New and Established Emerging Technologies to the Rescue:
A Global Overview of Sustainable/Good Governance with
Machine Learning/AI and Open Access

Falk Huettmann
-EWHALE lab-
419 Irving 1
Inst. of Arctic Biology
Biology & Wildlife Dept
Fairbanks, Alaska 99775 USA
fhuettmann@alaska.edu

(All materials from WWW, authors and public sources; only for teaching purposes, non-commercial use)
Overview of this presentation

Machine Learning (ML)/Artificial Intelligence (AI) Introduction

Examples of using ML/AI

Generic Use Categories of ML/AI

Teaching ML/AI

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Some references used in this talk

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Linear (Multiple) Regression as something to improve...

Data: R

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Linear (Multiple) Regression as something to improve...

- Summary statistics & metrics are not always unique...

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
What is ML/AI

A machine learns, summarizes, mimics, generalizes and predicts data

A heap of data

PS. Virtually no statistical theory involved, nor needed

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
What is ML/AI

A machine learns, summarizes, mimics, generalizes and predicts data

A heap of data

PS. Virtually no statistical theory involved, nor needed
How many ML/AI algorithms/methods are out there...

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Software Code (selection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Networks</td>
<td>Python</td>
</tr>
<tr>
<td>CARTs</td>
<td>JAVA</td>
</tr>
<tr>
<td>Boosting</td>
<td>R</td>
</tr>
<tr>
<td>Bagging (RandomForest)</td>
<td>Salford Systems Ltd</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>Mathlab</td>
</tr>
<tr>
<td>Maxent</td>
<td>SAS</td>
</tr>
<tr>
<td>Ensemble Models</td>
<td></td>
</tr>
</tbody>
</table>

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
How many ML/AI algorithms/methods are out there...100s!
How many ML/AI algorithms/methods are out there

The keys are

...the algorithm...

...the data...

...the knowhow/expertise
(Virtually) any problem can be tackled with ML/AI...
and it (usually) can be (greatly) improved with ML/AI

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Examples of use: Mineral Resource Prediction...

Aspects of Regional and Worldwide Mineral Resource Prediction

doi: 10.1007/s12583-020-1397-4

Frits Agterberg

Geological Survey of Canada, 601 Booth Street, Ottawa K1A 0E8, Canada

Abstract

The purpose of this contribution is to highlight four topics of regional and worldwide mineral resource prediction: (1) use of the jackknife for bias elimination in regional mineral potential assessments; (2) estimating total amounts of metal from mineral potential maps; (3) fractal/multifractal modeling of mineral deposit density data in permissive areas; and (4) worldwide and large-areas metal size-frequency distribution modeling. The techniques described in this paper...
Examples of use: Global Pandemics...

A global model of avian influenza prediction in wild birds: the importance of northern regions

Koko A. Herrick, Falk Huettmann, and Michael A. Lindgren

Abstract

Avian influenza virus (AIV) is endemic to wild birds, which is its natural reservoir. The virus exhibits a large degree of genetic diversity and most of the isolated strains are of low pathogenicity to poultry. Although AI is relatively ubiquitous in wild bird populations, highly pathogenic H5N1 subtypes have been the focus of most modeling efforts. To better understand viral ecologies of AIV, a predictive model should: 1) include wild birds, 2) include all isolated subtypes, and 3) cover the host’s natural range, unbounded by artificial country borders. As of this writing, there are few large-scale predictive models of AIV in wild birds. We used the Random Forests algorithm, an ensemble data-mining machine-learning method, to develop a global-scale predictive map of AIV, identify important predictors, and describe the environmental niche of AIV in wild bird populations. The model has an accuracy of 0.79 and identified northern areas as having the highest relative predicted risk of outbreak. The primary niche was described as regions of low annual rainfall and low temperatures. This study is the first global-scale model of low-pathogenicity avian influenza in wild birds and underscores the importance of largely unstudied northern regions in the persistence of AIV.

Figure 3 Global map of the predicted relative occurrence of avian influenza virus (AIV) in wild birds. The predictive model was constructed using the Random Forests algorithm on 41 predictor variables. The dots on the map represent all samples in both the testing and training databases. Locations where one or more AIV-positive samples were collected are shown as black dots; locations where no positive samples were collected are marked with white. A single dot may represent multiple samples taken at that location. This map is presented in Robinson (sphere) projection, central meridian 145°.
Examples of use: Find Disease Reservoirs ...(from a WEBPORTAL)

Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Influenza (AI) in the wider Pacific rim using big data sets

Marina Gulyaeva1, Falk Huettemann2, Alexander Shestopalov, Masatoshi Okamatsu, Keita Matsuno1, Duc-Huy Chu, Yoshitaka Sakuda, Alexandra Glyushchenko, Elaina Milton1 & Eric Bortz1

Avian influenza (AI) is a complex but still poorly understood disease; specifically when it comes to reservoirs, co-infections, connectedness and wider landscape perspectives. Low pathogenic (low-path LP) AI in chickens caused by less virulent strains of AI viruses (AIVs)—when compared with highly pathogenic AIVs (HPAIVs)—are not even well-described yet or known how they contribute to wider AI and immune system issues. Co-circulation of LPAIVs with HPAIVs suggests their interactions in their ecological aspects. Here we show for the Pacific Rim an international approach how to data mine and model-predict LP AI and its ecological niche with machine learning and open access data sets and geographic information systems (GIS) on 3 km pixel size for best possible inference. This is based on the best available data on the issue. 40,827 records of lab-analyzed field data from Japan, Russia, Vietnam, Mongolia, Alaska, and Influenza Research Database (IRD) and U.S. Department of Agriculture (USDA) database sets, as well as 12 GIS data layers. We sampled 137 hosts and 328 low-path AIVs

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettemann
Examples of use: Biomass at the World Seafloor ...
Examples of use: National Data Analysis...
Classification and regression with random forests as a standard method for presence-only data SDMs: A future conservation example using China tree species

Lei Zhang, Falk Huettmann, Shiromg Liu, Pengsen Sun, Zhen Yu, Xudong Zhang, Chunrong Mi

Fig. 4 Differences in prediction maps between random forests (RF) regression tree (RT) and classification tree (CT) algorithms for Quercus serrata. Numerical predictions were converted to binary predictions through objective threshold-setting methods (MaxTSS).

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Examples of use: National Data Analysis...

Chapter 7
Mapping Aboveground Biomass of Trees Using Forest Inventory Data and Public Environmental Variables within the Alaskan Boreal Forest

Brian D. Young, John Yarie, David Verbyla, Falk Huettmann, and E. Stuart Chapin III

in Humphries et al. (2018)
Examples of use: Polar Regions (Arctic)
Examples of use: Polar Regions (Antarctica)

http://atlas.biodiversity.aq/

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Examples of use: Hindu Kush Himalaya

Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (*Ailurus fulgens*) in the Hindu-Kush Himalaya region


**Show more**

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Examples of use: Climate Change ...

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Can ML/AI fail?

Of course, mostly when applied wrongly and by not trained non-experts.

Not all algorithms perform the same way and for different questions.

Examples: Predictions, (Autocorrelation), Robustness.
10 (simple) Use Categories of ML/AI
Data Mining (e.g. Find the Signal)

Chapter 12
Using TreeNet, a Machine Learning Approach to Better Understand Factors that Influence Elevated Blood Lead Levels in Wintering Golden Eagles in the Western United States

Erin H. Craig, Tim H. Craig, and Mark R. Fuller

Investigating the effects of environmental contaminants on individuals and, ultimately, on wildlife populations is challenging. This is partly because it is difficult to interpret results potentially influenced by many interacting variables. The complexity of such datasets can constrain the ability of researchers to obtain meaningful biological results using traditional statistical approaches. This can result in under-utilization of available information that is potentially useful to management decision makers (Craig and Huettmann 2009). Machine learning (ML) algorithms provide powerful tools that are increasingly seen as practical solutions for helping to address complex problems. They can be used as an end-product for prediction, to reveal patterns in data related to the incidence of contaminants in target species, and to guide future research efforts, or alternatively, to aid in hypothesis development or in conjunction with conventional statistical tools. ML has been used for decades in investigating the causes of disease in human populations (Colli et al. 2016; see e.g., Cooper et al. 1997; Coore and Wishart 2006; Moradi et al. 2015; Shipp et al. 2002; Serram et al. 2013) and the incidence of contaminants in the environment (see e.g.,

in Humphries et al. (2018)

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
10 (simple) Use Categories of ML/AI

Classification (Profiling)

A machine can count and learn better, and more, than a human, e.g. supervised learning

Friend/Enemy recognition, e.g. on radar screens!

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
10 (simple) Use Categories of ML/AI
Prediction (Current, Past, Future/Impact/Scenarios)

Fig. 1. Pixel-based map of the (A) Large Decline Model, (B) Moderate Decline Model, and (C) No Decline Model of passerine migratory birds for the 1970–1990 period. The map shows the modelled response variable as a relative index of change for the respective model. In all maps, dark-brown colors refer to a large response and green colors refer to a small response. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
10 (simple) Use Categories of ML/AI

Outlier Detection

In RandomForest (Minitab-Salford)

⇒ Flag outliers and data qualities

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
A Global Model of Predicted Peregrine Falcon (*Falco peregrinus*) Distribution with Open Source GIS Code and 104 Open Access Layers for use by the global public

Sumithra Sriram and Falk Huettmann

Received: 09 Dec 2016 – Accepted for review: 02 Feb 2017 – Discussion started: 13 Feb 2017
10 (simple) Use Categories of ML/AI
Overcome Data Gaps (e.g. Imputation)
10 (simple) Use Categories of ML/AI

Data Cloning

e.g. make a poor table with few rows and many columns better for a ~robust inference
ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann

10 (simple) Use Categories of ML/AI

Batteries/Shaving

Drop ‘the best’ (or worst etc) predictor each time

e.g.

\[ y = X_1 + X_2 + X_3 + X_4 + X_5 \]
\[ y = X_1 + X_2 + X_3 + X_4 \]
\[ y = X_1 + X_2 + X_3 \]
\[ y = X_1 + X_2 \]
\[ Y = X_1 \]

Provides another, unique, view on predictors, interactions and model subsets (=your project)

=> Identifies powerful predictors

in Humphries et al. (2018)
10 (simple) Use Categories of ML/AI

Batteries/Swapping

Swap out ‘the best’ (or worse etc) predictor each time with another subset

e.g.
\[ y = X_1 + X_2 + X_3 + X_4 + X_5 \]
\[ y = X_1 + X_2 + X_3 + X_4 \]
\[ y = X_1 + X_2 + X_3 \]
\[ y = X_1 + X_2 \]
\[ Y = X_1 \]

Provides another, unique, view on predictor contribution, replacement options and model subsets (=your project) => **Replaces** (best) predictors!
A power of using ML/AI: Combined Approaches across disciplines

Data Table

GIS

Complex and Social Science e.g. telecoupling

Fig. 7. Map of sending, receiving and spillover system as well as major flows under the telecoupling framework for the fisheries in the Patagonian Shelf. Black arrows indicate the relative size of fish products flows between sending (Argentina) and receiving countries (see details in Results). Flows from the Falkland/Malvinas fisheries and the countries that they sell license to fish (e.g., Spain) are not shown because information were not available in open access databases.

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Teaching ML/AI for wider implementation and use, e.g. for a good governance ?!
Teaching ML/AI as a MOOC

Use automated Online ML/AI Grading tools

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
But where are our universities re. ML/AI?
But where are our public schools re. ML/AI?

Inference ?!

Field Data

Application/Management

ISIM 2021: ‘Machine Learning/AI and Open Access’ keynote by Falk Huettmann
Where one may be ...

https://www.ted.com/talks/garry_kasparov_don_t_fear_intelligent_machines_work_with_them/up-next?language=en
Thanks & Questions

Thanks to the organizers and good colleagues!