AI for Earth Grantee Profile
WikiNet
Improving contaminated site cleanup with AI

Summary

Contamination of soil and groundwater by toxic pollutants is a serious problem worldwide, causing malnutrition, diseases, and death among many people. The degradation rate of natural environments by these contaminants reduces exploitable farmland and increases the extinction rate of fauna and flora species. As industrialization spreads globally, it brings these problems to newly developing nations that lack the experience and expertise to regulate, prevent, or clean up contaminated sites. WikiNet is developing a solution that draws upon the accumulated knowledge from past cleanup efforts to provide automated recommendations for more efficient and effective remediation methods.

Providing expert recommendations for toxic site cleanup with AI

Many common human activities, from agriculture to mining to manufacturing and even to waste removal, create pollution that contaminates the soil and through that the groundwater supply. This contamination can have serious immediate and long-term effects upon human health and the surrounding ecosystem. A 2007 survey of over 120 research papers concluded that 40 percent of (human) deaths worldwide were caused by water, soil, and air pollution, which also contributed to increases in disease and malnutrition. More recently,

> As of 2007, 40 percent of deaths worldwide were caused by water, soil, and air pollution, which also contributed to increases in disease and malnutrition.

in 2018 the Food and Agriculture Organization of the United Nations (UN FAO) issued a report that the serious threat of soil pollution globally is not well enough known and should be a top priority worldwide. Since the Industrial Revolution in the 18th century, the growth of industrialized manufacturing and agriculture has corresponded to increases in this pollution and contamination. The UN FAO report noted that most studies of the problem to date have focused on developed economies where strict regulations were put into place over the 20th century to today. Nations that have been developing industry more recently tend to lack experience both with the pollution and the use of regulations to manage and mitigate the problems.

Extending the experience and expertise on contamination cleanup is the purpose of [WatRem](#), a "remediation cognitive assistant" solution developed by Canadian firm [WikiNet](#). As co-founder and vice president of WikiNet Daniel Fortin explains, one of the most common remediation methods is simply to dig up the contaminated soil and ship it to a treatment facility. But this method can itself generate a lot of greenhouse gases, and just shifts the problem from one location to another—sometimes illegally taking valuable agricultural land as a dump site. It can also risk wider dispersal of contaminants during the transportation process. By helping to identify more efficient solutions that have been applied elsewhere, WikiNet can suggest greener remediation technologies that in the long term reduce the environmental impact of site cleanup.

Improving contaminated site cleanup decisions

Despite the urgency of health and environmental issues, contaminated site remediation projects often have limited budget and resources, especially in developing nations, while at the same time facing demands to be efficient and cost-effective. Those conditions can limit the time available for the cleanup organizations to consider all the options available for a particular site's situation. Although many case studies of remediation techniques are available in scientific journals or conference reports, the data in the studies is highly

"If we help make this data more structured, extract the relevant information, it could help for many different problems in environmental remediation."— Amélie Rolland, Lead AI Developer, WikiNet

unstructured, making it difficult to rapidly assess and consider various techniques. And it can be challenging to keep up with the continual output of new research. That combination of factors makes it impossible for government agencies and environmental experts to consider all the research and past cleanup efforts in their decisions.

WikiNet's aim is to make it easier for both agencies and experts to process this information quickly and get better recommendations for treatment. Lead AI Developer Amélie Rolland says, "Right now, if you want to make decisions and you have all these unstructured documents, it takes a long time to go through them, it might even be impossible for the time you have. If we help to make this data more structured, extract the information that's really relevant, then you can ask your own questions and get insights from this data. I think it could help for so many different problems in environmental remediation."

Making better recommendations through machine learning

The WatRem solution from WikiNet uses AI and machine learning to help make these recommendations happen. After importing a case study, the solution applies a pair of machine learning algorithms using natural language processing to identify and extract the useful data. Key words that refer to relevant topics and categories such as specific contaminants, soil types, or treatment methodologies are tagged and associated—

"Instead of waiting on our own computers, we can use the Microsoft cloud and have these results much faster."—Rolland

for instance, the contaminant trichloroethylene, a concentration of 500 mg per kilogram in the soil, and the technology used to treat the soil can all be grouped as structured data. By creating these structured datasets, machine learning algorithms can be trained to recommend treatments in various site conditions. Expert rules can also eliminate the treatments that are ineffective or inapplicable to a given situation.

WatRem builds on existing research for its recommendations. WikiNet has already labeled a dataset of 2,500 contaminated sites with over 100 treatments and is expecting to reach 5,000 labeled sites by the end of 2019. The solution also uses thousands of documents, over 15,000 labeled sentences and over 200,000 unlabeled sentences to train different AI components and extract the relevant data. That much data and AI algorithms requires a lot of computational power, which is where the Microsoft Azure cloud platform comes in. Thanks to a Microsoft AI for Earth grant, WikiNet is training its machine learning models in Azure Data Science Virtual Machines, which greatly speed up the process. Rolland says, "Instead of waiting on our own computers, we can use the Microsoft cloud and have these results much faster. I don't have to wait a month to test ideas, I can just have a couple hours and I'll have some results."

Looking forward

Currently, WikiNet is using the Azure platform and Azure DevOps services to host the code base, prototype and compare different approaches for its algorithms, and develop automated pipelines for model training and deployment. WikiNet's goal is to deploy a new version of each model every month, so that their predictions are always based on the latest dataset. With three years of development already done on the project, Fortin expects to have the new solution ready for the market by the end of 2019.

About Daniel Fortin

Daniel Fortin is the co-founder and vice president in charge of strategic partnerships at WikiNet. Daniel has held senior positions for more than 38 years in international companies such as IBM. He is an entrepreneur and the founding president of a company publishing software. He was business developer and sales and marketing vice president for several companies in small and medium-sized markets in North America (Canada and USA) and Asia. As a partner, Daniel has been involved for over 20 years with IBM and is still in various advisory councils. His priority with WikiNet is to guide both the selection of technological solutions and the application development supported by AI technologies.

About Amélie Rolland

Amélie Rolland holds a Master's degree in Computer Science from Laval University in Quebec City. Her work on structured prediction was published at the International Conference on Machine Learning (ICML), and she also contributed to a chapter in the book Advanced Structured Prediction. She has worked as a Research Assistant on machine learning projects in the health, security, forestry, and financial sectors, in collaboration with enterprises such as Desjardins, the largest federation of credit unions in North America. She currently leads the Artificial Intelligence team at WikiNet which focuses on multi-label classification, pairwise ranking, and natural language processing techniques to support environmental site cleanup.

Resources

Websites
[WikiNet](#) home site

Documentation
"Report sounds alarm on soil pollution." Food and Agriculture Organization of the United Nations. May 2, 2018. http://www.fao.org/news/story/en/item/1126971/icode/

Susan S. Lang. "Water, air and soil pollution causes 40 percent of deaths worldwide, Cornell research survey finds." Cornell Chronicle. August 2, 2007. http://news.cornell.edu/stories/2007/08/pollution-causes-40-percent-deaths-worldwide-study-finds