Machine Learning in Automating Carbon Sequestration Site Assessment

Jay Chen, Ligang Lu, Mohamed Sidahmed, Taixu Bai, Ilyana Folmar, Puneet Seth,
Manoj Sarfare, Duane Mikulencak, Ihab Akil

Shell International E&P Inc.

Abstract

Carbon capture and sequestration are viewed as an indispensable component to achieve the Paris Agreement climate goal, i.e, keep the global warming within 2 degrees celsius from pre-industrial levels. Once captured, most CO_2 needs to be stored securely for at least decades, preferably in deep underground geological formations. It is economic to inject and store CO_2 near/around a depleted gas/oil reservoir or well, where a geological trap for CO_2 with good sealing properties and some minimum infrastructure exist. In this proposal, with our preliminary work, it is shown that Machine Learning tools like Optical Character Recognition and Natural Language Processing can aid in screening and selection of injection sites for CO_2 storage, facilitate identification of possible CO_2 leakage paths in the subsurface, and assist in locating a depleted gas/oil well suitable for CO_2 injection and long term storage. The automated process based on ML tools can also drastically decrease the decision making cycle time in the site selection and assessment phase by reducing human effort. In the longer term, we expect ML tools like Deep Neural Network can also be utilized in CO_2 storage monitoring, injection optimization etc. By injecting CO_2 into a trapping geological underground formation in a safe and sustainable manner, the oil and gas industry can contribute substantially in reducing global warming and achieving the goals of the Paris Agreement by the end of this century.

1 Introduction

According to Intergovernmental Panel on Climate Change's Fifth Assessment Report [1] , 14 percent of the global greenhouse gas emissions reductions needed by 2050 can be achieved through carbon capture. Specifically in the industrial sector (fuel burning power plants, Oil & Gas refinery, etc.), carbon capture is the only viable way to achieve full decarbonization. Various carbon capture technology (physical, membrane and Cryogenic, etc.) are under active development [2]. Once captured, the carbon dioxide can be put to productive use in enhanced oil recovery, manufacturing fuels, and building materials in limited quantities. Most of captured CO_2 will have to be stored in underground geological formations such as a structural trap.

As shown in Fig. 1, the geological structure trapping mechanism is exactly the same as for gas and oil reservoirs. Therefore, the oil & gas industry is well positioned to handle this CO_2 sequestration challenge, based on years of experience in hydrocarbon exploration/extraction/injection from various underground geological formations (oil/gas trapping). With the possibility of implementation of carbon tax regulations in the near term, the economic incentives for the oil and gas industry to work on carbon sequestration projects will increase.

Tackling Climate Change with Machine Learning workshop at NeurIPS 2021

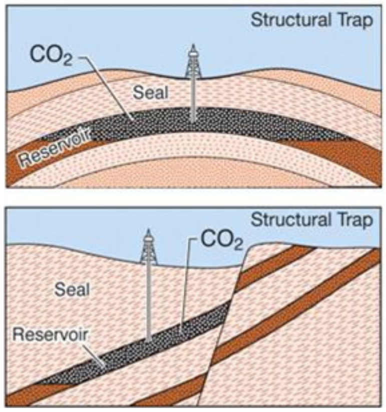


Figure 1: illustration showing structural traps for storing CO_2 in underground geological formations. Gas and oil reservoirs share exact same trapping mechanisms. Image is retrieved from DOE Carbon Storage Program [3].

Carbon sequestration is generally separated into four pillars: Capacity, Transport & Injection, Containment, and finally Monitoring & Corrective Measures. Capacity is the very first step which evaluates the storage capacity at a certain sequestration site, depending primarily on the geological formation at the sequestration site. Five types of underground formations for geologic carbon storage are currently under active investigation: saline formations, oil and natural gas reservoirs, un-mineable coal seams, organic-rich shales, and basalt formations [3]. Among these options, depleted oil/gas reservoir is the most economically viable sites to store CO_2, provided the depleted reservoirs have suitable properties: a geological trap with good seal to avoid leakage in long term. Therefore, choosing a proper site for carbon storage is crucial in realizing both the near and long term goal of carbon sequestration. To do this, geologists have to manually read through hundreds (sometimes thousands) of drilling & completion reports, and abandonment status reports (some can be dated back to late 1930s and written by hand) and extract relevant information in the Region of Interest (ROI), then summarize and synthesize all the findings and evaluate the over-all quality of storage potential for each CO_2 injection site. Then an estimation of potential storage capacity can be made for this ROI, followed by a feasibility study and decision to either further develop the ROI or abandon it if it is not practical.

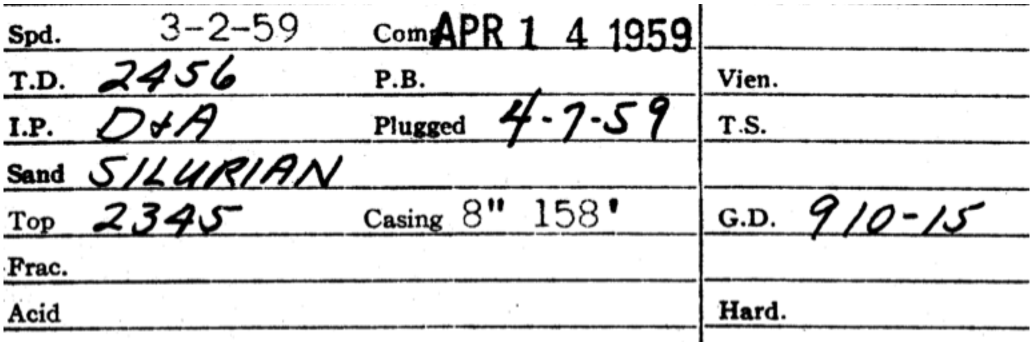
Machine Learning (ML) technologies like Boosted Decision Tree [4], and Deep Neural Network (DNN) [5] have been used in various O&G applications, such as rock type determination [6, 7], geophysical work flows like well tie [8] or salt picking [9], reservoir property analysis [10], and field offset well pressure data analysis [11]. In this study, we propose to apply Optical Character Recognition (OCR) [12] and Natural Language Processing (NLP) [13] ML tools to the carbon sequestration problem, specifically, in the sequestration site assessment process. The goal for OCR is to automate the reading process for hundreds of well reports, and automatically extract textual information from those well logs. Following this, NLP will be used to tokenize and extract relevant informations like abundant status, casing, cementing and plug information. Together with geological and geo-mechanical criteria obtained from Subject Matter Experts (SME), an automatic feasibility matrix of field development can be made by these ML tools. In short, the goal is to develop an automatic recommendation system with the help of ML to facilitate site selection and field development of CO_2 sequestration.

2 Preliminary Work

The process of injection site assessment for CO_2 storage requires a variety of information to be collected, analyzed and synthesized a priori. This normally includes daily drilling reports, mud log, completion reports, scout reports, etc. Formats of these files can include png images, pdfs, csv, data base output, excel spreadsheet, etc. For standard csv, pdf, and spreadsheet formats, we use the standard python packages to read the information in. However, old hand-written or typewriter typed reports (some that date back to 1930s) scanned to images and pdfs can be challenging to extract information. To this end, we utilize the state of art google Tesseract 4.1.1 OCR engine [12] to automatically recognize English and number characters in various well logs & reports. Keywords filtering is then applied to these outputs to locate the matches to an expected list of words generated by SMEs. Some example key words are "cemented", "casing", and "abandoned". These are important because if we see such words in a well report, it's highly indicative the well is probably a depleted gas or oil field, thus a possible candidate for CO_2 sequestration. Some other keywords like "porosity" and "permeability" will also be critical as they are related to rock properties which are very necessary to evaluate the geological trapping capacity and top-seal effectiveness.

As a pilot test, we developed a preliminary tool which converts scanned images (in PDF format) to Python Imaging Library (PIL) images, while each PIL image represents one page of the PDF document.The PIL image is then fed to the Tesseract OCR engine through $pytesseract$ interface. OCR returns the recognized words line by line. Keyword filtering is applied here to keep only the line containing at least one of the expected keywords. Input PDF file name, page number are also tracked and finally are put into a pandas DataFrame together with identified strings. This allows SME to quickly know which file and where the keyword is located, so that a quick quality check and validation can be performed.

During the pilot test, we gathered 122 well log/drilling report files from a region of interest in a North American field. These well reports have been studied by SMEs for quite some time and all the keywords in these files have been thoroughly identified and validated by SMEs. The scanned files also include 34 files which contain hand-written words, which are harder for OCR to recognize. An example image is shown in Fig. 2, where both typewriter typed words and hand-written words are present in a year 1959 report.



Spd.	3-2-59	Com	APR 1 4 1959		
T.D.	2456	P.B.		Vien.	
I.P.	D&A	Plugged	4-7-59	T.S.	
Sand	SILURIAN				
Top	2345	Casing	8" 158'	G.D.	910-15
Frac.					
Acid				Hard.	

Figure 2: A small part of an input well log scanned image, with both typewriter typed words and hand-written words.

For the example in Fig. 2, the tool correctly identified the key word $Casing$ and its corresponding contextual information: casing size value $8"158'$.

In total, the tool was able to locate keywords in 61 files. Among those 34 files containing hand-written words, the tool was able to identify 13 of them containing desired keywords. The rest $\sim$70 files indeed did not contain any keywords as also confirmed by the SMEs. The overall successful key words identification rate is 80% as reported by the SMEs who have the correct keyword list, which the tool never has access to. This result is rather encouraging, because it only took 10 minutes to identify these many keywords by the tool, compared to days of work by SMEs to read through these files page by page, file by file.

3 Proposed Future Work

Based on the encouraging pilot result, we proposed to do the following in the immediate near future:

3.1 Accuracy Improvement on Keyword Identification

This will be important especially for old hand-written reports. We will preprocess OCR image input to recognize more hand-written words, Techniques like Binarisation, Image Inverting, Erosion and Dilation will definitely be tested. We will also gather more variations of inputs. As retraining the OCR model from scratch is not feasible for this work, we can utilize transfer learning by only retraining the top several layers of the neural network model with our limited training samples.

To recognize more keywords, we can also apply the standard NLP [13] processing flows like punctuation removing, tokenization, stopwords removal, and lemmatizing.

3.2 Geological Information for the Recommendation System

As mentioned in the introduction, our goal is to provide a system which recommends whether a ROI is suitable for CO_2 storage. Therefore, keywords identification is only the first step. Additionally, we will need a text mining process to associate the numerical values to the identified keywords.

This is important as numerical value for certain keywords are always referred to a well and/or geological related information, thus critical in CO_2 trap assessment. For example, for keyword $plug$, the depth related information is important. If the plug is too shallow, meaning the reservoir is too shallow (the depth is less than critical depth) to keep the CO_2 in liquid state, which won't be ideal for CO_2 storage. As CO_2 in gaseous state occupies a larger volume in the reservoir (see Fig. 3), this will result in a diminished storage capacity for the sequestration site.

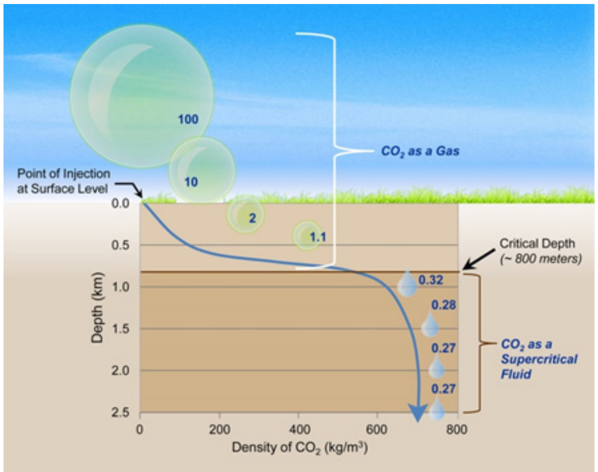

100
10
2
1.1
Point of Injection at Surface Level
CO_2 as a Gas
Critical Depth (~ 800 meters)
0.32
0.28
0.27
0.27
CO_2 as a Supercritical Fluid
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
0
200
400
600
800
Density of CO_2 (kg/m^3)

Figure 3: Pressure effect on CO_2, with increasing depth. Image is retrieved from DOE Carbon Storage Program [3].

Numerical value extractions can be done using heuristics methods like regex and distance measure to the keyword. For instance, in the Fig.. 2 example, we need to associate the $8"158'$ to possibly a "casing size" attribute. Once value extraction is done, we can construct an overall recommendation score for each storage site, based on available keyword information (well and geological information). If the result is positive, a carbon sequestration site development decision might be made.

3.3 Tool Development for more Usability

We want to make the assessment process independent of how many opportunities/ROIs we have. Ideally, this tool should allow screening ROIs in real time, so that we can make the CO_2 storage development decision faster. Currently, on average, it takes 6 months to identify a suitable CO_2 injection site. With the help of the proposed tool, we expect the site assessment process to be completed in less than one month. To do this, we need to extend the preliminary tool to a docker container, so anyone can run the tool easily with his/er own ROI input files. Furthermore, we can deploy the container to cloud as a web-based application that enables screening multiple ROIs in parallel, supported by a backend Kubernetes cluster, saving even more time for more SMEs.

OCR and NLP are the most critical parts of the proposed work, because they are the very first steps for information extraction. The extraction completeness directly determines how much data coverage we can have as input to the recommendation logic. Therefore, OCR and NLP have the largest impact to the accuracy and reliability of this recommendation tool.

4 Conclusion and Discussion

In this proposal, we demonstrated preliminary usage of ML tools like OCR and NLP in automating the carbon sequestration site optimization process. It is evident that OCR and NLP can support the currently human-interpretation-dominant decision making process to select a more objective site for CO_2 storage, to help identify possible leak paths in the CO_2 storage trap under investigation, to assist in locating a depleted oil/gas reservoir suitable for CO_2 injection and long term storage. By automating these tasks, ML tools can drastically decrease the time spent on site assessment, thus enabling a faster implementation of the CO_2 storage facility. In return, the proposed automated workflow will enable us to decarbonize faster and in larger quantities.

In the longer term, we expect ML tools like deep neural network (DNN) can also be utilized in CO_2 storage monitoring, injection optimization etc. For example, it has been shown that ML is promising in interpreting the flow rate and pressure in oil reservoirs [10] through inversion from raw field measurements. We can apply similar methodology to monitor CO_2 fluid properties in the storage trap to detect possible CO_2 leaking back to the surface, and take proactive measures to fix the leak.

By injecting CO_2 into a suitable geological formation in a safe and sustainable manner, the O&G industry can contribute substantially in reducing the global warming and achieving the Paris Agreement goal by the end of the century.

Acknowledgments and Disclosure of Funding

The authors would like to thank Shell for permission to show this work.

References

[1] IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

[2] Wilberforce, T., et.al., 2019, Outlook of carbon capture technology and challenges: Science of the Total Environment **657**, 56–72.

[3] U.S. Department of Energy (DOE) Carbon Storage Program: `https://www.netl.doe.gov/coal/carbon-storage`, retrieved on 08/31/2021.

[4] Chen, T. and Guestrin C., 2016, `XGBoost`: A scalable tree boosting system, arXiv: 1603.02754.

[5] Szegedy, C., et.al., 2017, Deep neural networks for object detection: Conference on Neural Information Processing Systems.

[6] Sidahmed, M., Roy A., and Sayed A., 2017, Streamline rock facies classification with deep learning cognitive process: SPE Annual Technical Conference and Exhibition.

[7] Chen J. and Zeng Y., 2018, Application of machine learning in rock facies classification with physics-motivated feature augmentation, arXiv:1808.09856.

[8] Thanoon, D., et.al., 2021, Deep Seismic2Well Tie: A Physics-guided CNN approach to a classic geophysical workflow: International Meeting for Applied Geoscience & Energy.

[9] Zeng Y., Jiang K., and Chen J., 2018, Automatic seismic salt interpretation with deep convolutional neural networks, arXiv:1812.01101.

[10] Roland, H., 2021, Big data and machine learning in reservoir analysis: SPE distinguished lecturers series `https://webevents.spe.org/products/big-data-and-machine-learning-in-reservoir-analysis#tab-product_tab_overview`, retrieved on 08/31/2021.

[11] Seth P., Elliott B., and Sharma M. M., 2020, Rapid analysis of offset well pressure response during fracturing: distinguishing between poroelastic, hydraulic and Frac-Hit responses in field data using pattern recognition: Unconventional Resources Technology Conference `https://doi.org/10.15530/urtec-2020-3129`.

[12] Smith, R., 2007, An overview of the Tesseract OCR engine, Proceedings of the Ninth International Conference on Document Analysis and Recognition **02** 629–633.

[13] Young T., et.al., 2017, Recent trends in deep learning based natural language processing, arXiv:1708.02709.