

You Don't Have to Have a PhD to Use Machine Learning

... (but it helps)!

Ed Beroset
Principal Technical Leader

MIPSYCON
3 November 2020



What is Machine Learning?

“Machine learning is a field of computer science that aims to teach computers how to learn and act without being explicitly programmed. More specifically, machine learning is an approach to data analysis that involves building and adapting models, which allow programs to "learn" through experience.”

– Dr. Andrew Ng

Why use Machine Learning?

- Let machines do more of our work, saving creative work for humans
- Efficiently extract more value from *data you probably already have*
- While the algorithms and theory are complex, *using* machine learning has become much easier

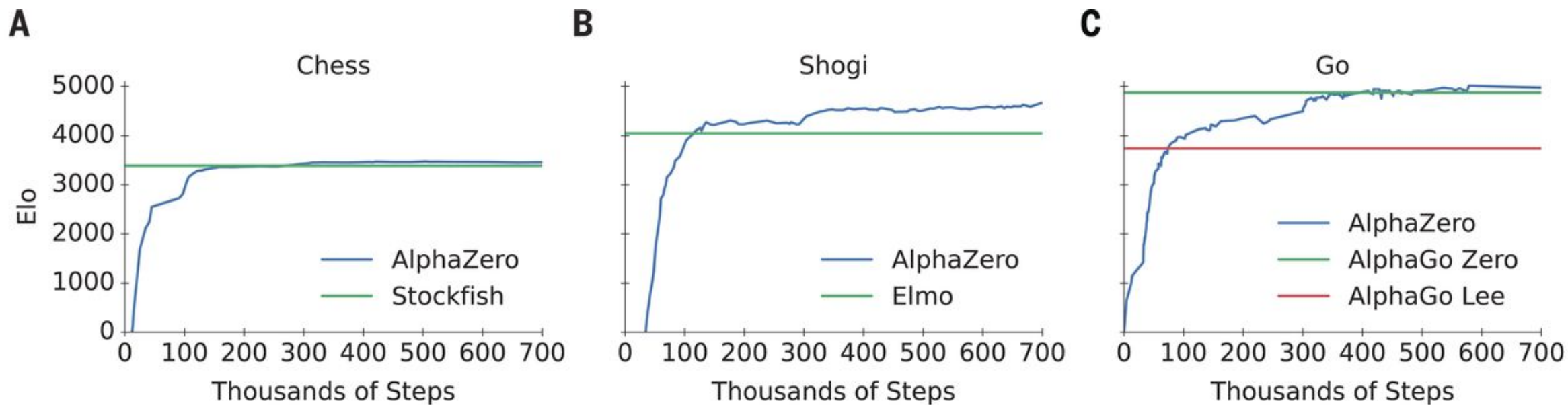
Photo: Thomas Depenbusch
License: cc-by-2.0



Reinforcement learning example: AlphaZero

■ **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play**

BY DAVID SILVER, THOMAS HUBERT, JULIAN SCHRITTWIESER, IOANNIS ANTONOGLU, MATTHEW LAI, ARTHUR GUEZ, MARC LANCTOT, LAURENT SIFRE, DHARSHAN KUMARAN, THORE GRAEPEL, TIMOTHY LILICRAP, KAREN SIMONYAN, DEMIS HASSABIS
SCIENCE | 07 DEC 2018 : 1140-1144



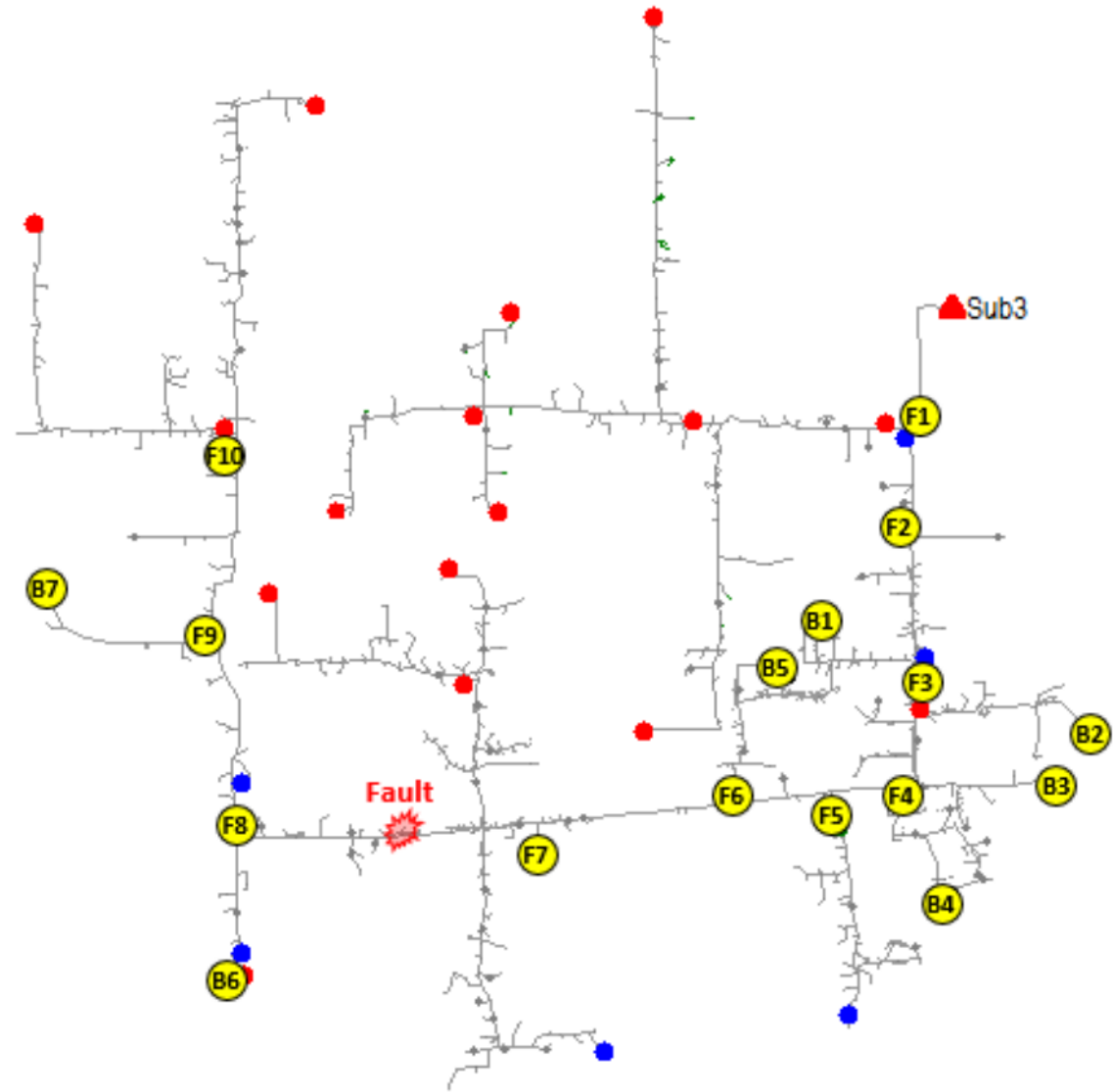
Can you identify the fault location?



Could a
computer?

What if...

- Can meters learn to autonomously identify the location of distribution faults?
- Plan: use reinforcement learning algorithm and OpenDSS modeling



Open Source is Everywhere

- Linux
 - GNU compiler collection (gcc, ld, make)
 - OpenSSL/OpenSSH
 - MySQL
 - BIND
 - Apache
-
- *Check the fine print for your TV, phone, or automobile*

Why Open Source?

- “More eyeballs” theory of software quality
- Lower cost
- No proprietary lock-in
- Free to modify, fix, improve
- Better security via transparency
- Sometimes academic or non-profit origins
- Each person contributes a small amount
- The result is much larger and shared by all
- Example:
 - 100 developers each contribute 1%
 - All get 100% of the software

Some open source software is simply the best in class regardless of origin or license



["Stone Soup"](#) by Qù F Meltingcardford is licensed under [CC BY-SA 3.0](#)

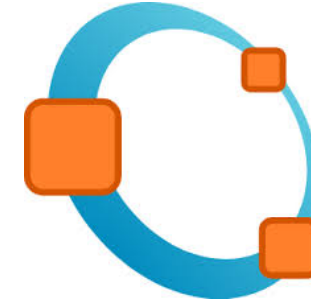
This Presentation



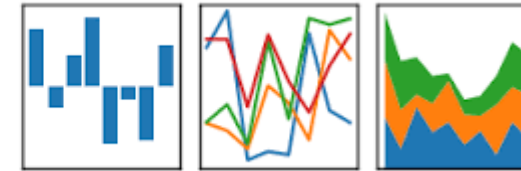
- Neither nutritionally sound nor satisfying meal
- A sweet taste of a few select morsels
- If you don't like one, it won't last too long anyway
- If you do like one, links to explore more

The List

- LibreOffice Calc
- Octave
- Python/Numpy/Pandas
- Apache Spark
- R
- TensorFlow
- Conda



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

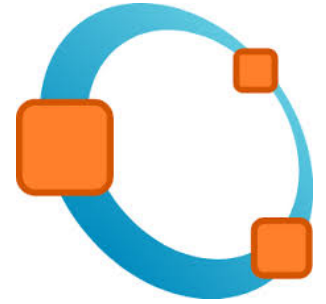


LibreOffice Calc



- <https://www.libreoffice.org/discover/calc/>
- LibreOffice is an office suite similar to Microsoft Office
- Calc is an Excel work-alike except free and open source
- Use it for:
 - Simple analysis of limited amounts of data
 - Quick graphs with familiar interface
- Example:
 - Municipal checking firmware versions vs. comms quality

Octave



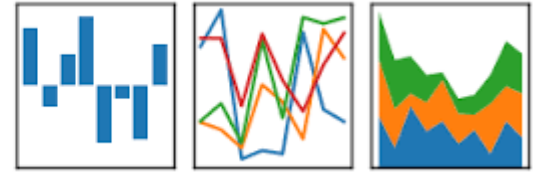
- <https://www.gnu.org/software/octave/>
- “Powerful mathematics-oriented syntax with built-in plotting and visualization tools
- Drop-in compatible with many Matlab scripts”
- Use it for:
 - Exploratory graphing, both 2D and 3D
 - Exploratory algorithm development
- Example:
 - Heat map for max current vs. avg demand vs. meter failure count

Python/Numpy/Pandas

- <https://pandas.pydata.org/>

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



- “*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.”
- Use it for:
 - Python-assisted correlation of large data sets
 - Data cleaning
- Example:
 - Correlation of weather, kWh data

Apache Spark

- <https://spark.apache.org/>
- Spark vs. Hadoop MapReduce
- Spark is in-memory while Hadoop uses disk IO
- Spark has MLlib - a machine learning library
- Use it for:
 - Very large data sets
 - Experimenting with distributed machine learning
- Example:



R



- <https://www.r-project.org/>
- “R is a free software environment for statistical computing and graphics.”
- Use it for:
 - Descriptive statistics of large data sets
 - Publication quality graphics
- Example:
 - Scatter plot of current, voltage, power factor, demand, energy

TensorFlow



- <https://www.tensorflow.org/>
- "An end-to-end open source machine learning platform"
- Use it for:
 - Exploring machine learning algorithms
 - Deployment of machine learning apps
- Example:
 - Smart phone based visual meter tamper app

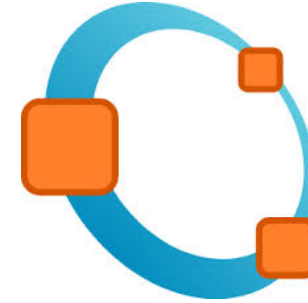
Conda



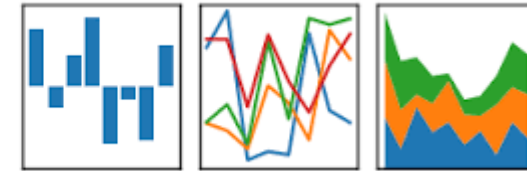
- <https://conda.io/en/latest/>
- "Conda is an open source package management system and environment management system."
- Use it for:
 - Managing environments for development or deployment
 - Distributing Python-based cross-platform application
- Example:
 - Deploy pandas-based tamper analysis program to engineers

The List

- LibreOffice Calc
- Octave
- Python/Numpy/Pandas
- Apache Spark
- R
- TensorFlow
- Conda



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Together...Shaping the Future of Electricity