

Hail Forecasting Project

Hokyung (Andy) Lee, Dominique Brunet, Robert Crawford

Motivation



Each year in Canada, hail causes an average \$134M in damage (2004-2013). Thus, it is very important to accurately predict hail events to minimize damage

With this project, we aim to achieve:

- 1 Improvement in quality of hail report database
- 2 Improve hail prediction accuracy

Models & Datasets

10K observations of hail vs no hail, previously prepared for CodeML hackathon, were utilized to train the following XGBoost models:

- 1 No Hail vs Hail (e.g. Binary)
- 2 Severe Weather vs Non-Severe Weather (e.g. Binary)
- 3 No Hail vs Non-Severe Hail vs Severe Hail vs Seriously Severe Hail (e.g. Categorical)

Training Results

Model Type	POD	POFD	PSS
Hail vs No Hail	0.989	0	0.989
Severe Weather vs Non-Severe Weather	0.911	0.41	0.502

Fig 1. Training Results for Binary Classification Models

Category	POD	POFD	PSS
No Hail	0.951	0.0314	0.92
Non-Severe Hail	0.407	0.133	0.274
Severe Hail	0.624	0.341	0.283
Seriously Severe Hail	0.25	0.0798	0.17

Fig 2. Training Results for Categorical Classification Model

Correlated Features

Lighting density and temperature at 650-700hPa pressure level before hail events correlated with the occurrences and severity of the hail the most

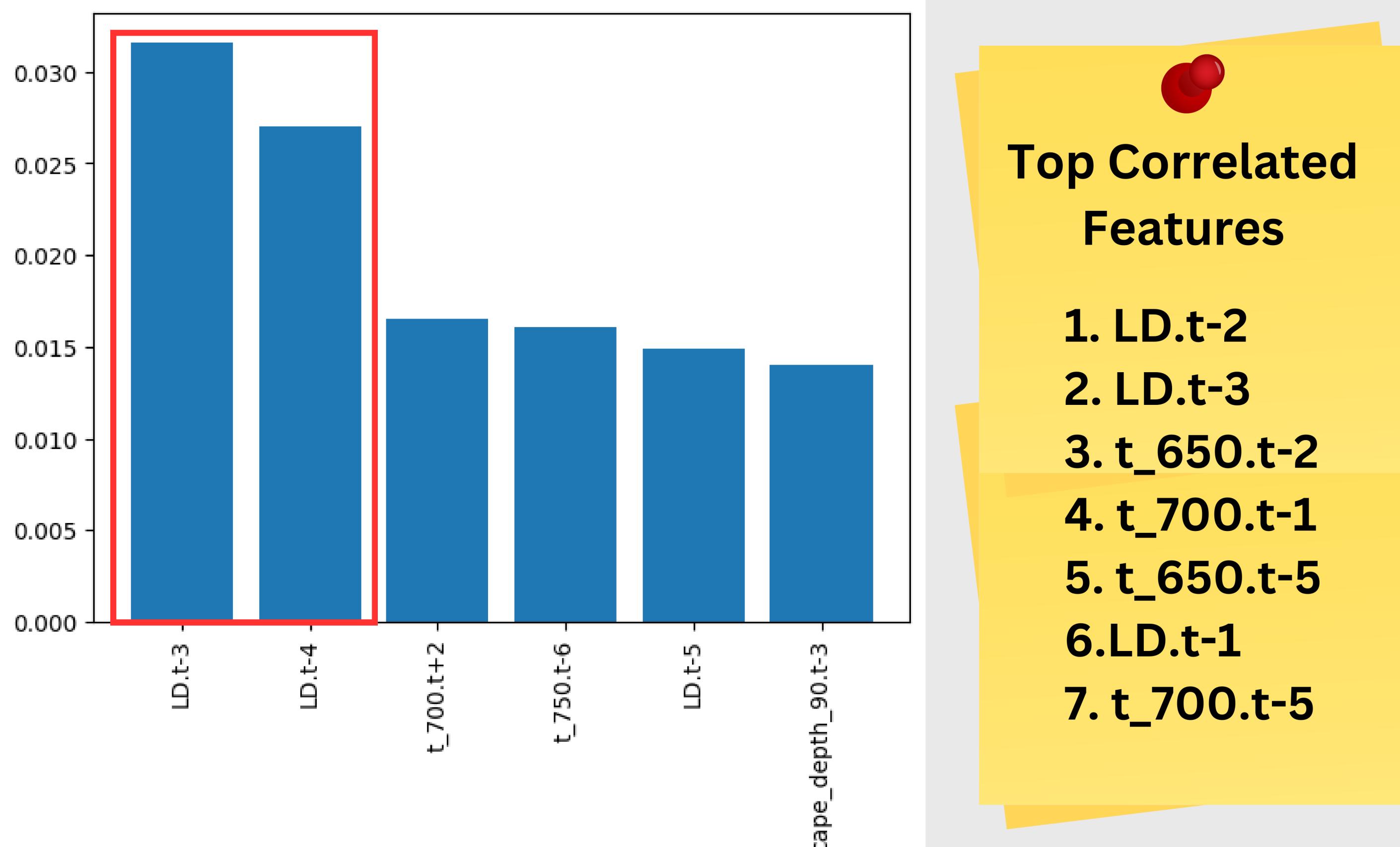


Fig 3. Top Correlated Features to Categorical Hail Size Prediction

*LD prefix stands for lighting density
*t_x prefix stands for temperature at xhPa pressure level
*t-y postfix represents y hours before hail observation

Data Extraction via LLM

To accelerate the process of extracting individual features (e.g. reference objects, locations) from the observation notes provided in the dataset, we utilized Large Language Models (LLMs), specifically GPT-4 through Azure API

Similarly, LLM was utilized to filter out confidential information (e.g. name, email, phone number) from observation notes to replace with generic identifiers

Prompt

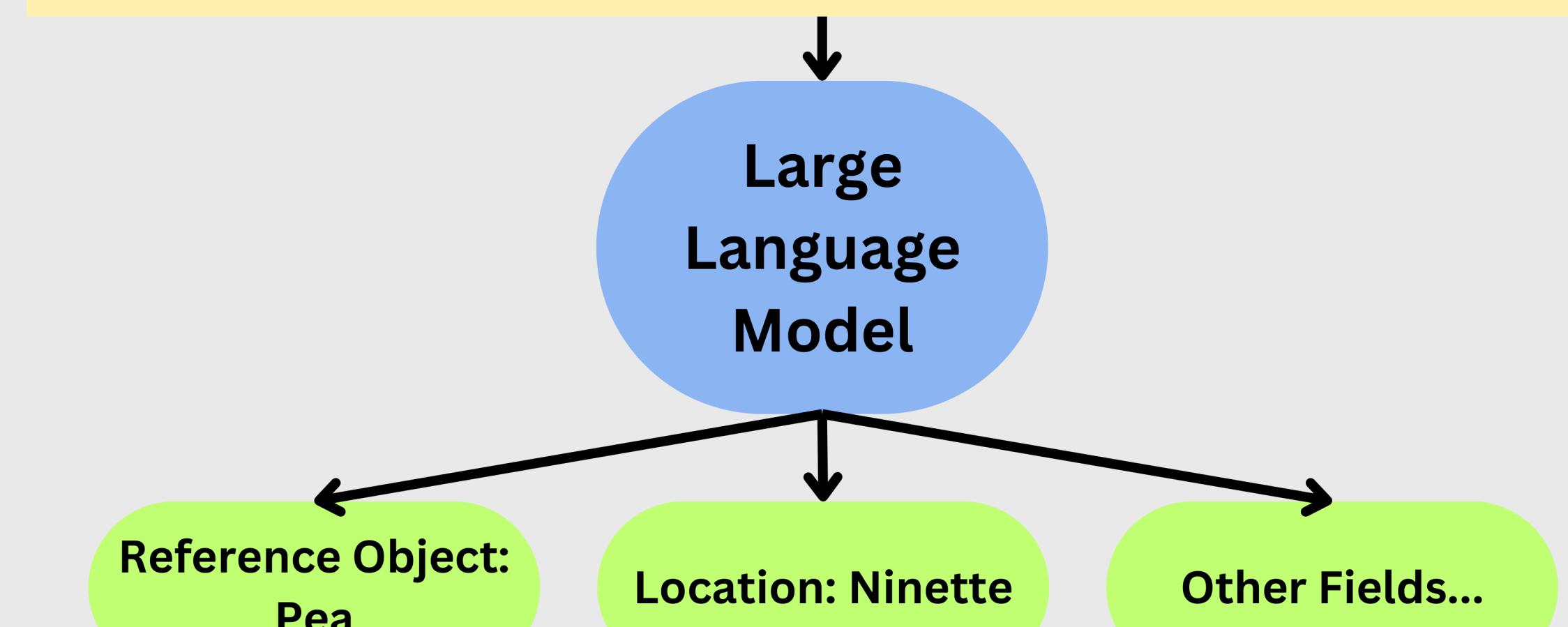
- Given the provided list of reference objects, extract the value from the provided observation note
- ...

Observation Note

<person> spotted a tornado on the ground SW of Ninette lasting about 3 minutes on the ground. It was small and rope like in appearance. Lifted just after report. Pea sized hail associated with it.

Reference Objects

- pea
- marble
- ...



The LLM-extracted information was compared with the human-extracted truth values using syntactic match, LLM-assisted eval, and Google Maps API (for locations)

Extracted Data	Accuracy
Reference Object	83.00%
Hail Diameter	91.40%
Location	93.04%

Fig 4. Accuracy of Each Features Extracted from Observation Notes