Explainable AI in Urban Applications

Ziqi Li

School of Geographical and Earth Sciences
Today

• What is eXplainable AI (XAI) and why does it matter?

• A case study of modelling ride-sharing preferences using XAI and big data.
Emergence of AI

Big Data

Computing

ML Algorithm
He et al., 2021
Road Map Segmentation

Historical Accidents (2 years)

Satellite Imagery

GPS Trajectories

(a) Madrid
(b) Barcelona
(c) Sevilla
(d) Zaragoza
(e) Valladolid
(f) A Coruña

He et al., 2021

Arribas-Bel et al., 2021
Why AI/ML

- Unstructured data (e.g. image, video, text, speech, GPS, point clouds, etc.)
- Scalability for big data
- Fewer assumptions (distribution, relationship)
- Automated model selection
- Superior predictive accuracy
Tradeoff between accuracy and interpretability

- Linear Regression
- Decision Trees
- Support Vector Machine
- Ensemble Methods
- Deep Learning
Black box of AI

- AI models are intrinsically hard to interpret due to a huge number of parameters being estimated.

Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance
Monday, April 4, 2022

In recent years, large neural networks trained for lang achieved impressive results across a wide range of tasks. In models (LLMs) can be used for few-shot learning and scale tasks; specific data collection or model parametere LaMDA, Gopher, and Megatron-Turing NLG, achieved tasks by scaling model size, using sparsely activated more diverse sources. Yet much work remains in und few-shot learning as we push the limits of model scal

GPT-4 Will Have 100 Trillion Parameters — 500x the Size of GPT-3
Are there any limits to large neural networks?
Black box of AI

Input → BLACK BOX → Output

"Why did you predict 42 for this data point?"

*awkward silence*
Trust issues with black-box AI

- Technical:
  - Why a certain decision is made; when does the system work/fail, how to correct the error and improve the model?

- Ethical:
  - Critical decisions are made by AI: healthcare, finance, security, etc.
  - Discrimination and biases
Racial bias in COMPAS

Angwin et al. (2016)

Black offenders were seen almost twice as likely as white offenders to be labeled a higher risk but not actually re-offend.
Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford
Explainable AI (XAI)

- Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms.

- Improve understanding of the underlying decision processes.

- Provide credibility and confidence of the model parameters and outcome.

https://miro.medium.com/max/1400/1*-vNI5N7f1GUBWFZUfwd8w.png
Tradeoff between accuracy and interpretability

Accuracy

Interpretability

- Linear Regression
- Decision Trees
- Support Vector Machine
- Ensemble Methods
- Deep Learning

XAI
Google trends searching “XAI”
XAI in urban applications

Interpretable machine learning models for crime prediction
Xu Zhang, Lin Liu, Minzun Lan, Guangle Song, Lixi Xiao, Jianguo Chen

Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach
Runzi Wang, Jun-Hyun Kim, Minghan Li

Decoding pedestrian and automated vehicle interactions using immersive virtual reality and interpretable deep learning
Arsah Khatizian, Bilal Farooq

Hybrid interpretable predictive machine learning model for air pollution prediction
Yuanlin Gu, Bihua Li, Qinggang Meng
Fig. 9. Local explanations of an image classification prediction described using LIME [65]. Here, top three classes are "electric guitar" ($p = 0.32$), "acoustic guitar" ($p = 0.24$) and "labrador" ($p = 0.21$). By selecting a group of "superpixels" from the input image, the classifier provides visual explanations to the top predicted labels.
XAI for text

https://pair-code.github.io/lit/demos/
XAI for tabular data

- Feature “importance”
  - What are the major contributors to the model.
- Partial dependence plot
  - Relationships between \( X \) and \( y \).
Two types of model explanations

Model-based

Post-hoc

Murdoch et al., 2019. PNAS
Post-hoc explanation

\[ f(x) = x \]

\[ f(x) = x^2 \]
SHAP

- **SHAP (Shapley Additive Explanations):** is a game theoretic approach to quantify the contribution of each feature in the model that collectively makes the prediction.

\[ \hat{y}_i = \text{shap}_0 + \text{shap}(X_{1i}) + \text{shap}(X_{2i}) + \ldots + \text{shap}(X_{pi}) \]

SHAP

A unified approach to interpreting model predictions

Authors: Scott M. Lundberg, Su-In Lee

Publication date: 2017

Journal: Advances in neural information processing systems

Volume: 30

Description: Understanding why a model makes a certain prediction can be as crucial as the prediction’s accuracy in many applications. However, the highest accuracy for large modern datasets is often achieved by complex models that even experts struggle to interpret, such as ensemble or deep learning models, creating a tension between accuracy and interpretability. In response, various methods have recently been proposed to help users interpret the predictions of complex models, but it is often unclear how these methods are related and when one method is preferable over another. To address this problem, we present a unified framework for interpreting predictions, SHAP (SHapley Additive exPlanations). SHAP assigns each feature an importance value for a particular prediction. Its novel components include: (1) the identification of a new class of additive feature importance measures, and (2) theoretical results showing there is a unique solution in this class with a set of desirable properties. The new class unifies six existing methods, notable because several recent methods in the class lack the proposed desirable properties. Based on insights from this unification, we present new methods that show improved computational performance and/or better consistency with human intuition than previous approaches.

Total citations: Cited by 8013

Shapley

Lloyd Shapley (1923-2016)
Nobel Prize in Economics (2012)

\[ \varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \frac{(v(S \cup \{i\}) - v(S))}{(n - |S| - 1)!} \]

Possible permutations

\[
\text{Marginal contribution}
\]

Shapley value measures -
the average of marginal contribution of a player in a game over all possible different permutations in which the coalition can be formed.
A Shapley value example

• Null: 0
• Ziqi: 5
• Qunshan: 10
• Nick: 100

• \{Qunshan + Ziqi\}: 5
• \{Nick + Ziqi\}: 120
• \{Nick + Qunshan\}: 140
• \{Nick + Qunshan + Ziqi\}: 150

• So, what is the contribution of each of us?
A Shapley value example

- Null: 0
- Ziqi: 5
- Qunshan: 10
- Nick: 100

- {Qunshan + Ziqi}: 5
- {Nick + Ziqi}: 120
- {Nick + Qunshan}: 140
- {Nick + Qunshan + Ziqi}: 150

So, what is the contribution of each us?

\[
\text{Shapley(Ziqi)} = \frac{v(Ziqi) - v(Null)}{3} + \\
\frac{v(Ziqi + Qunshan) - v(Qunshan)}{6} + \\
\frac{v(Ziqi + Nick) - v(Nick)}{6} + \\
\frac{v(Ziqi + Qunshan + Nick) - v(Qunshan + Nick)}{3}
\]

\[
= \frac{5}{3} + \frac{-5}{6} + \frac{20}{6} + \frac{10}{3} = \frac{45}{6} = 7.5
\]
A Shapley value example

- Null: 0
- Ziqi: 5
- Qunshan: 10
- Nick: 100

- \{Qunshan + Ziqi\}: 5
- \{Nick + Ziqi\}: 120
- \{Nick + Qunshan\}: 140
- \{Nick + Qunshan + Ziqi\}: 150

- So, what is the contribution of each of us?

\[
\text{Shapley(Ziqi)} = \frac{v(Ziqi) - v(Null)}{3} + \frac{v(Ziqi + Qunshan) - v(Qunshan)}{6} + \frac{v(Ziqi + Nick) - v(Nick)}{6} + \frac{v(Ziqi + Qunshan + Nick) - v(Qunshan + Nick)}{3} \\
= \frac{5}{3} + \frac{-5}{6} + \frac{20}{6} + \frac{10}{3} = \frac{45}{6} = 7.5
\]

- Shapley(Qunshan): 20
- Shapley(Nick): 122.5
Shapley

• Shapley value is the unique solution for a fair distribution that has certain properties:
  • Null, additivity, efficiency, symmetry properties

• Shapley in ML:
  • Game→ model, player -> feature; outcome→prediction
  • Shapley value of a feature is the feature’s contribution to the model prediction
Shapley

- The computation of Shapley value is NP-hard.
- SHAP provides different approximation methods to estimate Shapley value.
  - Sampling based, Kernel based, Tree based, etc.
- SHAP also unifies some other XAI methods for text and image explanations.
- SHAP has been integrated into industry XAI software such as Amazon SageMaker and Google Vertex Explainable AI.
SHAP

\[ \hat{y}_i = shap_0 + shap(X_{1i}) + shap(X_{2i}) + \ldots + shap(X_{pi}) \]

\[ 0.4 = 0.1 + (0.1) + (0.1) + (-0.3) + (0.4) \]
Feature Importance

Partial Dependence Plot
Today

• What is eXplainable AI (XAI) and why it matters?

• A case study of modelling ridesharing preferences using XAI and big data.
Ride-hailing market

- Uber and Lyft (and others) are common in cities.

- Jupiter Research (Dec 2021):
  - Consumer spending on ride-hailing will approach US $937 billion by 2026.
  - = 50 times the total annual revenue of Transport for London, New York City's MTA, and Beijing Metro in 2021.
Impacts of transportation network companies on urban mobility

Mi Diao, Hui Kong and Jinhua Zhao

The role of transportation network companies (TNCs) in the urban transport system is under intense debate. In this study, we systematically assess three aspects of the net impacts of TNCs on urban mobility in the United States—road congestion, transit ridership and private vehicle ownership—and examine how these impacts have evolved over time. Based on a set of fixed-effect panel models estimated using metropolitan statistical area level data, we find that the introduction of TNCs has been associated with increased congestion in terms of both intensity (by 0.9%) and duration (by 4.5%), and a change in vehicle ownership. Despite the ideal of providing a sustainable mobility option, our analysis suggests that TNCs have intensified urban transport challenges.

The hidden cost of ride-hailing: is Uber bad for the environment?

Research shows that Uber’s carbon footprint depends largely on the characteristics of individual cities.

Los Angeles Times

Why Uber is bad for cities

Column: Uber and Lyft increase traffic and pollution. Why do cities let it happen?
Ride-hailing services (solo vs. shared)

- Solo rides: Single-occupancy. e.g. UberX and Lyft

- Shared rides: shared carpool style. e.g. UberPool and Lyft Shared

- Shared ride benefits: shared rides can reduce traffic congestion, cut per-passenger carbon emissions, reduce parking infrastructure, and provide a more cost-effective way to travel (Shaheen and Cohen, 2019)

- Limited availability/popularity: Available only in selected cities, and occupies 15–25% of total trips in cities such as London, Hangzhou, Toronto, and Chicago.
Understanding willingness to share

• To promote the proportion of shared trips, we need to understand the factors influencing people’s willingness/reluctance to share.
Big trip data

- City: Chicago, USA
- Trip records are time-stamped.
- O-D are geo-referenced at the census tract level.
- Label: shared trip? 1 : 0

https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p
Big trip data

- City: Chicago, USA
- Trip records are time-stamped.
- O-D are geo-referenced at the census tract level.
- Label: shared trip? 1 : 0

https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p
Big trip data

- City: Chicago, USA
- Trip records are time-stamped.
- O-D are geo-referenced at the census tract level.
- Label: shared trip? 1 : 0
- Date: whole year of 2019
  - Shared rides were suspended due to COVID.
- Total records: >10M
Modelling willingness to share

No post-trip information
- Actual trip distance
- Actual trip duration
- Total cost (incl. tip, late fees, etc.)

Trip base fare
Additional fees

Distances:
- Trip distance (straight)
- Distance to downtown Built-environment:
- Walkability
- Distance to public transportation Socio-economics
- Population density
- % no car
- Education (% College)
- Income (median household)
- Race/Ethnicity (% non white)
- Age (% 19-29)

Hour of the day
Week/weekend Holiday
Weather:
- Rain
- Temperature
- Wind
XGBoost

• XGBoost is a gradient boosting method that uses a gradient descent optimisation algorithm to sequentially ensemble decision trees to minimise model error (Chen and Guestrin, 2016).

• XGBoost typically outperforms deep learning, random forests and other alternatives when handling tabular data (Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022).

• XGBoost is highly scalable.
## Model accuracy

**Table 2.** Training and testing accuracy of the XGBoost model.

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td>86.8%</td>
<td>86.0%</td>
</tr>
<tr>
<td>True Negative</td>
<td>95.8%</td>
<td>95.0%</td>
</tr>
<tr>
<td>False Positive</td>
<td>13.2%</td>
<td>14.0%</td>
</tr>
<tr>
<td>False Negative</td>
<td>4.2%</td>
<td>5.1%</td>
</tr>
<tr>
<td>F1</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td><strong>91.3%</strong></td>
<td><strong>90.5%</strong></td>
</tr>
</tbody>
</table>
Feature group feature importance

Figure 3 SHAP-based global feature importance ranking for five groups of features.

SHAP values have been converted to probability.
SHAP-based partial dependence plots

- People favour shared rides for longer distance trips while holding everything else constant.

- People tend to favour shared rides during morning and afternoon commute hours.

- Solo rides are preferred in the evening.

- Willingness has a higher variance on a non-working day, with a preference to take more solo rides.
SHAP-based partial dependence plots

- Shared rides are preferred in low-income, low-education, non-white, young, high vehicle ownership census tracts.
Trip-level explanation analysis

Table 3. Feature values and their SHAP contributes to three example trips in Figure 8.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Trip</th>
<th>Shared</th>
<th>Time</th>
<th>SHAP (Hour)</th>
<th>Median Income</th>
<th>SHAP (Median)</th>
<th>Pct Bach</th>
<th>SHAP (Pct Bach)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A to Downtown</td>
<td>Yes</td>
<td>8am</td>
<td>+22.4%</td>
<td>168,352</td>
<td>-0.2%</td>
<td>36.3%</td>
<td>+0.4%</td>
</tr>
<tr>
<td>Downtown to A</td>
<td>No</td>
<td></td>
<td>10pm</td>
<td>-15.5%</td>
<td></td>
<td>-2.6%</td>
<td></td>
<td>-1.7%</td>
</tr>
<tr>
<td>B</td>
<td>B to Downtown</td>
<td>Yes</td>
<td>7am</td>
<td>+16.2%</td>
<td>32,308</td>
<td>+13.0%</td>
<td>2.0%</td>
<td>+17.4%</td>
</tr>
</tbody>
</table>
Conclusions

• Decision to rideshare is largely driven by economic considerations.

• User tend to prefer ridesharing during am/pm rush hours.

• Socio-economic disparities.

• How will COVID impact the ridesharing market?

• How does ridesharing compete with public transport?
Some thoughts on XAI

- We should adopt a simpler and more interpretable model if it has similar performance.

- If AI model is used, some degree of XAI should be presented to provide credibility of the model.

- XAI can be a good alternative to statistical approaches especially when data is large.

- Explanations are skewed if either data or model is biased.
Some further thoughts on XAI

- How do we evaluate explanation accuracy/faithfulness?
  - Ground truth validations.
  - Simulating simple data generating processes to validate model and explanations.
Some further thoughts on XAI

- How do we evaluate explanation accuracy/faithfulness?
  - Ground truth validations.
  - Simulating simple data generating processes to validate model and explanations.
- How can XAI inform us with the development of (Geo)AI?
  - XAI provides a great way to diagnose and improve model.
Some further thoughts on XAI

• How do we evaluate explanation accuracy/faithfulness?
  • Ground truth validations.
  • Simulating simple data generating processes to validate model and explanations.

• How can XAI inform us with the development of (Geo)AI?
  • XAI provides a great way to diagnose and improve model.

• Can XAI give us spatial explanations?
Some thoughts on XAI

Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost

- Benchmark with spatial models
- Validate explanations
- Discuss the use of XAI with spatial data
Interpretable Machine Learning

A Guide for Making Black Box Models Explainable

Christoph Molnar

2022-07-12

Summary

Thank you!

• Email: Ziqi.Li@glasgow.ac.uk
• Twitter: @geo_ziqi