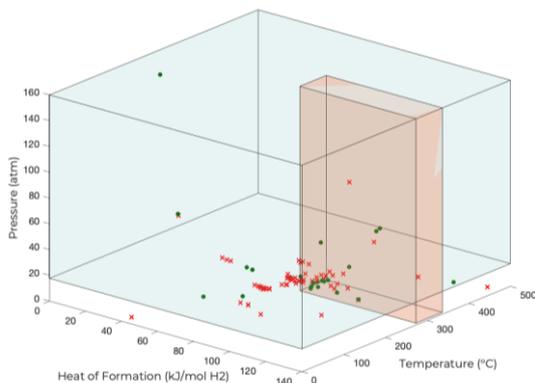

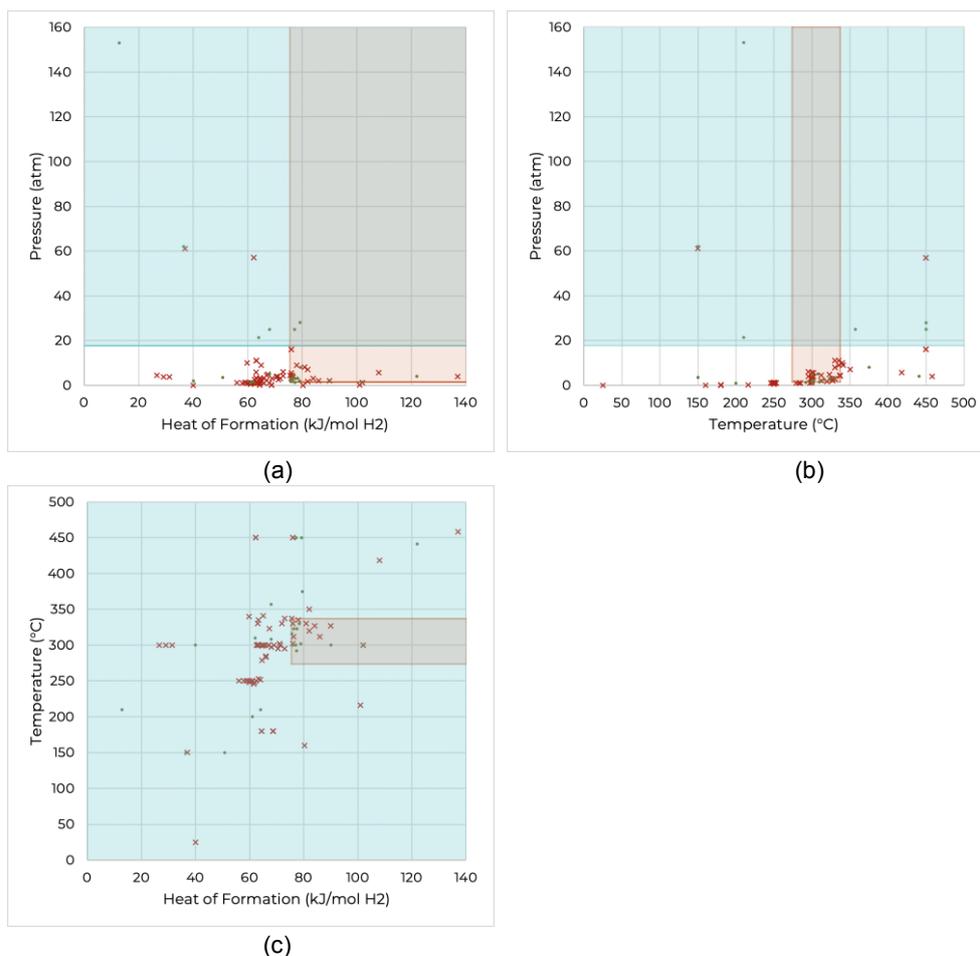Figure 2: Hyperboxes with respect to heat of formation, pressure, and temperature

Figure 3: (a) two-dimensional projection showing pressure and heat of formation (b) two-dimensional projection showing pressure and temperature (c) two-dimensional projection showing temperature and heat of formation

It can be seen that the majority of the points are concentrated at lower pressures. With this, the conditions of Rule A are more precise and aim to capture the desired datapoints at lower pressures. Rule B on the other hand is unable to capture a majority of the points focusing on only capturing all higher-pressure hydrides. This result implies that higher pressure hydrides are better hydrogen storage materials.

Andreasen (2004) stated that at 1 bar, the Mg hydrides must be heated to around 280 °C in order to release the hydrogen from the compound. Additionally, the magnesium at this state is said to be at its most stable, with a heat of formation of approximately 75 kJ/mol H_2. These values are consistent with the rules generated from the data, specifically the lower limits for the heat of formation and temperature for Rule A. Zhang et al. (2019) discussed how the crystalline structure of pure magnesium requires a temperature range of 300 to 400 °C in order to react with hydrogen, relating to the upper limit for temperature in Box A. With regards to the rules for Box B, the same study by Zhang et al. (2019) provides more information regarding high pressure Mg hydrides. It was stated that these hydrides are able to exist at different phases (crystal lattice) at different pressures; these higher-pressure Mg hydrides exist in the beta or gamma phase. Zhou et al. (2015) discussed how the beta and gamma phases of these hydrides are able to have a theoretical maximum hydrogen storage capacity of 7.6 wt %. These findings indicate that the positive hydrides captured by Rule B are these high pressure, high capacity, beta-Mg hydrides or gamma-Mg hydrides.

5. Conclusions

This study has developed a rule-based classifier model for predicting hydrogen storage capacity of Mg and complex hydrides from heat of formation, temperature, and pressure. The classifier was generated using the hyperbox ML approach. A training set consisting of 25 datapoints and a validation set of 69 datapoints were used. The best performing model produced two hyperboxes. Box A had: for the heat of formation a lower limit of 75.5 kJ/mol with no upper limit, for the temperature, a lower limit of 274 °C and an upper limit of 337 °C, and

for the pressure, a lower limit of 1.4 MPa without an upper limit. Box B, on the other hand, consisted of only one constraint which was for pressure with a lower limit value of 17.6 MPa. The rules generated are consistent with the laws of physics and chemistry, and have good predictive and explanatory power resulting in 22.0 % false positives and 36.8 % false negatives with the validation data set. The rules formed can then be used to inform future researchers regarding the hydrogen capacity of a given hydride, giving them guidance on which materials to pursue further research on.

Future research can apply the classifier from this study for the analysis of possible hydrides for hydrogen storage, or they may even adapt the algorithm to other green engineering systems. There are still improvements that can be made to the algorithm in order to improve its performance. One promising direction is to integrate an automated system that is able to randomize the selection of training sets and parameter settings, run the algorithm, and record the data, all without human supervision.

References

Andreasen A., 2004, Predicting formation enthalpies of metal hydrides, Risø National Laboratory, Roskilde, Denmark

Dincer I., Siddiqui O., 2020, Types of fuels, Chapter In: In Ammonia Fuel Cells, Elsevier Inc., Amsterdam, Netherlands.

Falcone P.M., Hiete M., Sapio A., 2021, Hydrogen economy and sustainable development goals: Review and policy insights, Current Opinion in Green and Sustainable Chemistry, 31, 100506.

Hui-Min L.I., Xue-Chun W.A.N.G., Xiao-Fan Z.H.A.O., Ye Q.I., 2021, Understanding systematic risk induced by climate change, Advances in Climate Change Research <DOI:10.1016/j.accre.2021.05.006> accessed 10.06.2021.

IPCC, 2018, An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)] <www.ipcc.ch/sr15/> accessed 10.06.2021.

Mukhopadhyay R., Karisiddaiah S.M., Mukhopadhyay J., 2018, Threat to opportunity, Climate Change: Alternate Governance Policy for South Asia, 99-117.

Rahnama A., Zepon G., Sridhar S., 2019a, Machine learning based prediction of metal hydrides for hydrogen storage, part I: Prediction of hydrogen weight percent, International Journal of Hydrogen Energy, 44, 7337–7344.

Rahnama A., Zepon G., Sridhar S, 2019b, Machine learning based prediction of metal hydrides for hydrogen storage, part II: Prediction of material class, International Journal of Hydrogen Energy, 44, 7345–7353.

Rudin C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence, 1, 206–215.

Tan R.R., Aviso K.B., Janairo J.I.B., Promentilla M.A.B., 2020, A hyperbox classifier model for identifying secure carbon dioxide reservoirs, Journal of Cleaner Production, 272, Article 122181

US Department of Energy, Hydrogen Storage Materials Database, < hydrogenmaterialssearch.govtools.us> accessed 10.06.2020.

von Colbe J.B., Ares J., Barale J., Baricco M., Buckley C., Capurso G., Gallandat N., Grant D.M., Guzik M.N., Jacob I., Jensen E.H., Jensen T., Jepsen J., Klassen T., Lototskyy M.V., Manickam K., Montone A., Puszkiel J., Sartori S., Sheppard D.A., Stuart A., Walker G., Webb C.J., Yang H., Yartys V., Züttel A., Dornheim M., 2019, Application of hydrides in hydrogen storage and compression: achievements, outlook and perspectives, International Journal of Hydrogen Energy, 44, 7780-7808.

Xu G., Papageorgiou L.G., 2009, A mixed integer optimisation model for data classification, Computers and Industrial Engineering, 56, 1205–1215.

Young K, 2018, Metal Hydrides, Chapter in: Chemistry, Molecular Sciences and Chemical Engineering, Elsevier Inc. <doi.org/10.1016/B978-0-12-409547-2.05894-7> accessed 10.06.2021.

Zhang J., Li Z., Wu Y., Guo X., Ye J., Yuan B., Wang S., Jiang L., 2019, Recent advances on the thermal destabilization of Mg-based hydrogen storage materials, RSC Advances, 9, 408–428.

Zhou S., Zhang Q., Chen H., Zang X., Zhou X., Wang R., Jiang X., Yang B., Jiang R., 2015, Crystalline structure, energy calculation and dehydriding thermodynamics of magnesium hydride from reactive milling, International Journal of Hydrogen Energy, 40, 11484–11490.