ClimateNode: Using Natural Language Processing (NLP) for climate risk intelligence

Helen Jackson, Director, ClimateNode, Cambridge Institute for Sustainability Leadership Canopy Workspace

Problem Statement

Climate change is placing strains on the systems underpinning human welfare, including agriculture, infrastructure and natural capital. Information on evolving climate risks is needed to plan and make adaptation decisions in public policy, finance and business.

Frameworks and tools have been created to assess climate risk. However, there is growing recognition that there are aspects of climate risk which quantitative techniques do not always capture well, including:

- second-order effects (e.g. supply chain impacts, conflict, migration, cascading impacts from infrastructure failure)
- effectiveness of adaptation efforts
- how climate-related risks/events interact with other risks, events and trends

Risk assessment tools based on climate projections are also not open to new information on how impacts are materialising.

We need additional methods to help us monitor and understand how climate change is affecting or could affect economic activity and human welfare.

Insights on the actual or potential impacts of climate-related hazards on individual places, sectors, assets and companies are continually being generated by news articles, scientific papers and reports but are spread across hundreds of thousands of documents.

Harnessing unstructured data with NLP

In order to harness large volumes of unstructured data, we need techniques to extract, compile, store, organize and retrieve it. Natural Language Processing (NLP) – the branch of AI tasked with getting computers to understand and generate human language – can help.

Knowledge extraction

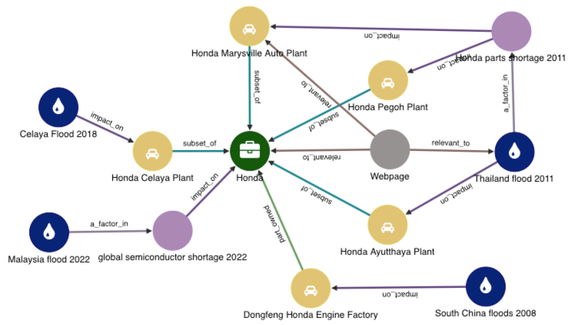
Named Entity Recognition (NER) is one NLP technique. An entity is a 'type of thing', such as person or company, which is used consistently within sentences.

An NER pipeline might consist of doing systematic, automated internet searches for particular news articles, identifying place, asset and company names via an NER model, then storing the details of the entities mentioned in a database.

Production was halted at a major car plant in Guangzhou$_{PLACE}$ after it was flooded by torrential rain on Wednesday. Rain flooded workshops and caused a power cut at the Dongfeng Honda Engine Factory$_{ASSET}$. The joint venture between Honda$_{COMPANY}$ and Dongfeng$_{COMPANY}$ makes more than 400,000 cars a year.

Knowledge representation

A graph database can be used to flexibly link entities together, forming a knowledge graph; a structured representation of knowledge that allows entities and the relationships between them. to be organised and tracked.



LLMs offer new techniques

Flexible knowledge extraction

Large language models (LLMs) such as GPT-4 offer more flexibility over the entities, relationships and information it is possible to extract from texts. In addition, they can recognise entities without specific training data.

PROMPT = "Do the search results below help answer the following question? Answer YES/NO. If the answer is YES, compile a JSON dataset of the distinct incidents from the search results .

QUESTION: "Which {INSERT COMPANY NAME} factories have been affected by flooding?"

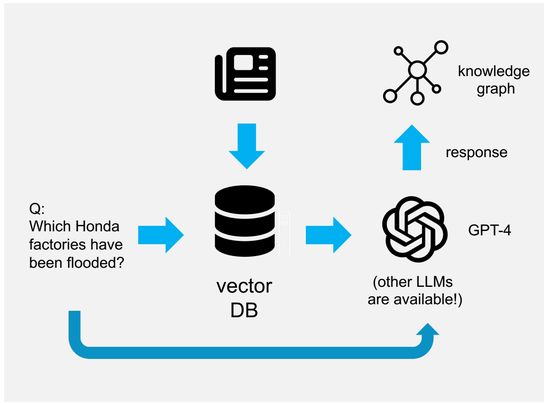
SEARCH RESULTS:

⋮

```
[
  {
    "company":"Honda",
    "factory_location": "Central Thailand",
    "year": 2011,
    "impact": "Factory shut down, 4.7% of global output
               affected."
  },
  {
    "company":"Honda",
    "factory_location": "US and Canada",
    "year": 2011,
    "impact": "Production cut by 50% due to parts shortage
               from floods in Thailand."
  },
  {
    "company":"Honda",
    "factory_location": "Celaya, Mexico",
    "year": 2018,
    "impact": "Auto and component production suspended
               due to river overflow."
  },
⋮
```

Retrieval Augmented Generation

LLMs can also be used to interrogate users' own library of documents. Texts are converted into numerical representations (embeddings vectors) which are stored in a specialised vector database. When the user asks a question, the texts which are most semantically similar to the question, (therefore most likely to be able to help answer it), are fed to the LLM as "context" along with the question. This technique provides much more accurate and domain-specific answers than relying on the model's intrinsic knowledge.



Generating reports

Outputs of systematic searches and LLMs can be stored on the knowledge graph and used to generate reports (subject to human quality control) relating to individual companies, assets, locations and sectors.

Physical Impact Report: Honda

Overview

At least 3 of Honda's facilities have directly experienced flooding, while others have experienced supply chain disruption due to flood events. During flooding in Thailand in 2011, Honda's international production was disrupted, including all six plants in North America.[1] At a global level, Honda's output was down by approx. 150,000 cars and it lost 55% of its operating profit ($1.4bn).[2]

Honda's 2022 Sustainability Report acknowledges that increased severity of floods and supply chain disruption are among the physical climate risks potentially affecting its operations. It mentions that a Business Continuity Plan has been formulated to counter these risks. [3]

Identified incidents

Direct impacts

Ayutthaya Plant, Thailand, 2011: Honda was the only automotive company to be directly impacted by major floods in Thailand in 2011[4]. The plant, which represented 4.7% of Honda's global output at the time, was completely inundated and shut down[5,6]; flood defences built around the factory were breached[6]; over 1,000 flood–damaged cars were scrapped.[7]; and 80% of employees were also affected by the floods[6]. The factory was back in partial operation within 3-4 months of engineers being able to access the site.[8] The site's levee system was upgraded to prevent similar future impacts.[8,9]

Celaya Plant, Mexico, 2018: All auto and component production was suspended at the plant in 2018 when a local river overflowed, making entry into the factory impossible[10,11] Full operations were not resumed for several months[12]. The flooding cost Honda approx. $450 million.[12,13]

⋮

Supply chain impacts

2021 Malaysia floods: Severe flooding in Malaysia in December 2021 disrupted the semiconductor supply chain, compounding chip shortages caused by the COVID-19 pandemic. Honda responded by allocating chips to its most in-demand/profitable models.[14]

Contact information

ClimateNode, The Canopy Workspace, 1 Regent Street, Cambridge CB2 1GG

Email: Helen Jackson helen@climatenode.org

www.climatenode.org

https://www.linkedin.com/company/76378034