



Universiteit  
Leiden  
The Netherlands

## Massively collaborative machine learning

Rijn, J.N. van

### Citation

Rijn, J. N. van. (2016, December 19). *Massively collaborative machine learning*. IPA Dissertation Series. Retrieved from <https://hdl.handle.net/1887/44814>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/44814>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/44814> holds various files of this Leiden University dissertation

**Author:** Rijn, Jan van

**Title:** Massively collaborative machine learning

**Issue Date:** 2016-12-19

# Stellingen

behorende bij het proefschrift  
*Massively Collaborative Machine Learning*  
van Jan N. van Rijn

1. Meta-learning technieken kunnen rechtstreeks worden toegepast op verzamelde experimentele data. Dit kan tot veel nieuwe inzichten leiden (Hoofdstuk 4).
2. Meta-learning technieken kunnen bij het modelleren van data streams worden gebruikt om ensembles te maken die bestaan uit diverse soorten classificatie algoritmes (Hoofdstuk 5).
3. In de traditionele meta-learning omgeving zijn meta-features die iets over over de prestatie van een algoritme zeggen (zogenaamde landmarks) het meest voorspellend. Analoog hieraan zijn in de data stream omgeving landmarks met een tijdscomponent het meest voorspellend (Hoofdstuk 5).
4. Wanneer het doel van een algoritme selectie procedure zowel prestatie als efficiëntie bevat, moet de algoritme selectie procedure hier rekening mee houden om competitief te zijn, bijvoorbeeld door gebruik te maken van A3R (Hoofdstuk 6).
5. Alle output van Machine Learning onderzoek gefinancierd door belastinggeld, zou publiekelijk toegankelijk gemaakt moeten worden.
6. In tegenstelling tot veel andere wetenschappen, is onderzoek naar Machine Learning vaak toegespitst op het bedenken van nieuwe methoden en hoe deze presteren. In plaats daarvan zou het zich meer moeten focussen op bestaande methoden en onder welke omstandigheden deze goed presteren.
7. De ‘No Free Lunch Theorem’ (NFL) wordt vaak als argument gebruikt tegen meta-learning. Dit is onjuist, aangezien de uniformiteitsaannname die NFL maakt in de praktijk niet opgaat. Als dat wel het geval was, zouden mensen noch machines kunnen leren.
8. Wetenschappelijke samenwerking met commerciële partners kan alleen succesvol zijn als de verwachte uitkomst van zowel wetenschappelijke als commerciële waarde is.
9. De computationele complexiteit van een Rummukub puzzel is  $\mathcal{O}(n)$ . Het oplossen van zo een puzzel is hierdoor makkelijker dan het sorteren (in: J. N. van Rijn, F. W. Takes and J. K. Vis: The Complexity of Rummikub Problems, *Proceedings of the 27th Benelux Conference on Artificial Intelligence*).
10. Het belangrijkste doel van wetenschappelijke conferenties is netwerken, het succes hiervan wordt gemeten in de hoeveelheid nieuwe samenwerkingen.

# Propositions

for the PhD thesis

*Massively Collaborative Machine Learning*

by Jan N. van Rijn

1. Meta-learning techniques can be directly applied to collaboratively generated Machine Learning experiments, leading to novel insights (Chapter 4).
2. In the data stream setting, standard meta-learning techniques can be applied to construct heterogeneous ensembles of classifiers (Chapter 5).
3. In the traditional batch meta-learning setting, performance estimating meta-features (also called landmarks) are the most predictive meta-features. Analogous to this, performance estimating meta-features with a temporal component are highly effective in the data stream setting (Chapter 5).
4. When the goal of an algorithm selection procedure is multi-objective, such as area under the loss-time curves, algorithm selection procedures must be aware of this (e.g., by using A3R) in order to be competitive (Chapter 6).
5. All output of Machine Learning research funded by public money should be made available to the public.
6. In contrast to many other sciences, Machine Learning research is heavily focused on creating new methods and reporting how well these perform. Instead, there should be more emphasis on utilizing existing methods and reporting under which conditions these perform well.
7. The ‘No Free Lunch Theorem’ (NFL) is often used erroneously as an argument against meta-learning, as the uniformity assumption that NFL makes does not hold. If it did hold, there would be no learning at all, for humans nor machines.
8. Scientific collaborations with commercial partners can only be successful if the expected output is both of scientific and commercial value.
9. The computational complexity of solving a Rummikub puzzle is  $\mathcal{O}(n)$ . Effectively, solving this puzzle is easier than sorting (in: J. N. van Rijn, F. W. Takes and J. K. Vis: The Complexity of Rummikub Problems, *Proceedings of the 27th Benelux Conference on Artificial Intelligence*).
10. The most important purpose of scientific conferences is networking, and the most important output is measured in newly forged collaborations.