Transforming Wikipedia into a Search Engine for Local Experts

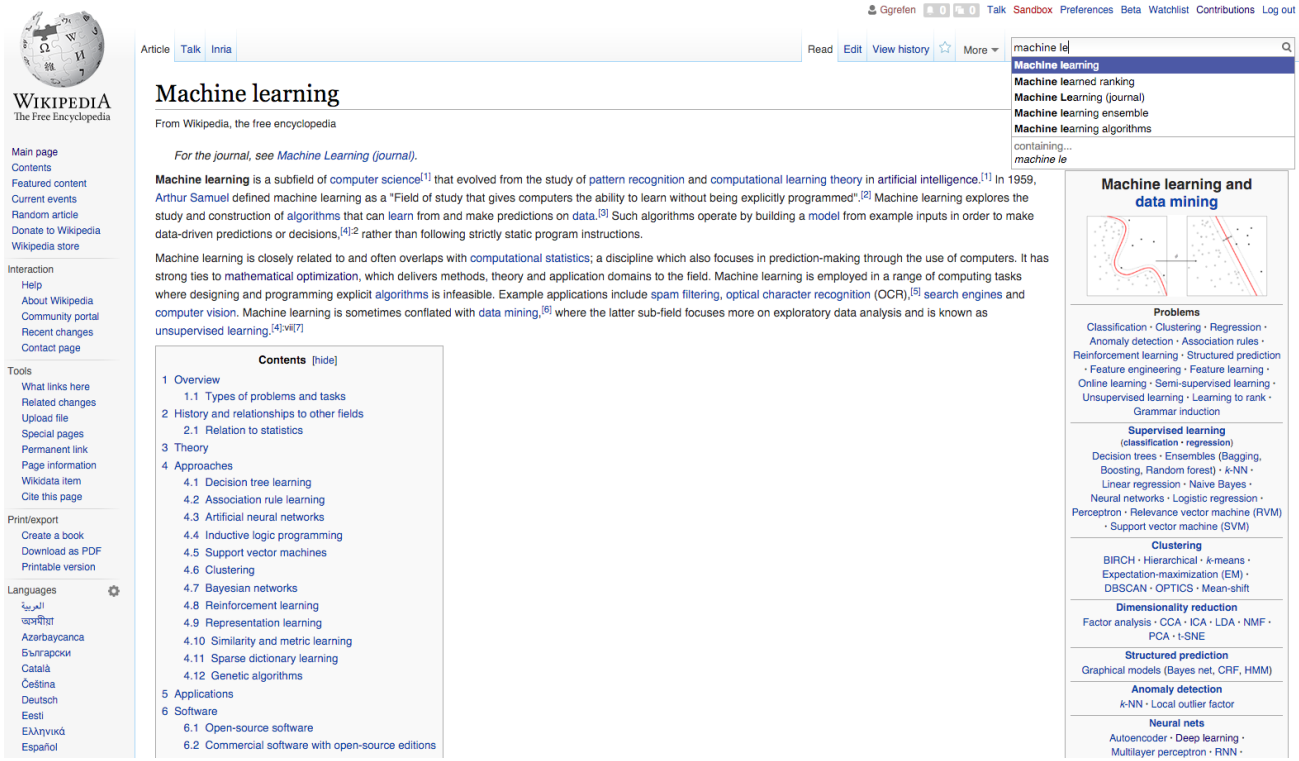
Gregory Grefenstette

WikiData Day

Febraury 2, 2016

Inria

Familiar way to look for knowledge



Ggrefen 0 0 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk Inria | Read Edit View history More | machine le

Machine learning
Machine learned ranking
Machine Learning (journal)
Machine learning ensemble
Machine learning algorithms
containing...
machine le

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
العربية
অসমীয়া
Azərbaycanca
Български
Català
Čeština
Deutsch
Eesti
Ελληνικά
Español

Machine learning

From Wikipedia, the free encyclopedia

For the journal, see Machine Learning (journal).

Machine learning is a subfield of computer science[1] that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.[1] In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed".[2] Machine learning explores the study and construction of algorithms that can learn from and make predictions on data.[3] Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions,[4]:2 rather than following strictly static program instructions.

Machine learning is closely related to and often overlaps with computational statistics; a discipline which also focuses in prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible. Example applications include spam filtering, optical character recognition (OCR),[5] search engines and computer vision. Machine learning is sometimes conflated with data mining,[6] where the latter sub-field focuses more on exploratory data analysis and is known as unsupervised learning.[4]:vii[7]

Contents [hide]

1 Overview
  1.1 Types of problems and tasks
2 History and relationships to other fields
  2.1 Relation to statistics
3 Theory
4 Approaches
  4.1 Decision tree learning
  4.2 Association rule learning
  4.3 Artificial neural networks
  4.4 Inductive logic programming
  4.5 Support vector machines
  4.6 Clustering
  4.7 Bayesian networks
  4.8 Reinforcement learning
  4.9 Representation learning
  4.10 Similarity and metric learning
  4.11 Sparse dictionary learning
  4.12 Genetic algorithms
5 Applications
6 Software
  6.1 Open-source software
  6.2 Commercial software with open-source editions

Machine learning and data mining

Problems
Classification · Clustering · Regression · Anomaly detection · Association rules · Reinforcement learning · Structured prediction · Feature engineering · Feature learning · Online learning · Semi-supervised learning · Unsupervised learning · Learning to rank · Grammar induction

Supervised learning (classification · regression)
Decision trees · Ensembles (Bagging, Boosting, Random forest) · *k*-NN · Linear regression · Naive Bayes · Neural networks · Logistic regression · Perceptron · Relevance vector machine (RVM) · Support vector machine (SVM)

Clustering
BIRCH · Hierarchical · *k*-means · Expectation-maximization (EM) · DBSCAN · OPTICS · Mean-shift

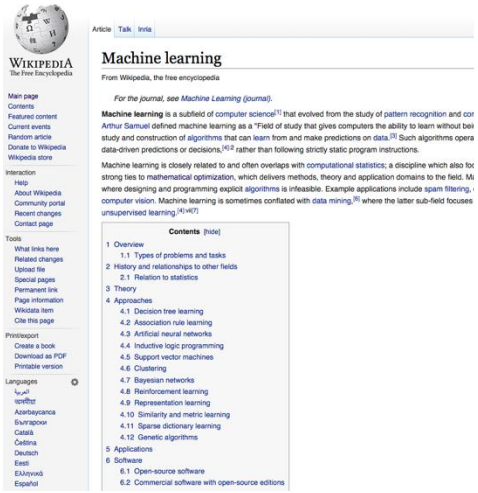
Dimensionality reduction
Factor analysis · CCA · ICA · LDA · NMF · PCA · t-SNE

Structured prediction
Graphical models (Bayes net, CRF, HMM)

Anomaly detection
k-NN · Local outlier factor

Neural nets
Autoencoder · Deep learning · Multilayer perceptron · RNN ·

Unfamiliar knowledge source



informatics mathematics

Explo-Raweb

Explo-RAweb is an experimental browsing application that allows to navigate within the activity reports of the Inria's research teams.

This application is not updated. In particular, it doesn't contain the highlighting of the "Best-papers"

Accès Intranet

PDF par sections

Infos pratiques

Contact:
- Direction générale déléguée à la science

Assistance technique:
- DSI-SEISM

Archives 2013 | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 | 1995 | 1994

Inria - 2014 Teams Activity Report

The Inria's Research Teams produce an annual Activity Report presenting their activities and their results of the year. This reports includes the m
by the team and the new results of the year. The report also describes the grants and contracts and the activities of dissemination and teaching. F

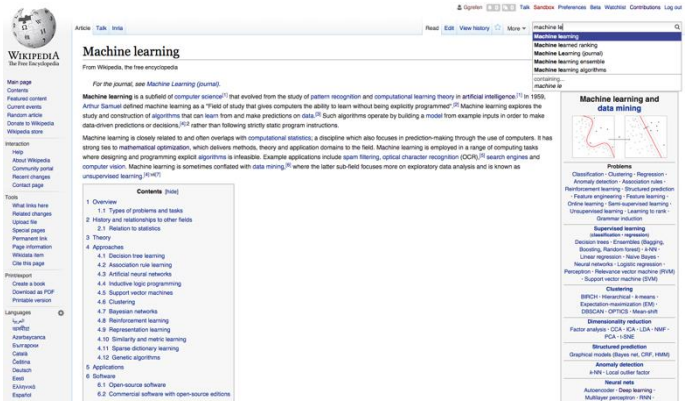
A B C D E F G H I K L M N O P Q R S T U V W Z

Abs - Algorithms, Biology, Structure
Alf - Amdahl's Law is Forever
Algorille - Algorithms for the Grid
Alice - Geometry and Lighting
Alpage - Large-scale deep linguistic processing
Alpines - Algorithms and parallel tools for integrated numerical simulations
Amib - Algorithms and Models for Integrative Biology
Ange - Numerical Analysis, Geophysics and Ecology
Antique - Analyse Statique par Interprétation Abstraite
Aoste - Models and methods of analysis and optimization for systems with real-time and embedding constraints
Apics - Analysis and Problems of Inverse type in Control and Signal processing
Aramis - Algorithms, models and methods for images and signals of the human brain
Aric - Arithmetic and Computing
Asap - As Scalable As Possible: foundations of large scale dynamic distributed systems
Asclepios - Analysis and Simulation of Biomedical Images
Ascola - Aspect and composition languages
Aspi - Applications of interacting particle systems to statistics
Ateams - Analysis and Transformation based on rEliAble tool coMpositionS
Athena - Computational Imaging of the Central Nervous System
Atlanmod - Modeling Technologies for Software Production, Operation, and Evolution
Avalon - Algorithms and Software Architectures for Distributed and HPC Platforms
Aviz - Analysis and Visualization
Ayin - Models of spatio-temporal structure for high-resolution image processing

Bacchus - Parallel tools for Numerical Algorithms and Resolution of essentially Hyperbolic problems
Bamboo - An algorithmic view on genomes, cells, and environments
Beagle - Artificial Evolution and Computational Biology
Bigs - Biology, genetics and statistics
Biocore - Biological control of artificial ecosystems
Bipop - Modelling, Simulation, Control and Optimization of Non-Smooth Dynamical Systems
Bonsai - Bioinformatics and Sequence Analysis

WIKIPEDIA
The Free Encyclopedia

informatics mathematics
Inria



Explo-Raweb

Explo-RAweb is an experimental browsing application that allows to navigate within the activity reports of the Inria's research teams.

This application is not updated. In particular, it doesn't contain the highlighting of the "Best-papers"

Accès Intranet

PDF par sections

Infos pratiques

Contact:
- Direction générale déléguée à la science

Assistance technique:
- DSI-SEISM

Archives 2013 | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 | 2000 | 1999 | 1998 | 1997 | 1996 | 1995 | 1994

Inria - 2014 Teams Activity Report

The Inria's Research Teams produce an annual Activity Report presenting their activities and their results of the year. This reports includes the m
by the team and the new results of the year. The report also describes the grants and contracts and the activities of dissemination and teaching. F

A B C D E F G H I K L M N O P Q R S T U V W Z

Abs - Algorithms, Biology, Structure
Alf - Amdahl's Law is Forever
Algorille - Algorithms for the Grid
Alice - Geometry and Lighting
Alpage - Large-scale deep linguistic processing
Alpines - Algorithms and parallel tools for integrated numerical simulations
Amib - Algorithms and Models for Integrative Biology
Ange - Numerical Analysis, Geophysics and Ecology
Antique - Analyse Statique par Interprétation Abstraite
Aoste - Models and methods of analysis and optimization for systems with real-time and embedding constraints
Apics - Analysis and Problems of Inverse type in Control and Signal processing
Aramis - Algorithms, models and methods for images and signals of the human brain
Aric - Arithmetic and Computing
Asap - As Scalable As Possible: foundations of large scale dynamic distributed systems
Asclepios - Analysis and Simulation of Biomedical Images
Ascola - Aspect and composition languages
Aspi - Applications of interacting particle systems to statistics
Ateams - Analysis and Transformation based on rElIAble tool coMpositionS
Athena - Computational Imaging of the Central Nervous System
Atlanmod - Modeling Technologies for Software Production, Operation, and Evolution
Avalon - Algorithms and Software Architectures for Distributed and HPC Platforms
Aviz - Analysis and Visualization
Ayin - Models of spatio-temporal structure for high-resolution image processing

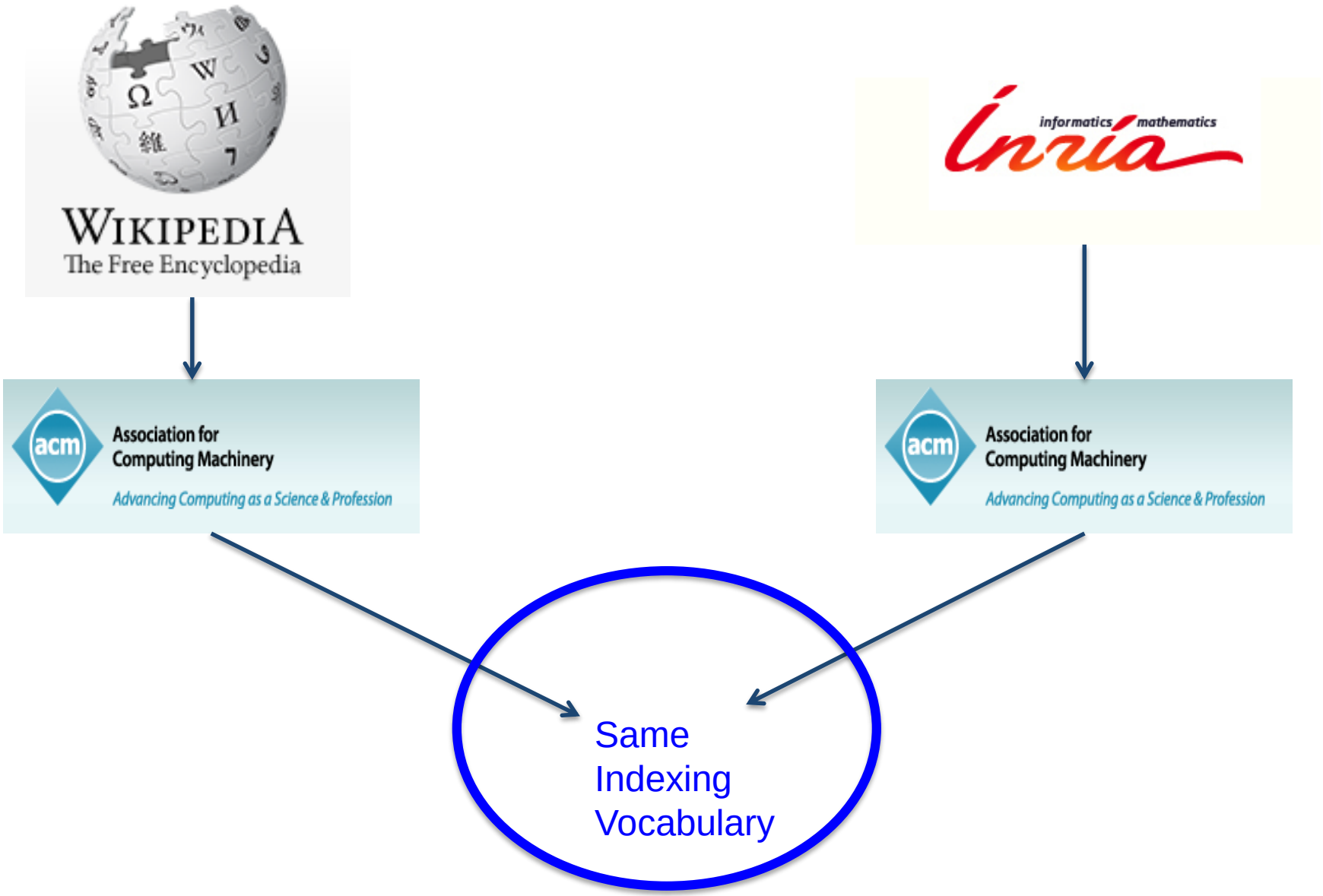
Bacchus - Parallel tools for Numerical Algorithms and Resolution of essentially Hyperbolic problems
Bamboo - An algorithmic view on genomes, cells, and environments
Beagle - Artificial Evolution and Computational Biology
Bigs - Biology, genetics and statistics
Biocore - Biological control of artificial ecosystems
Bipop - Modelling, Simulation, Control and Optimization of Non-Smooth Dynamical Systems
Bonsai - Bioinformatics and Sequence Analysis

Domain vocabulary - pivot

Association for Computing Machinery
Advancing Computing as a Science & Profession

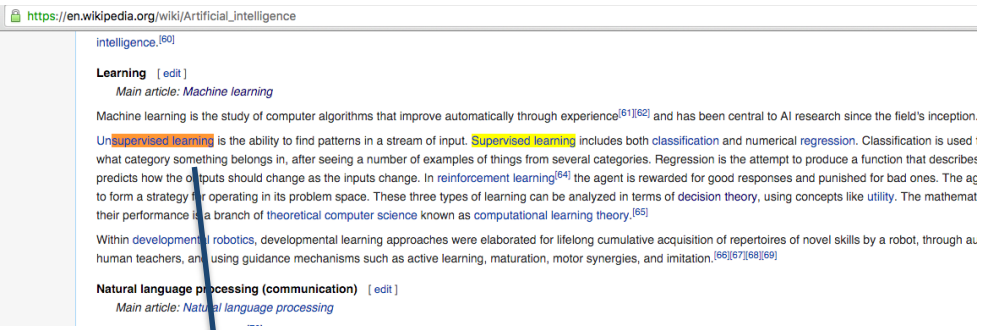
CCS 2012 - Table of Contents

- General and reference
  - Document types
    - Surveys and overviews
    - Reference works
    - General conference proceedings
    - Biographies
    - General literature
    - Computing standards, RFCs and guidelines
  - Cross-computing tools and techniques
    - Reliability
    - Empirical studies
    - Measurement
    - Metrics
    - Evaluation
    - Experimentation
    - Estimation
    - Design
    - Performance
    - Validation
    - Verification
- Hardware
  - Printed circuit boards
    - Electromagnetic interference and compatibility
    - PCB design and layout
  - Communication hardware, interfaces and storage
    - Signal processing systems
      - Digital signal processing
      - Beamforming
      - Noise reduction
    - Sensors and actuators
    - Buses and high-speed links
    - Distributed programming languages
  - Concurrent computing methodologies
    - Concurrent programming languages
    - Concurrent algorithms
- Applied computing
  - Electronic commerce
    - Digital cash
    - E-commerce infrastructure
    - Electronic data interchange
    - Electronic funds transfer
    - Online shopping
    - Online banking
    - Secure online transactions
    - Online auctions
  - Enterprise computing
    - Enterprise information systems
      - Intranets
      - Extranets
      - Enterprise resource planning
      - Enterprise applications
      - Data centers
    - Business process management
      - Business process modeling
      - Business process management systems
      - Business process monitoring
      - Cross-organizational business processes
      - Business intelligence
    - Enterprise architectures
      - Enterprise architecture management
      - Enterprise architecture frameworks
      - Enterprise architecture modeling
    - Service-oriented architectures
    - Event-driven architectures
    - Business rules
    - Enterprise modeling
    - Enterprise ontologies, taxonomies and vocabularies
    - Enterprise data management
    - Reference models
    - Business-IT alignment
    - IT architectures
    - IT governance
    - Enterprise computing infrastructures
    - Enterprise interoperability
      - Enterprise application integration
      - Information integration and interoperability
  - Physical sciences and engineering
    - Aerospace
      - Avionics
    - Archaeology


WIKIPEDIA
The Free Encyclopedia
informatics mathematics
Inria
acm
Association for
Computing Machinery
Advancing Computing as a Science & Profession
acm
Association for
Computing Machinery
Advancing Computing as a Science & Profession
Same
Indexing
Vocabulary

Inria report from the TAI team
http://raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html

Artificial intelligence article in Wikipedia


https://en.wikipedia.org/wiki/Artificial_intelligence

intelligence.[60]

Learning [edit]

Main article: Machine learning

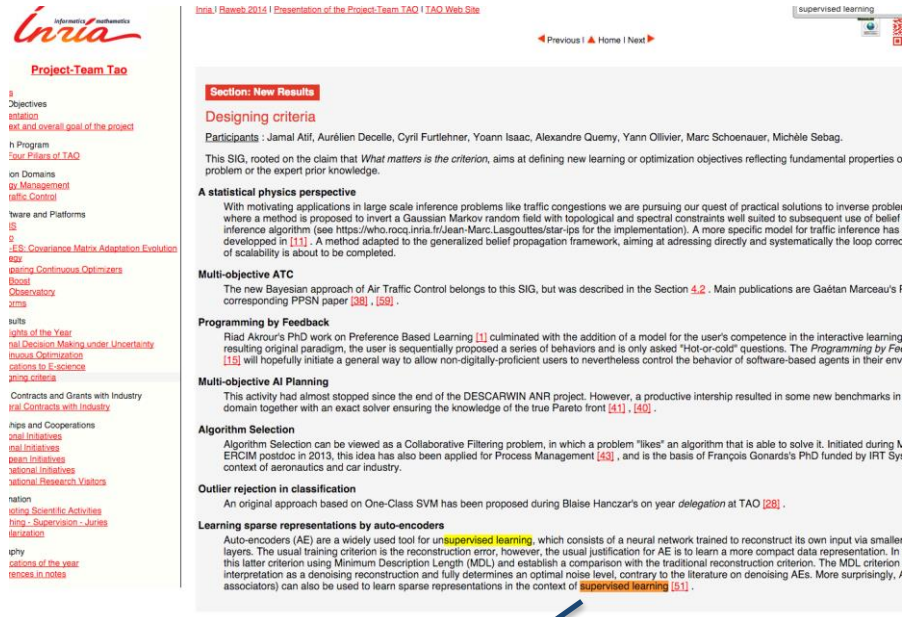
Machine learning is the study of computer algorithms that improve automatically through experience[61][62] and has been central to AI research since the field's inception.

Unsupervised learning is the ability to find patterns in a stream of input. Supervised learning includes both classification and numerical regression. Classification is used to determine what category something belongs in, after seeing a number of examples of things from several categories. Regression is the attempt to produce a function that describes the relationship between inputs and outputs and predicts how the outputs should change as the inputs change. In reinforcement learning[64] the agent is rewarded for good responses and punished for bad ones. The agent uses this sequence of rewards and punishments to form a strategy for operating in its problem space. These three types of learning can be analyzed in terms of decision theory, using concepts like utility. The mathematical analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory.[65]

Within developmental robotics, developmental learning approaches were elaborated for lifelong cumulative acquisition of repertoires of novel skills by a robot, through autonomous self-exploration and social interaction with human teachers, and using guidance mechanisms such as active learning, maturation, motor synergies, and imitation.[66][67][68][69]

Natural language processing (communication) [edit]

Main article: Natural language processing


Inria | Raweb 2014 | Presentation of the Project-Team TAO | TAO Web Site

Previous | Home | Next

supervised learning

Project-Team Tao

Section: New Results

Designing criteria

Participants : Jamal Atif, Aurélien Decelle, Cyril Furtlehner, Yoann Isaac, Alexandre Quemy, Yann Ollivier, Marc Schoenauer, Michèle Sebag.

This SIG, rooted on the claim that *What matters is the criterion*, aims at defining new learning or optimization objectives reflecting fundamental properties of the problem or the expert prior knowledge.

A statistical physics perspective

With motivating applications in large scale inference problems like traffic congestions we are pursuing our quest of practical solutions to inverse problems where a method is proposed to invert a Gaussian Markov random field with topological and spectral constraints well suited to subsequent use of belief inference algorithm (see https://who.rocq.inria.fr/Jean-Marc.Lasgouttes/star-ips for the implementation). A more specific model for traffic inference has been developped in [11] . A method adapted to the generalized belief propagation framework, aiming at adressing directly and systematically the loop correc of scalability is about to be completed.

Multi-objective ATC

The new Bayesian approach of Air Traffic Control belongs to this SIG, but was described in the Section 4.2 . Main publications are Gaëtan Marceau's corresponding PPSN paper [38] , [59] .

Programming by Feedback

Riad Akrour's PhD work on Preference Based Learning [1] culminated with the addition of a model for the user's competence in the interactive learning resulting original paradigm, the user is sequentially proposed a series of behaviors and is only asked "Hot-or-cold" questions. The *Programming by Fe* [15] will hopefully initiate a general way to allow non-digitally-proficient users to nevertheless control the behavior of software-based agents in their env

Multi-objective AI Planning

This activity had almost stopped since the end of the DESCARWIN ANR project. However, a productive intership resulted in some new benchmarks in domain together with an exact solver ensuring the knowledge of the true Pareto front [41] , [40] .

Algorithm Selection

Algorithm Selection can be viewed as a Collaborative Filtering problem, in which a problem "likes" an algorithm that is able to solve it. Initiated during ERCIM postdoc in 2013, this idea has also been applied for Process Management [43] , and is the basis of François Gonards's PhD funded by IRT Sy context of aeronautics and car industry.

Outlier rejection in classification

An original approach based on One-Class SVM has been proposed during Blaise Hanczar's on year *delegation* at TAO [28] .

Learning sparse representations by auto-encoders

Auto-encoders (AE) are a widely used tool for unsupervised learning, which consists of a neural network trained to reconstruct its own input via smaller layers. The usual training criterion is the reconstruction error, however, the usual justification for AE is to learn a more compact data representation. In this latter criterion using Minimum Description Length (MDL) and establish a comparison with the traditional reconstruction criterion. The MDL criterion interpretation as a denoising reconstruction and fully determines an optimal noise level, contrary to the literature on denoising AEs. More surprisingly, associators) can also be used to learn sparse representations in the context of supervised learning [61] .

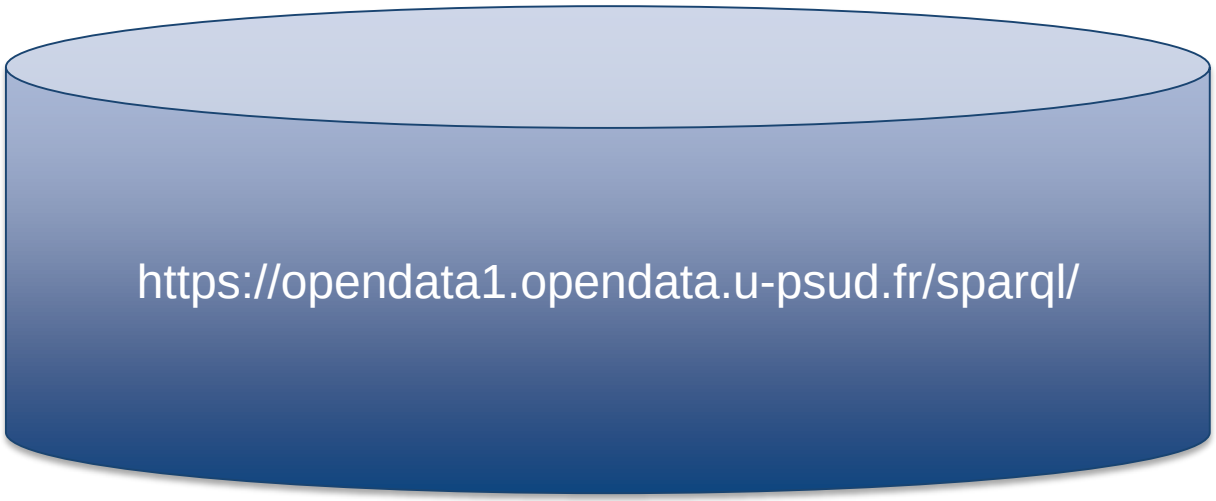
ACM vocabulary

```
<skos:Concept rdf:about="#10010259" xml:lang="en">
<skos:prefLabel xml:lang="en">Supervised learning</skos:prefLabel>
<skos:inScheme rdf:resource="http://totem.semedica.com/taxonomy/The ACM Computing Classification System (CCS)"/>
<skos:broader rdf:resource="#10010258"/>
<skos:narrower rdf:resource="#10003268"/>
<skos:narrower rdf:resource="#10003343"/>
<skos:narrower rdf:resource="#10010263"/>
<skos:narrower rdf:resource="#10010264"/>
<skos:narrower rdf:resource="#10010265"/>
<skos:narrower rdf:resource="#10010266"/>
</skos:Concept>
```

Artificial intelligence : Supervised learning : http://raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html : TAO

Artificial intelligence : Supervised learning : http://raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html : TAO

<http://dl.acm.org/buildccscode.cfm?id=10010259&lid=f> a <http://dl.acm.org/ontology#code> ;
rdfs:label '"Supervised learning"' ; wdt:P2179 10010259 .
<https://en.wikipedia.org/w/index.php?title="Artificial_intelligence"> rdfs:label '"Artificial intelligence"' ;
wdt:P2179 10010259 ;
rdfs:seeAlso <http://raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html> .
<http://raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html> a inria:team ;
rdfs:label '"TAO"' ; rdfs:comment '"Machine Learning and Optimisation"' ;
wdt:P2179 10010259 .



https://meta.wikimedia.org/wiki/Special:MyPage/global.js

```
// [[d:User:Ggrefen/DisplayInriaTeam2.js]]
mw.loader.load("//www.wikidata.org/w/index.php?title=User:Ggrefen/DisplayInriaTeam2.js&action=raw&ctype=text/javascript");
```

WIKIPEDIA
The Free Encyclopedia

Article Talk Inria

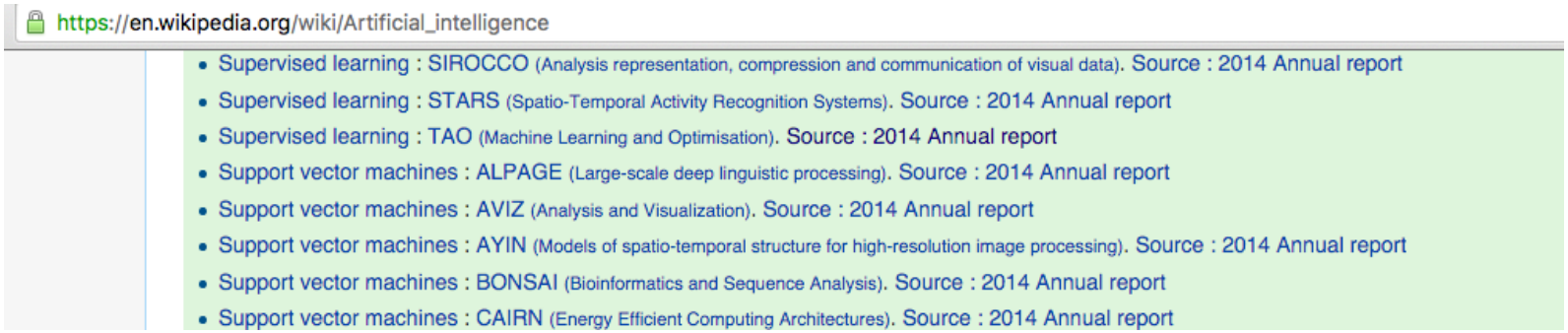
Artificial intelligence

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article

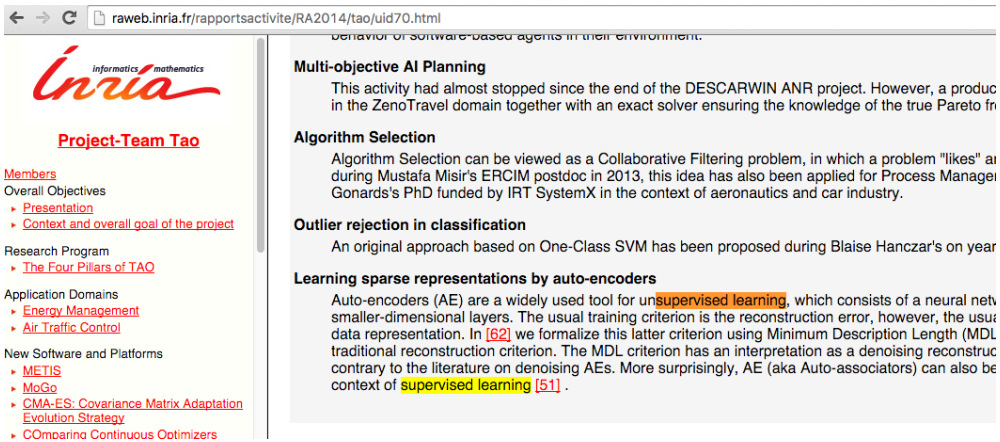
"AI" redirects here. For other uses, see Ai and Artificial intelligence (disambiguation)

Artificial intelligence (**AI**) is the intelligence exhibited by machines or software. It is als
intelligent behavior. Major AI researchers and textbooks define this field as "the study a
maximize its chances of success.[2] John McCarthy, who coined the term in 1955,[3] def



https://en.wikipedia.org/wiki/Artificial_intelligence

- Supervised learning : SIROCCO (Analysis representation, compression and communication of visual data). Source : 2014 Annual report
- Supervised learning : STARS (Spatio-Temporal Activity Recognition Systems). Source : 2014 Annual report
- Supervised learning : TAO (Machine Learning and Optimisation). Source : 2014 Annual report
- Support vector machines : ALPAGE (Large-scale deep linguistic processing). Source : 2014 Annual report
- Support vector machines : AVIZ (Analysis and Visualization). Source : 2014 Annual report
- Support vector machines : AYIN (Models of spatio-temporal structure for high-resolution image processing). Source : 2014 Annual report
- Support vector machines : BONSAI (Bioinformatics and Sequence Analysis). Source : 2014 Annual report
- Support vector machines : CAIRN (Energy Efficient Computing Architectures). Source : 2014 Annual report



raweb.inria.fr/rapportsactivite/RA2014/tao/uid70.html

Project-Team Tao

Members
Overall Objectives
- Presentation
- Context and overall goal of the project

Research Program
- The Four Pillars of TAO

Application Domains
- Energy Management
- Air Traffic Control

New Software and Platforms
- METIS
- MoGo
- CMA-ES: Covariance Matrix Adaptation Evolution Strategy
- COmparing Continuous Optimizers

Multi-objective AI Planning

This activity had almost stopped since the end of the DESCARWIN ANR project. However, a produc
in the ZenoTravel domain together with an exact solver ensuring the knowledge of the true Pareto fr

Algorithm Selection

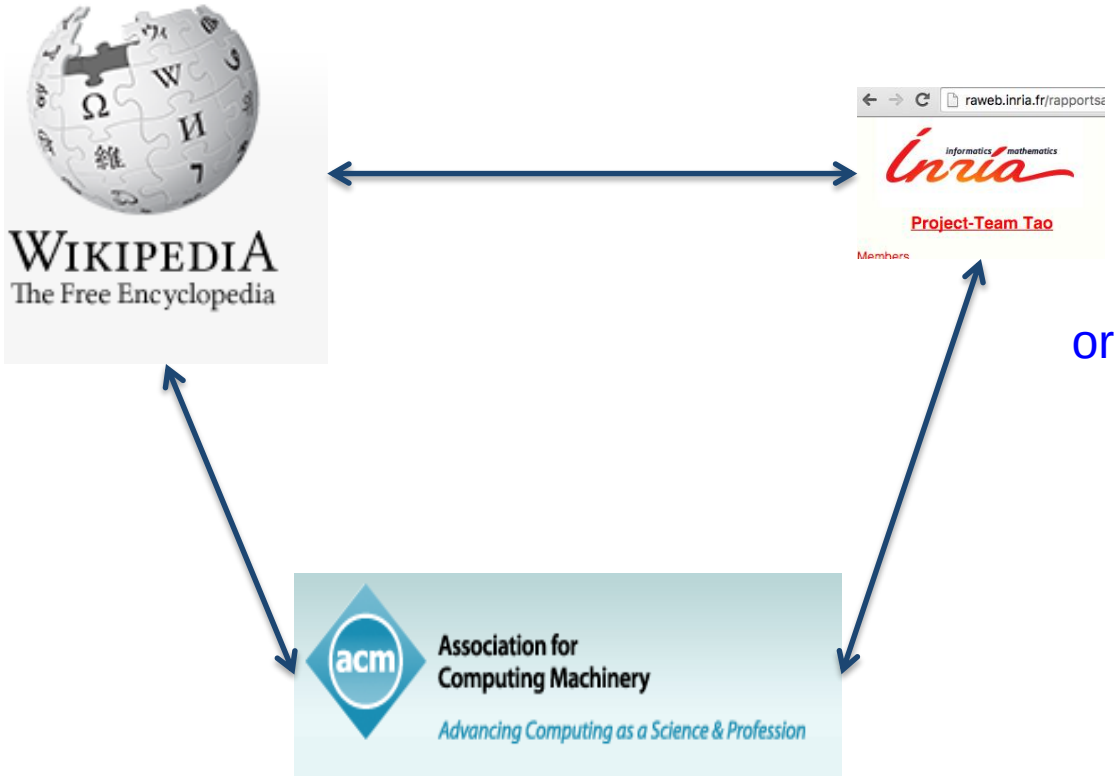
Algorithm Selection can be viewed as a Collaborative Filtering problem, in which a problem "likes" a
during Mustafa Misir's ERCIM postdoc in 2013, this idea has also been applied for Process Manage
Gonards's PhD funded by IRT SystemX in the context of aeronautics and car industry.

Outlier rejection in classification

An original approach based on One-Class SVM has been proposed during Blaise Hanczar's on year

Learning sparse representations by auto-encoders

Auto-encoders (AE) are a widely used tool for unsupervised learning, which consists of a neural net
smaller-dimensional layers. The usual training criterion is the reconstruction error, however, the usu
data representation. In [62] we formalize this latter criterion using Minimum Description Length (MDL
traditional reconstruction criterion. The MDL criterion has an interpretation as a denoising reconstruc
contrary to the literature on denoising AEs. More surprisingly, AE (aka Auto-associators) can also be
context of supervised learning [51] .



or any Corpus of text

or any domain vocabulary

Thank you !

www.inria.fr