A method based on machine learning to help determine the availability of hydrogen energy storage system in hydropower stations

Haoxiang Xu *

Nanjing Foreign Language School, Nanjing, China

* Corresponding author: steven950818@my.yosemite.edu

Abstract. Nowadays, the hydrogen energy storage technologies are being developed at a constant speed. The application of this technology in hydropower stations is now being discussed. To determine whether this technology can be used in real hydropower stations to store electrical energy in order to deal with the change of wet seasons and dry seasons in rivers, a function based on machine learning can be used. The program, written in python, uses '0' and '1' to determine whether the system is available. Sets of data of hydropower stations are used to train the model to make it predict the result. The passage introduces the hydrogen energy storage systems, the principle of machine learning and the structure of code.

Keywords: machine learning, hydrogen energy storage technologies, hydropower stations, big data.

1. Introduction

Along with the development of modern industry and technology, the demand for energy rapidly increases. Conventional energy sources such as petroleum, coal and natural gas contribute to polluting the environment by releasing air pollutants or greenhouse gases. Thus, renewable resources like solar, wind or hydro are more suggested to be used because of its environmental friendliness. Among all renewable energy sources, hydropower serves as a stable, effective and low-cost energy generation method. Hydropower stations work as main electricity generation facilities. A water power plant is usually constructed near a river, using flowing water to push electricity generator.

In a year's loop, there are a wet season and a dry season in rivers. The change of water flow velocity and water amount results in a difference in electricity production amount in hydropower stations near rivers [1]. Meanwhile, the demand of electricity slightly varies. Thus, it is necessary to store the electrical power generated in wet seasons to fill in the demand in dry seasons. Electricity can be stored by converting into other forms of energy. There are several common ways in electricity storage, including pumped hydroelectric energy storage, battery storage or hydrogen storage. Currently, pumped hydroelectric energy storage system is the most common, while hydrogen storage is increasingly being noticed because of the decarbonization trend [2]. Hydrogen is a clean fuel and a long-term energy source option. Hydrogen storage is also one of the potential lowest-cost electricity storage methods [2]. Thus, applying hydrogen storage method to hydropower station program is a suitable option to help solve the problem in demand difference.

Hydropower stations vary because of their location, facility size, river flow and other influential factors. Hydrogen storage program is not available in every situation. To evaluate whether hydrogen storage program can be used in a particular hydropower station, people normally investigate local condition of the station, measure the size of facilities and calculate the efficiencies. However, this process would need days or months to make evaluations. In modern time, artificial intelligence is popular among new developed technologies. Also, a large number of advising fee are needed. Machine learning, an important concept in AI field, can be used properly to help make evaluations on the availability of hydrogen storage programs.

This article introduces the structure of hydrogen storage program, illustrates the basic principle of machine learning, and explain how machine learning help determine the availability of using hydrogen to store electrical power. A set of code written in Python is provided to realize machine learning.

2. The Process of Energy Storage with Hydrogen

Hydrogen storage process mainly contains three parts: generation, storage and conversion to electricity.

2.1. Hydrogen Generation

For generation part, hydropower plants usually use water electrolysis since the generation process is renewable and clean. Water electrolysis means to produce hydrogen with an electrolytic cell. An electrolytic cell consists of two electrodes, anode and cathode. A direct current is applied to form a specific difference in electric potential, known as voltage. The minimum voltage needed can be calculated with Gibbs Free energy equation and Nernst equation in this reaction. Then a redox reaction happens at both sides and water is then decomposed to oxygen and hydrogen [3].

$$H_2O \rightleftharpoons 1/2\ O_2 + H_2 \quad \Delta H = +285.9\ kJ/mol \tag{1}$$

A more common electrolyzer cell contains a proton exchange membrane (PEM), which is a well-matured electrolysis technology for hydrogen production. It has been commonly used on a commercial scale. This method operates well when using intermittent energy sources. Fig.1 shows the basic structure and principle of PEM electrolyzer cells.

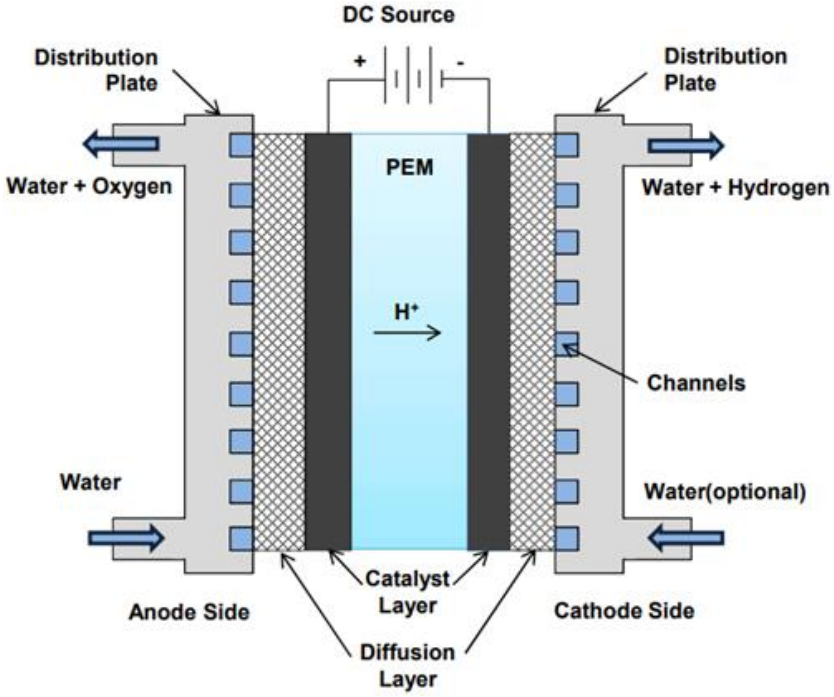


Figure 1. Structure of PEM electrolyzer cell [4]

2.2. Hydrogen Storage

Hydrogen needs to be compressed for storage because of its low density. Currently, compressed hydrogen gas storage mainly consists of two common ways: geological storage and storage vessels [5]. Geological storage means to inject hydrogen gases underground and store in spaces like saline aquifers, salt caverns or depleted natural gas and oil reservoirs. However, considering the fact that the types of landforms near hydropower stations vary, geological storage method might not be available everywhere. Storage vessels are more likely to be available and stable in various places. Storage vessels are commonly used in industrial, commercial and aerospace applications. They contain a variety of types, from small bottles to massive storage tanks. Normally, the pressure of hydrogen storage is 35 to 70 MPa, with the storage density of 23kg/m^3 to 38kg/m^3 [5].

2.3. Hydrogen‘s Conversion to Electricity

Methods of hydrogen’s conversion to electricity are mainly divided into 2 parts: using it as fuel to be burnt or using it as reductor in a cell. As fuel, hydrogen is widely used in hydrogen energy vehicles, which plays a significant role in sustainable energy applications. However, burning hydrogen is not an efficient method to produce electricity because it generates heat in the same time that remarkably reduces the efficiency. Using hydrogen as fuel cell is a more efficient method. The hydrogen fuel cell is a standard galvanic cell in which oxidizer and hydrogen are fed to the cathode and anode, respectively, often with PEM in the middle together with suitable catalysts.

3. Principles of Machine Learning

Machine learning is widely used in the development of artificial intelligence. Chat-GPT, an AI oriented language model generated by OpenAI, uses machine learning codes to generate response. Machine learning is a way for computer to predict the classification of a feature that contains a set of input. Currently, machine learning is developing rapidly because of the increasing notice of AI technologies in the world.

Machine learning is mainly separated into supervised machine learning and unsupervised machine learning [6]. The process of supervised machine learning isare as follows. First, the computer receives a large amount of database, which contains groups of sets of input and the output of the sets. Then, the computer learns the pattern of the database and forms a model by measuring the most suitable coefficients for all variables. Thus, after receiving a new set of variables, the computer is able to infer the output of the new set. The major difference of supervised machine learning and unsupervised machine learning is that in unsupervised machine learning, the computer is not given the category of the sets, which means that it is asked to find out the rule and present the interesting structure in the data [7]. After it is introduced a new set of input, it is asked to categorize the set.

One of the most common machine learning algorithms is linear regression. Independent variables are the inputs in the dataset, named “x1, x2, x3, …, xm”. The predicted output, named as y’, in linear regression is defined as follow:

$$y' = \sum_{i=0}^{m-1} w_i x_i + b \tag{2}$$

Where ‘m’ is the number of independent variables. All the ‘w’s are the coefficient in the function that determines the influence of ‘x’s to the result’s value. In order to match all the database as perfectly as possible, the values of ‘w’s and ‘b’ are continuously changing until the value of cost function reaches the minimum. Cost function is defined as follow:

$$\text{J(w, b)}=\frac{1}{2m}\sum_{i=0}^{m-1}(y' - y)^2 \tag{3}$$

Where ‘y’ is the actual value of the output for that set of data. The database should be as much as possible because with more data, the program can more precisely match suitable coefficients [6]. Moreover, other methods such as logistic regression and classification are also involved in supervised machine learning.

Obviously, the problem of whether the method of hydrogen energy storage system can be used in hydropower stations is able to be solved with supervised machine learning. The output is set to be a number between 0 to 1, which represents “not recommended to use hydrogen storage systems” or “recommended to use hydrogen storage systems”. If the data is above 0.5, then the system is more likely to be used in hydropower stations since it brings more benefits than costs. Reversely, if the output is below 0.5, it illustrates that the disadvantage is larger than the advantage.

Currently, Python is the coding language that support machine learning at a high extent. Thus, python is chosen to be the language used to encode the evaluation execution. Python environment

contains scikit-learn package that consists of a large number of algorithms that are commonly used in machine learning. Scikit-learn package is the most popular package in GitHub. Except this, package numpy and pandas are also imported.

First, the program is designed to read the database from the local absolute path of the data table. After reading, the data are normalized to the range of 0 to 1. The normalization process is listed below:

$$x = (x - x_{\min})/(x_{max} - x_{\min}) \tag{4}$$

Through this process, 'x' can be converted to a number between 0 to 1. Then, the program uses classification to group sets of data into 0 and 1. A model is trained and predicted with logistic regression. Lastly, all the coefficients of the function are printed, together with the accuracy of the model.

4. Independent Variables in the Database

In order to design a machine learning code to determine the availability of hydrogen storage program, the types of independent variables need to be determined. Water potential is converted into mechanical energy by rotating the turbine and mechanical energy is converted to electrical energy with electricity generator. The mechanical energy produced is calculated as follow [8]:

$$P = Q\rho ghη \tag{5}$$

Where Q is the flow rate of water in cubic meter per second, ρ is water density in kilograms per cubic meter, g is gravitational acceleration in meter per square second, h is the height difference, head, in meter, and η is the efficiency of the turbine [8]. Since ρ and g are certain numbers, the independent variables contain only Q, h and η. To determine the electrical power generated in the station, the efficiency of generator is also required, named as 'k'. Since the converting efficiency is normally provided on the generator, this passage is not going to distinguish different steps in the converting process.

The flow rate Q varies from season to season, and it causes the necessity to store electrical power. Thus, the data of average flow rate in low-flow periods and high-flow periods are needed, named as 'Q_L and Q_H'. Also, the number of days of low-flow days and high-flow days are necessary to determine the conversion rate, named as 'N_L and N_H'. Lastly, the data of average monthly demand of electrical power is needed, named as 'Amd'.

In order to make the problem predictable, geological storage method is not under consideration because the cost varies according to the location of the stations. The hydrogen storage part is united to be the storage vessel method. The generation, storage and conversion steps of hydrogen storage system is fixed since they all use the same facilities. Thus, no independent variables are needed to be added.

Furthermore, because of the difference in the scales of hydropower stations, the budget of the station should be considered. An independent variable named 'b' is added to involve the total amount of money that can be used to set up the hydrogen storage program. In conclusion, there are 9 independent variables to be inputted in the code in total: h, η, k, Q_L, Q_H, N_L, N_H, Amd and b. Other influential factors play a relatively slight role, so they are not under consideration.

5. The Source of Database

The lack of database is a serious problem to the program. Currently, there are few hydropower stations that have started completely using hydrogen storage technology, including Green Hydrogen Electrolysis Plant by Dubai Electricity and Water Authority (DEWA), United Arab Emirates [9]; La Romaine Hydroelectric Power Station, Quebec, Canada [10]; and Moss Landing Power Plant,

California, USA. Unfortunately, the types of data listed in part 4 of these stations are not available. Thus, the database is not enough for using machine learning to determine the availability of hydrogen power storage system.

However, database is not only the examples to evaluate the reliability of the machine learning program but also the necessary condition to train the model. Thus, several sets of data have been made up. The sets of artificial data satisfy all the following conditions: 1. The combination of data theoretically fits actual condition; 2. The independent variables vary in a relatively large scale respectively, leading to the change of dependent variable. The database is listed below in table 1.

Table 1. Data used in training the model

Set No.	x1	x2	x3	x4	x5	x6	x7	x8	x9	y
1	178	0.85	0.8	1120	7500	5	2	1875.1	2000	0.33
2	145	0.82	0.75	960	6330	6	2	1425.5	350	0.25
3	80	0.83	0.73	630	5320	5	1	320.4	40	0.79
4	75	0.8	0.65	640	5540	6	3	430.2	45	0.42
5	125	0.9	0.76	1000	6110	4	1	1103.3	60	0.38
6	60	0.85	0.69	510	4250	5	1	210.2	100	0.89
7	10	0.88	0.73	120	540	3	2	15	10	0.44
8	5	0.87	0.7	25	163	4	2	1.9	2	0.39
9	120	0.82	0.78	1020	5960	3	3	1095.6	65	0.62
10	170	0.8	0.8	920	8100	5	2	2021.2	2000	0.41

In the above table, x1 to x9 represents h (meter), η, k, Q_L (m^3/s), Q_H (m^3/s), N_L (months/year), N_H (months/year), Amd (Megawatt) and b (ten thousand dollars) respectively. ‘y’ is the score that evaluates the availability of hydrogen storage system. The more it is close to 1, the more system is considered to be effective.

The result is formed after fitting the data to the model. Since the model has nine dimensions, it is hard to use graphs to demonstrate the results. Because a small change can greatly affect the model and the coefficients, these numbers are not reliable in real conditions since the data are artificial. The application of hydrogen storage programs is increasingly globally. Many power plants are exploring the use of hydrogen storage technology to utilize renewable energy. Along with this process, the database will become more amplified, which improve the reliability of the code.

6. Reliability of the Program

In order to test the reliability of the program, a dataset of a particular hydropower station needs to be used. The data of the hydropower station is substitute into the function with all the coefficients printed.

The data chosen for testing should be easy to judge the class it belongs. Thus, I chose a relatively extreme set. The data is based on an extremely large-scale hydropower station, which has a height difference of 160 meters. The data set is [160, 0.88, 0.8, 1240, 5810, 5, 2, 1351.2, 400]. For a large-scale hydropower station, it is hard to apply hydrogen energy storage technologies because the space needed for storing vessels is too large and the cost is too high. Using coefficients got from the model, the value of ‘y’ is 0.13, perfectly fit the results in actual situation.

This test result proves that: 1. The code runs well and the coefficients in the model are reasonable; 2. The data used to train the model are basically suitable. Thus, after inputting enough real data, this code is definitely reliable to help make decisions.

7. Conclusion

In this passage, a method based on supervised machine learning has been suggested to help hydropower stations to decide whether to apply hydrogen power storage systems to solve the problem of different production rate of electricity. This method binds artificial intelligence algorithms with

problems in renewable energy fields. Different types of independent variables of conditions in stations are listed. After training artificial data to the program, it computes the most suitable parameters to form a function without compiling error. Then, an example of real situation is inputted, the output result fit the actual situation. However, the reliability still remains a problem because the sets of data fed are not in real conditions. In the future, after the successful applications of hydrogen power storage systems, the data of the applied stations can be used to train the model. Thus, the model would be more reliable and is able to be used by other stations to help make decisions. After the development and improvement of hydrogen energy storage technology, this code will be a time-saving and less money-consuming way for decision makers in hydropower stations to decide whether to apply hydrogen energy storage systems.

References

[1] Chang, J., Li, Y., Yuan, M., & Wang, Y. Efficiency evaluation of hydropower station operation: A case study of Longyangxia station in the Yellow River, China. Energy, 2017, 135, 23 - 31.

[2] Rehman, S., Al-Hadhrami, L. M., & Alam, M. M. Pumped hydro energy storage system: A technological review. Renewable and Sustainable Energy Reviews, 2015, 44, 586 - 598.

[3] O. Haill, "Green hydrogen versus blue hydrogen (and turquoise and grey hydrogen) - an explainer," Proactive investors NA, Sep. 22, 2021. https://www.proactiveinvestors.com/companies/news/961002/green-hydrogen-versus-blue-hydrogen-and-turquoise-and-grey-hydrogen-an-explainer-961002.html?region=ca (accessed Jul. 19, 2023).

[4] Sood, S., Prakash, O., Boukerdja, M., Dieulot, J. Y., Ould-Bouamama, B., Bressel, M., & Gehin, A. L. Generic dynamical model of PEM electrolyser under intermittent sources. Energies, 2020, 13 (24), 65 - 56.

[5] Elberry, A. M., Thakur, J., Santasalo-Aarnio, A., & Larmi, M. Large-scale compressed hydrogen storage as part of renewable electricity storage systems. International journal of hydrogen energy, 2021, 46 (29), 15671 - 15690.

[6] Jordan, M. I., & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. Science, 2015, 349 (6245), 255 - 260.

[7] Mahesh, B. Machine learning algorithms-a review. International Journal of Science and Research (IJSR). [Internet], 2020, 9 (1), 381 - 386.

[8] Singh, V. K., & Singal, S. K. Operation of hydro power plants-a review. Renewable and Sustainable Energy Reviews, 2017, 69, 610 - 619.

[9] Razi, F., & Dincer, I. Renewable energy development and hydrogen economy in MENA region: A review. Renewable and Sustainable Energy Reviews, 2022, 168, 112763.

[10] Karanasios, K., & Parker, P. Recent developments in renewable energy in remote aboriginal communities, Quebec, Canada. Papers in Canadian Economic Development, 2017, 16, 98 - 108.