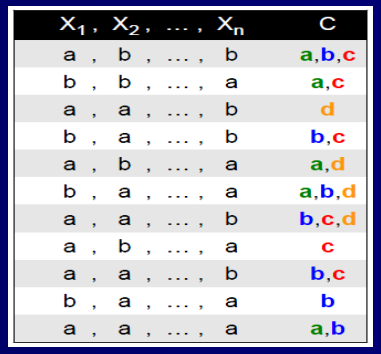
MACHINE LEARNING

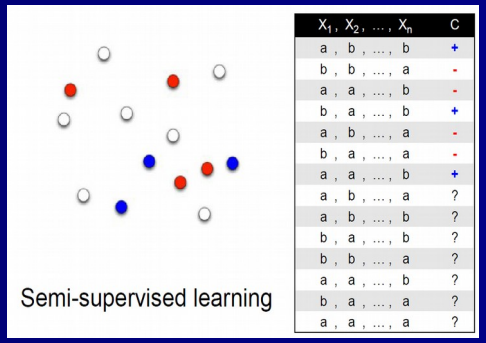
PRINCIPAL LEARNING SCENARIOS

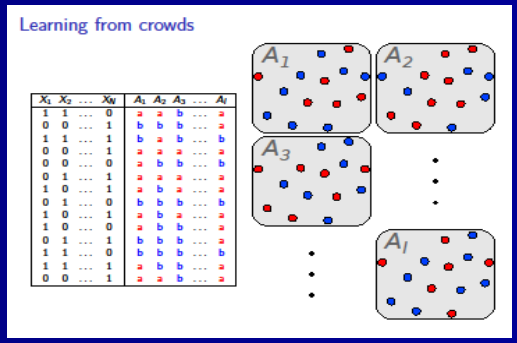
– NLP ORIENTED –

Iñaki Inza
Intelligent Systems Group, www.sc.ehu.es/isg
Computer Science Faculty
University of the Basque Country, Donostia - San Sebastian



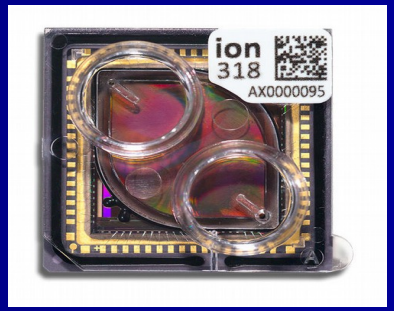
X_1 , X_2 , ... , X_n	C
a , b , ... , b	a,b,c
b , b , ... , a	a,c
a , a , ... , b	d
b , a , ... , b	b,c
a , b , ... , a	a,d
b , a , ... , a	a,b,d
a , a , ... , b	b,c,d
a , b , ... , a	c
a , a , ... , b	b,c
b , a , ... , a	b
a , a , ... , a	a,b





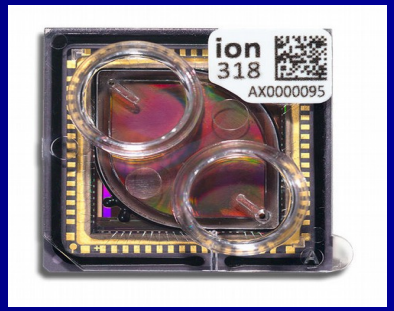
THE AGE OF DATA

- Data collected at enormous speeds in **everyday practice**:
  - social networks' activity
  - electronic purchases and transactions
  - monitored patients
  - bioinformatics gene expression data
  - traffic and people movement
  - Internet of Things
  - text everywhere
  - ...



THE AGE OF DATA

- Computers and storage systems have become **cheaper** and more **powerful**
- World's technological capacity to store info: **doubled** every 40 months since 80's
- Since 90's, much more data is being stored **than** analyzed (around 5-10%)
- By 2020 estimated... 30.6 exabytes through networks, 35 trillion GB stored, 11.6 billion connected devices
- **Traditiona**l data analysis techniques (70s-80s) **unfeasible** for "modern" data



THE AGE OF DATA

- Computers and storage systems have become **cheaper** and more **powerful**

- Hidden big data. Large quantities of useful data are in fact useless because they are untagged, file-based, and unstructured. The 2012 IDC study on big data [117] explained that, in 2012, 23% (643 exabytes) of the digital universe would be useful for big data if tagged and analyzed. However, at that time only 3% of the potentially useful data was tagged, and even less was analyzed. The figures have probably gotten worse in recent years. The Open Data and Semantic Web movements have emerged, in part, to make us aware and improve on this situation. No comments

- **Traditional** data analysis techniques (70s-80s) **unfeasible** for "modern" data

ARTIFICIAL INTELLIGENCE ~ DATA

"Almost current 95% of AI is based on machine learning from data"

Data: a cornerstone for AI – Toward a Common European Data Space

> *" Good quality shared data is essential to develop socially responsive AI "*

For an application of artificial intelligence (AI) to be ready for market entry it has to learn on the basis of training data. Once in use on the market, it should generate a sufficient amount of data as part of its use.

ARTIFICIAL INTELLIGENCE ~ DATA

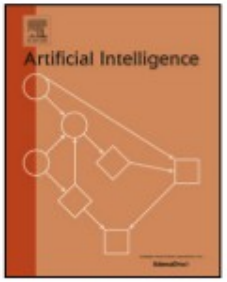
"Almost current 95% of AI is based on machine learning from data"

Artificial Intelligence 289 (2020) 103386

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



Artificial Intelligence requires more than deep learning – but what, exactly?

Michael Wooldridge

Department of Computer Science, University of Oxford, United Kingdom of Great Britain and Northern Ireland

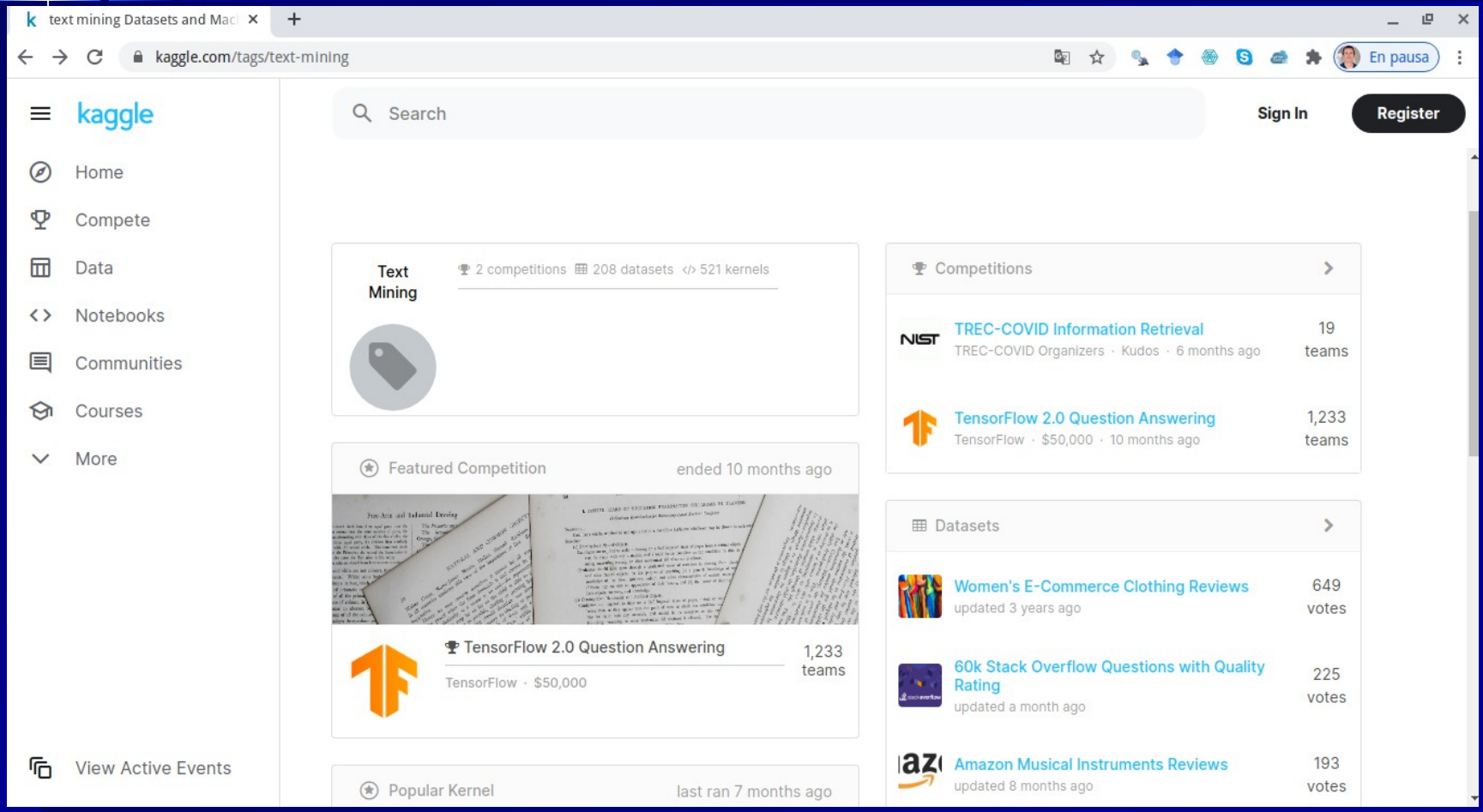
KAGGLE.COM

	**Online Product Sales** Predict the online sales of a consumer product based on a data set of product features.	$22,500	365	14 months ago
Boehringer Ingelheim	**Predicting a Biological Response** Predict a biological response of molecules from their chemical properties	$20,000	703	14 months ago
Ford	**Stay Alert! The Ford Challenge** Driving while not alert can be deadly. The objective is to design a classifier that will detect whether the driver is alert or not alert, employing data that are acquired while driving.	$950	176	2 years ago

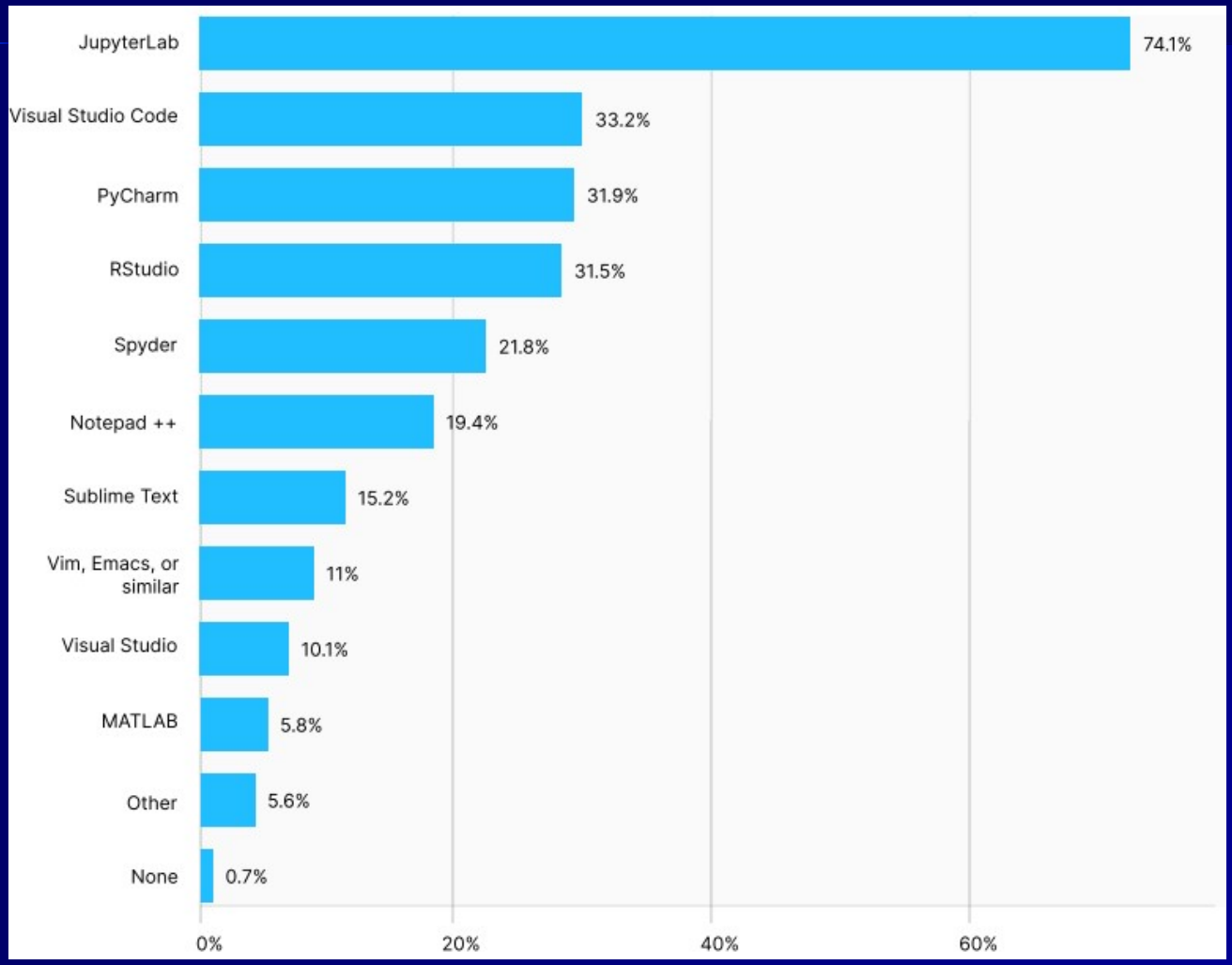
Kaggle competitions

Kaggle Open Datasets

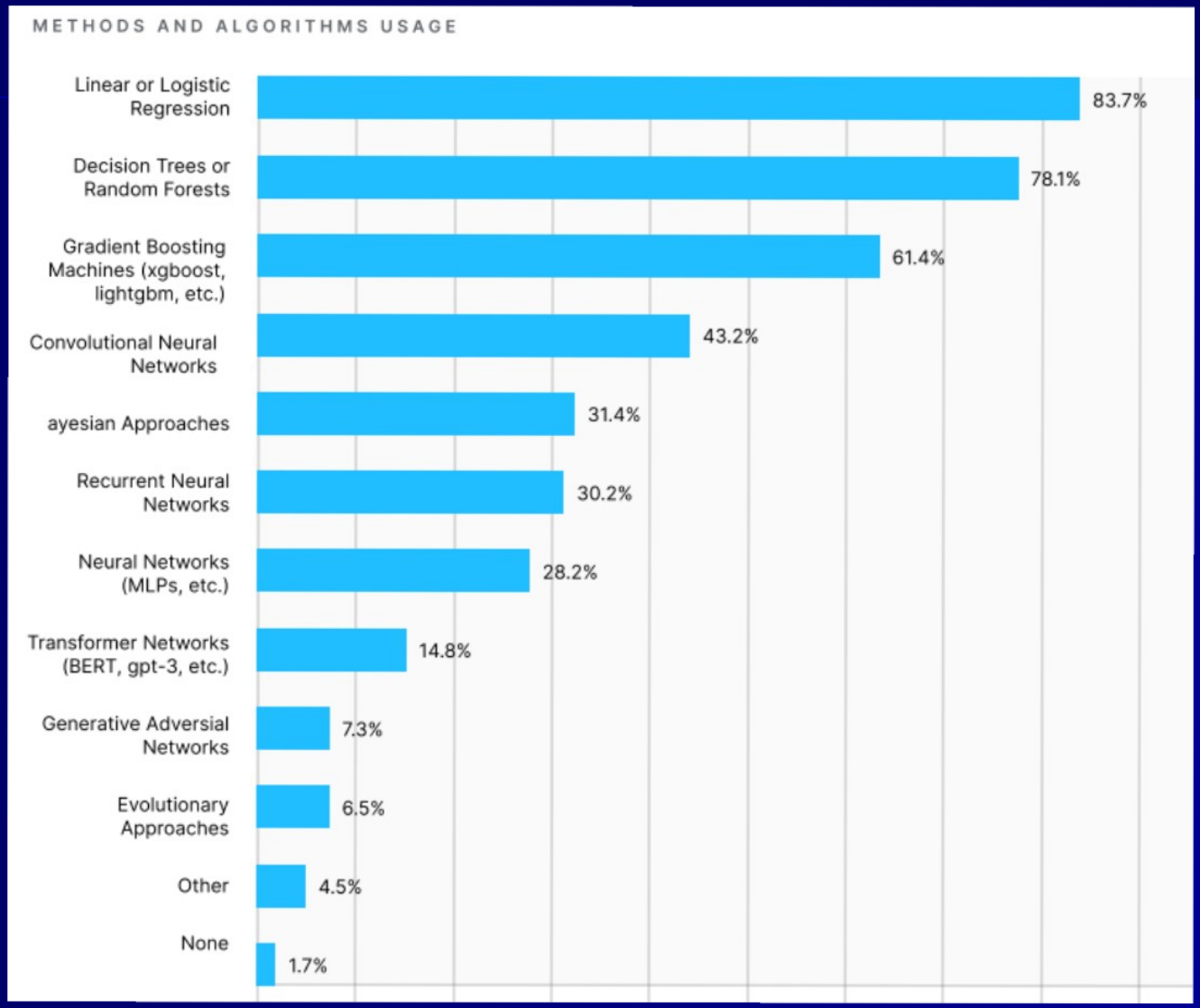
KAGGLE.COM
TEXT-MINING



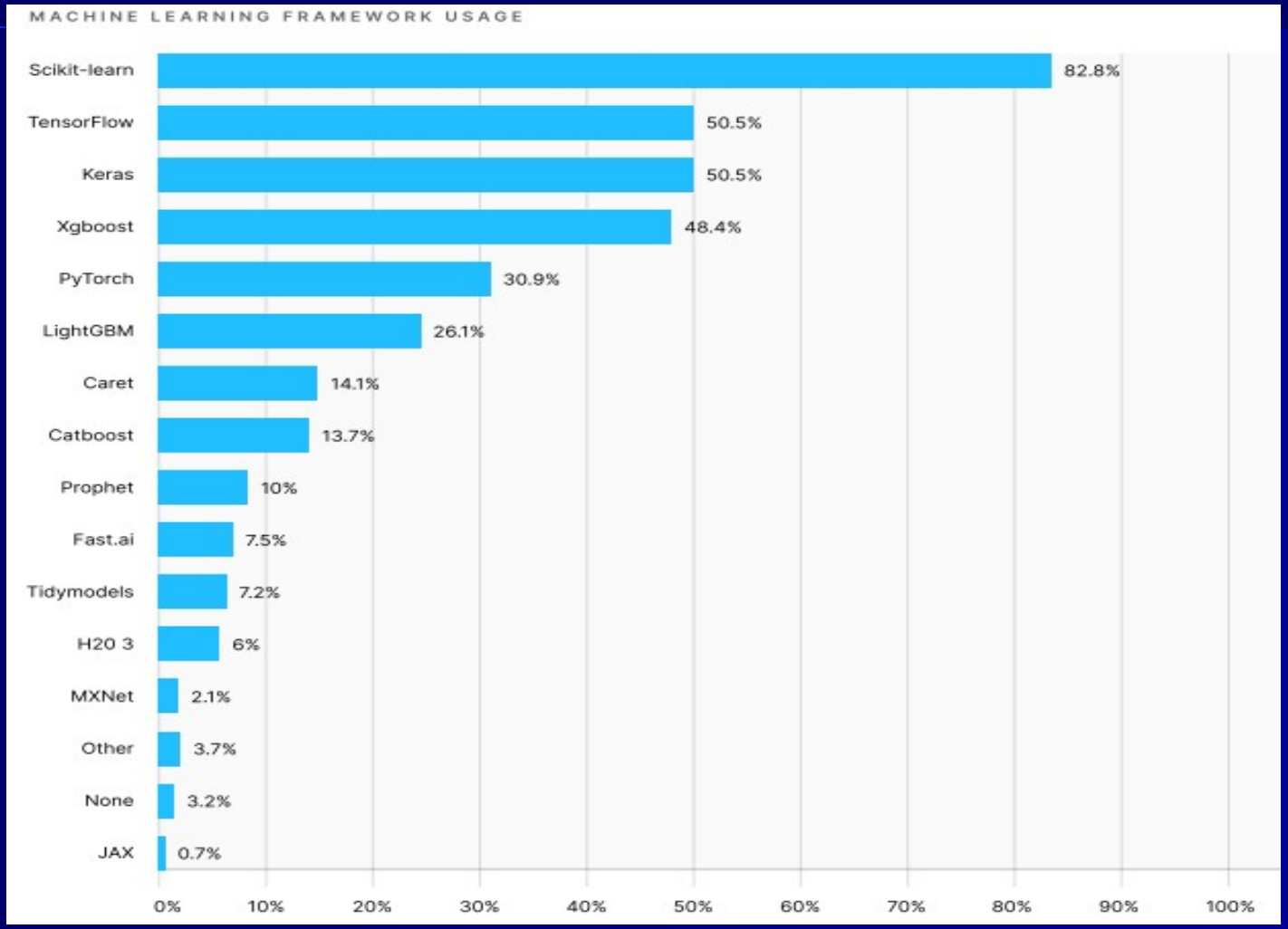
KAGGLE 2020 SURVEY



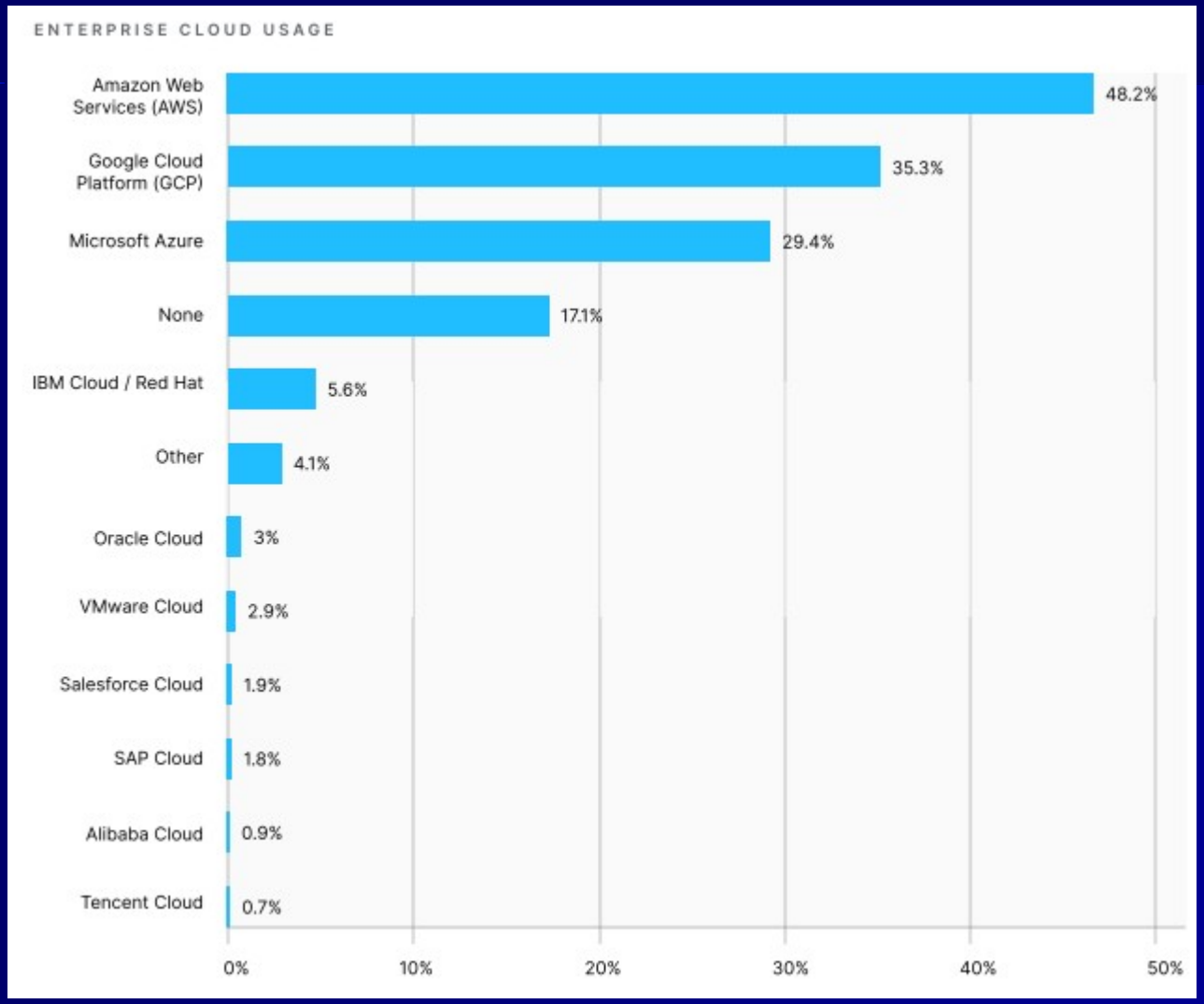
KAGGLE 2020 SURVEY



KAGGLE 2020 SURVEY



KAGGLE 2020 SURVEY



KEY RESOURCE
DATA MINING IN BUSINESS

POLLS' RESULTS
kdnuggets.com

- Top machine learning methods used
- Application field
- Analyzed data types / sources
- Primary programming language for data mining
- Used software tools
- Complete list of polls
- kaggle.com's survey: the state of data science'2017

DATA SCIENCE FOR A BETTER WORLD

- UN initiative
- Data: phones, meteo, networks...

- Humanitarian emergencies
- Pandemic diseases
- Sustainable development

- Quick answer
- Global answer

UNITED NATIONS GLOBAL PULSE

Harnessing big data for development and humanitarian action

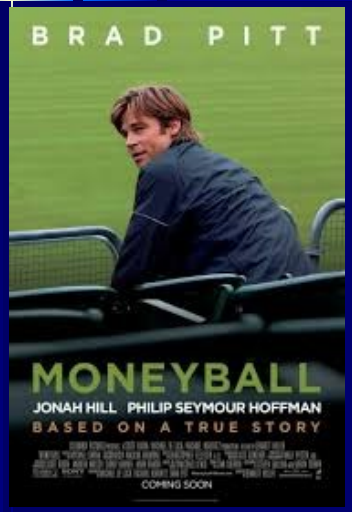
Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data. Its vision is a future in which big data is harnessed safely and responsibly as a public good. Its mission is to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action.
The initiative was established based on a recognition that digital data offers the opportunity to gain a better understanding of changes in human well-being, and to get real-time feedback on how well policy responses are working.

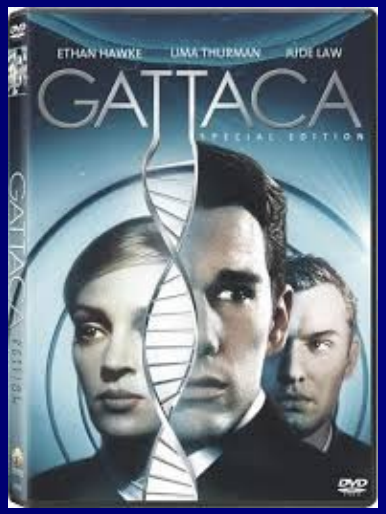
To this end, Global Pulse is working to promote awareness of the opportunities Big Data presents for sustainable development and humanitarian action, forge public-private data sharing partnerships, generate high-impact analytical tools and approaches through its network of Pulse Labs, and drive broad adoption of useful innovations across the UN System.

DATA-POP ALLIANCE

Data-Pop Alliance is a global coalition on Big Data and development created by the Harvard Humanitarian Initiative, MIT Media Lab, and Overseas Development Institute that brings together researchers, experts, practitioners, and activists to promote a people-centered Big Data revolution through collaborative research, capacity building, and community engagement. As of February 2016, Flowminder Foundation has joined Data-Pop Alliance as its fourth Core Member.

FILMS AND "AI"


BRAD PITT
MONEYBALL
JONAH HILL PHILIP SEYMOUR HOFFMAN
BASED ON A TRUE STORY
COMING SOON


ETHAN HAWKE UMA THURMAN JUDE LAW
GATTACA
SPECIAL EDITION
DVD


DESCIFRAR EL CÓDIGO. GANAR LA GUERRA
★★★★★ INTENSO, IMPECABLE, ÚNICO. DE ÓSCAR. VARIETY
★★★★★ GANADORA Y CONMOVEDORA. UNA GRAN PELÍCULA. HOLLYWOOD REPORTER
★★★★★ UNA MAGISTRAL INTERPRETACIÓN DE BENEDICT CUMBERBATCH. NEW YORK POST
EXCEPCIONAL. ELLE
★★★★★ THE INDEPENDENT
★★★★★ GLAMOUR
THE IMITATION GAME
BENEDICT CUMBERBATCH
KEIRA KNIGHTLEY
(DESCIFRANDO ENIGMA)


JOAQUIN PHOENIX AMY ADAMS ROONEY MARA
OLIVIA WILDE SCARLETT JOHANSSON
her
A SPIKE JONZE LOVE STORY
IN CINEMAS FEBRUARY 14TH


28 DAYS LATER
ES GIBT NICHTS MENSCHLICHERES
ALS DEN WILLEN ZU ÜBERLEBEN
EX_MACHINA
DEMNÄCHST IM KINO

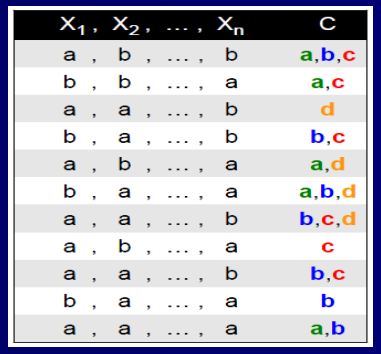

TOM CRUISE
MINORITY REPORT

MACHINE LEARNING

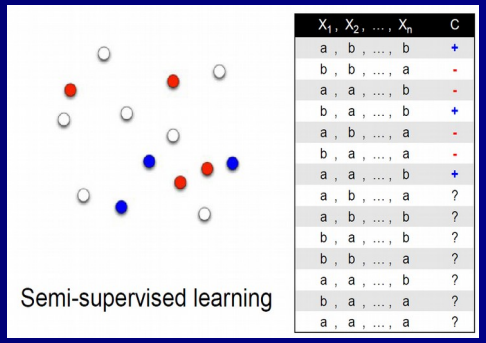
PRINCIPAL LEARNING SCENARIOS

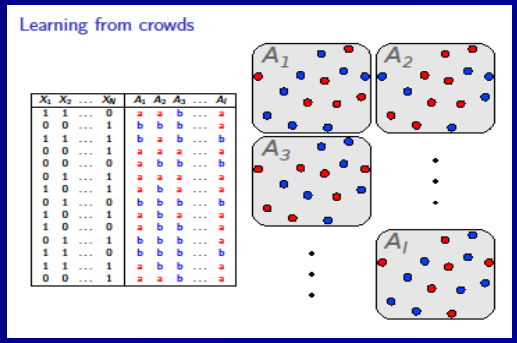
– NLP ORIENTED –

Iñaki Inza
Intelligent Systems Group, www.sc.ehu.es/isg
Computer Science Faculty
University of the Basque Country, Donostia - San Sebastian

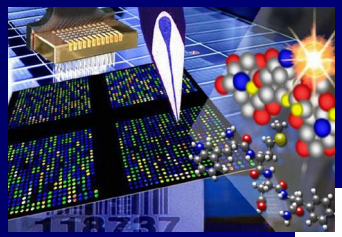


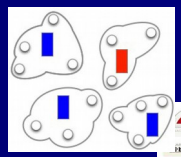
X_1 , X_2 , ... , X_n	C
a , b , ... , b	a,b,c
b , b , ... , a	a,c
a , a , ... , b	d
b , a , ... , b	b,c
a , b , ... , a	a,d
b , a , ... , a	a,b,d
a , a , ... , b	b,c,d
a , b , ... , a	c
a , a , ... , b	b,c
b , a , ... , a	b
a , a , ... , a	a,b

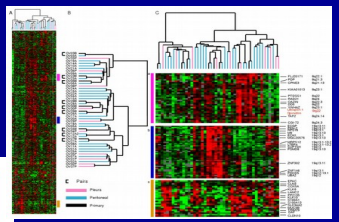


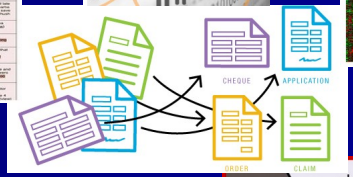


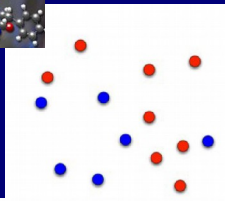
(a subset of) LEARNING SCENARIOS AND APPLICATIONS

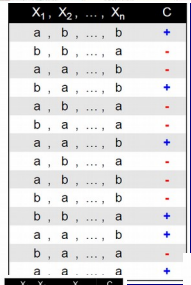


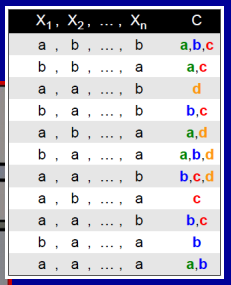


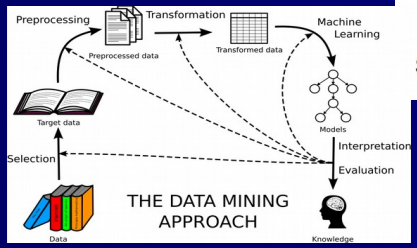


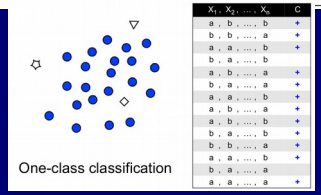


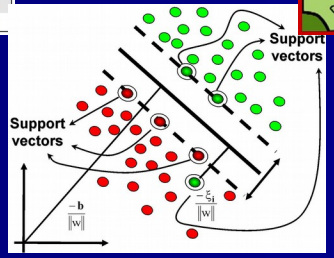








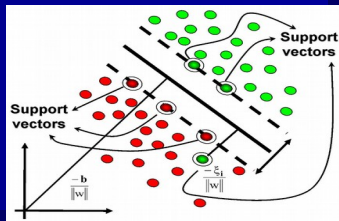




DATA MINING: MAIN TASKS

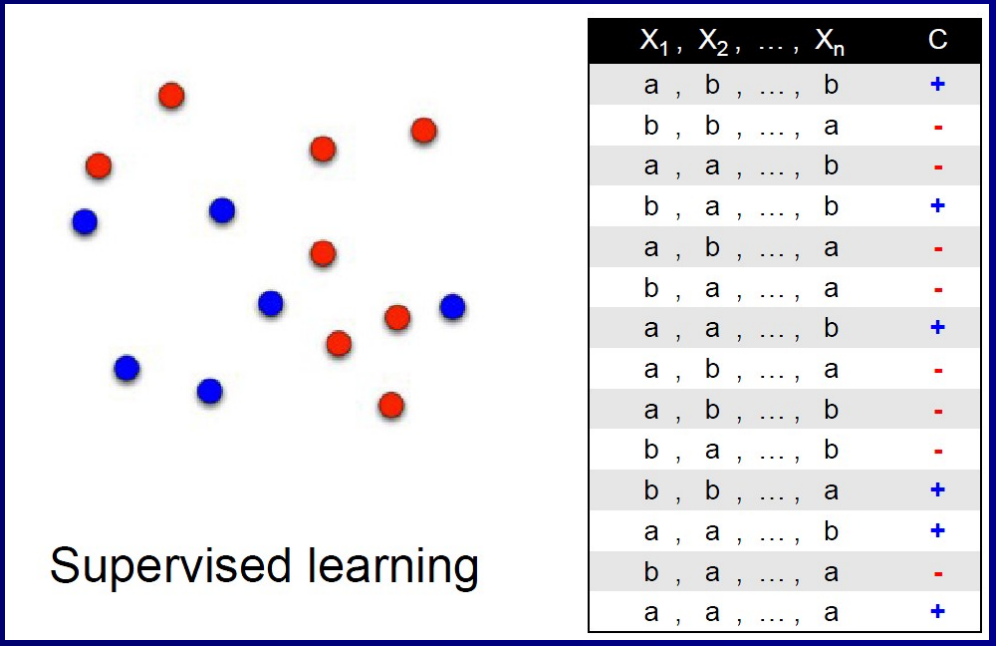
- Prediction Methods
  - Use some variables to predict unknown or future values of other variables
    - *Supervised classification: nominal variable to be predicted*
    - *Regression: quantitative variable to be predicted*

- Description Methods
  - Find human-interpretable patterns that describe the data
    - *Clustering – unsupervised classification*
    - *ANOVA – groups differences by variance analysis*
    - *Association rule discovery*
    - *Feature selection: discover the key predictors*
    - *Outlier detection*



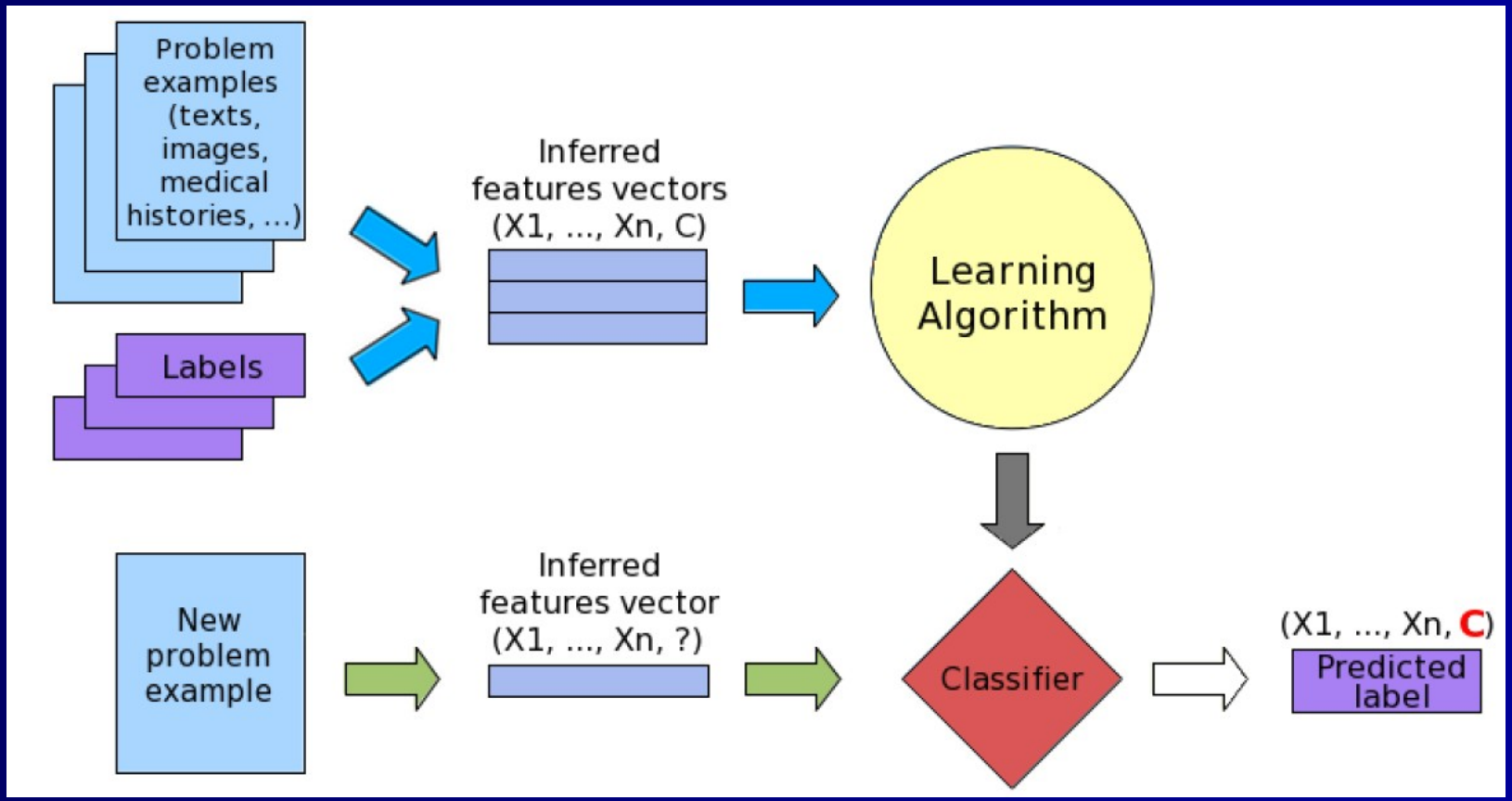
SUPERVISED CLASSIFICATION

- Given a collection of records-samples (*training set*)
  - Each record contains a set of *attributes-features-predictors*
  - Each record belongs to a *class, our variable of interest (variable to be predicted)*

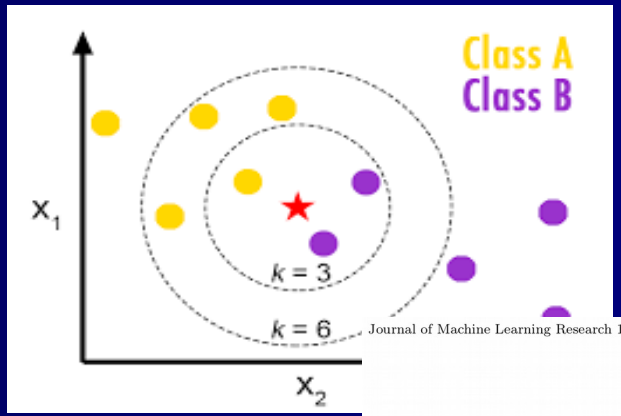


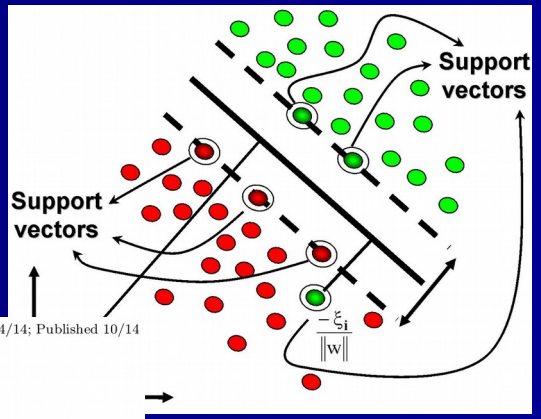
X_1 , X_2 , ... , X_n	C
a , b , ... , b	+
b , b , ... , a	-
a , a , ... , b	-
b , a , ... , b	+
a , b , ... , a	-
b , a , ... , a	-
a , a , ... , b	+
a , b , ... , a	-
a , b , ... , b	-
b , a , ... , b	-
b , b , ... , a	+
a , a , ... , b	+
b , a , ... , a	-
a , a , ... , a	+

SUPERVISED CLASSIFICATION: the standard scenario



SUPERVISED CLASSIFICATION: models



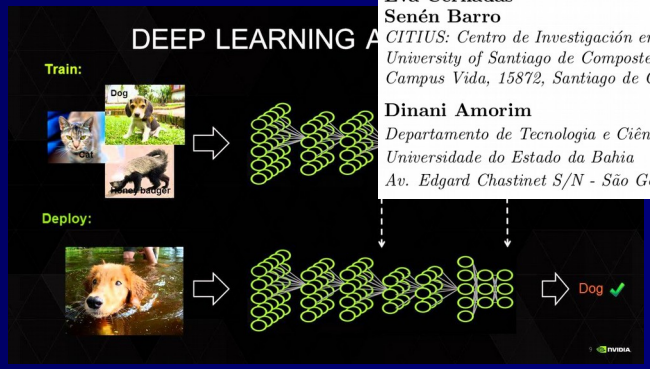


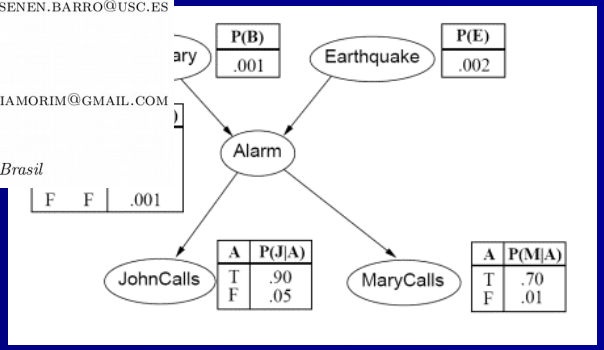
Journal of Machine Learning Research 15 (2014) 3133-3181 Submitted 11/13; Revised 4/14; Published 10/14

Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

Manuel Fernández-Delgado MANUEL.FERNANDEZ.DELGADO@USC.ES
Eva Cernadas EVA.CERNADAS@USC.ES
Senén Barro SENEN.BARRO@USC.ES
CITIUS: Centro de Investigación en Tecnoloxías da Información da USC
University of Santiago de Compostela
Campus Vida, 15872, Santiago de Compostela, Spain

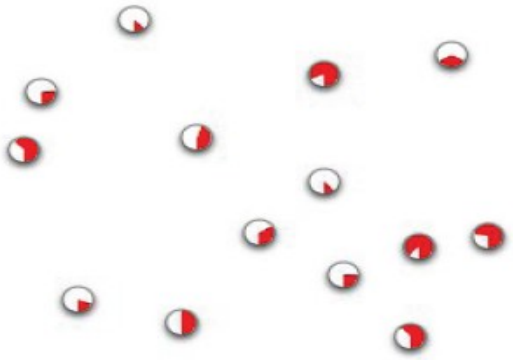
Dinani Amorim DINANIAMORIM@GMAIL.COM
Departamento de Tecnologia e Ciências Sociais- DTCS
Universidade do Estado da Bahia
Av. Edgard Chastinet S/N - São Geraldo - Juazeiro-BA, CEP: 48.305-680, Brasil





REGRESSION

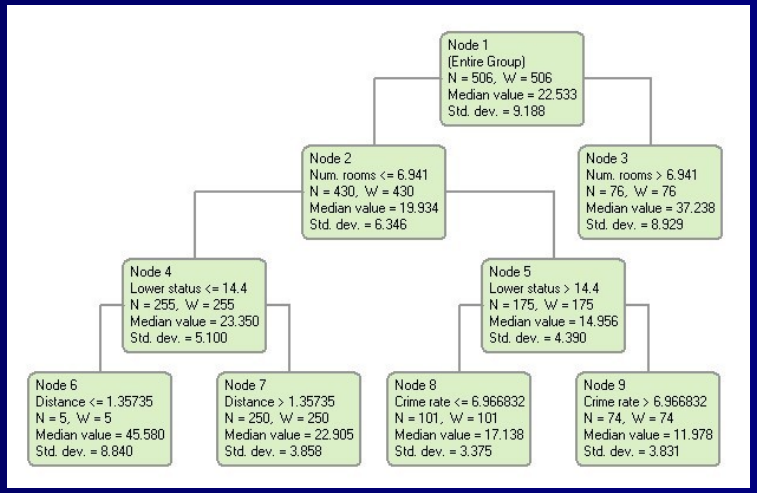
- The *variable of interest* to be predicted is *quantitative*



Regression

X_1 , X_2 , … , X_n	C
a , b , … , b	1.1
b , b , … , a	0.1
a , a , … , b	6.3
b , a , … , b	2.0
a , b , … , a	3.3
b , a , … , a	2.1
a , a , … , b	5.5
a , b , … , a	4.0
a , b , … , b	0.5
b , a , … , b	5.6
b , b , … , a	4.3
a , a , … , b	2.2
b , a , … , a	1.2
a , a , … , a	3.1

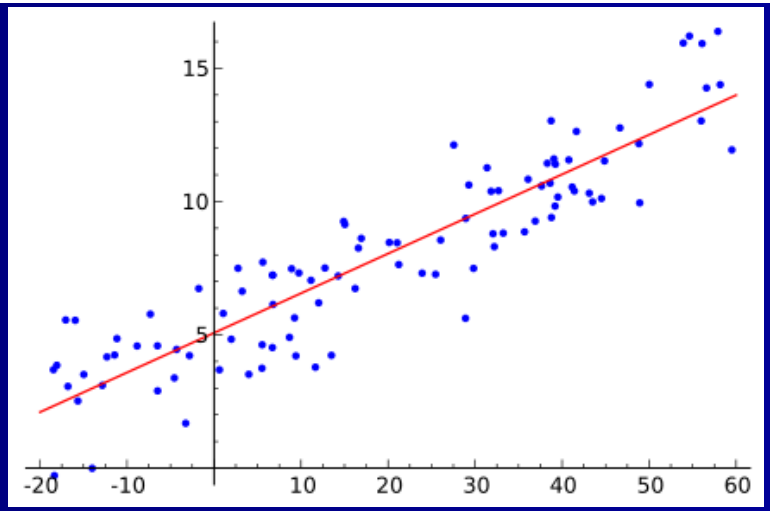
REGRESSION: models



Regression trees

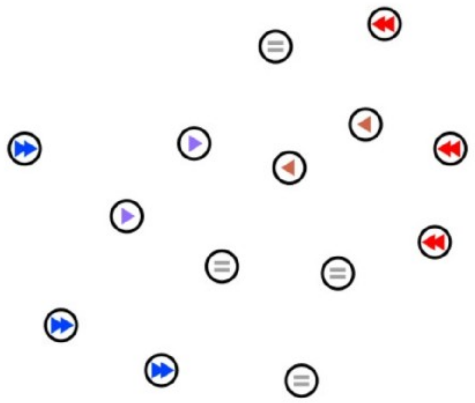
$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

Linear regression



ORDINAL CLASSIFICATION

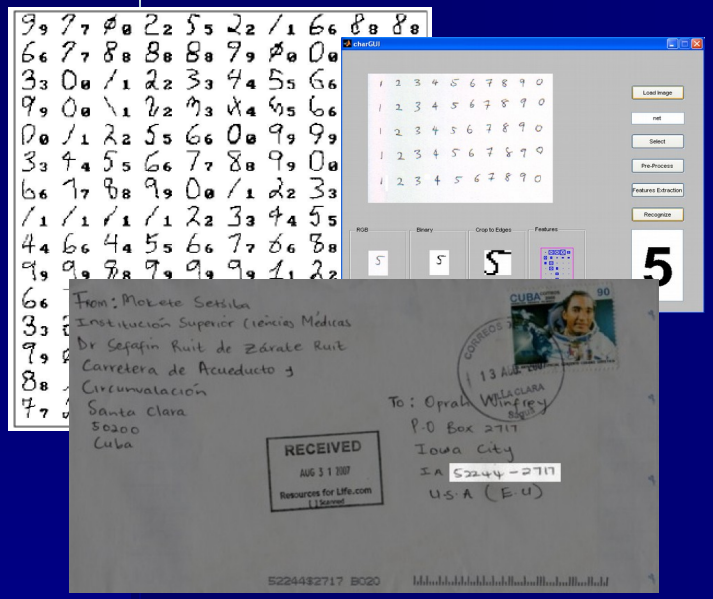
- The *variable of interest* to be predicted is *discrete, but ordered*

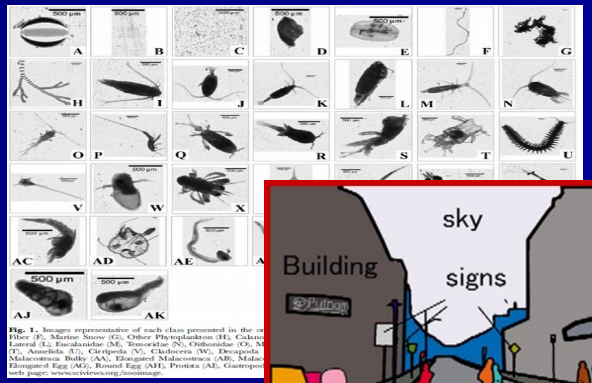


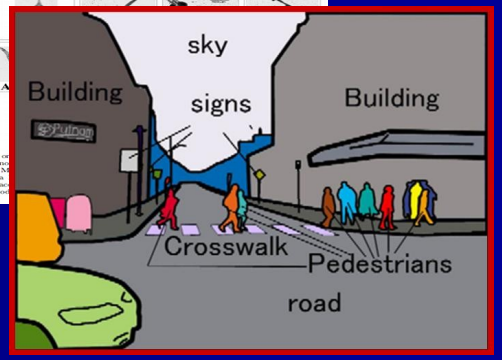
Ordinal learning

X_1 , X_2 , … , X_n	C
a , b , … , b	<<
b , b , … , a	>
a , a , … , b	=
b , a , … , b	<
a , b , … , a	>>
b , a , … , a	>
a , a , … , b	<
a , b , … , a	=
a , b , … , b	=
b , a , … , b	<
b , b , … , a	>>
a , a , … , b	>
b , a , … , a	<<
a , a , … , a	=

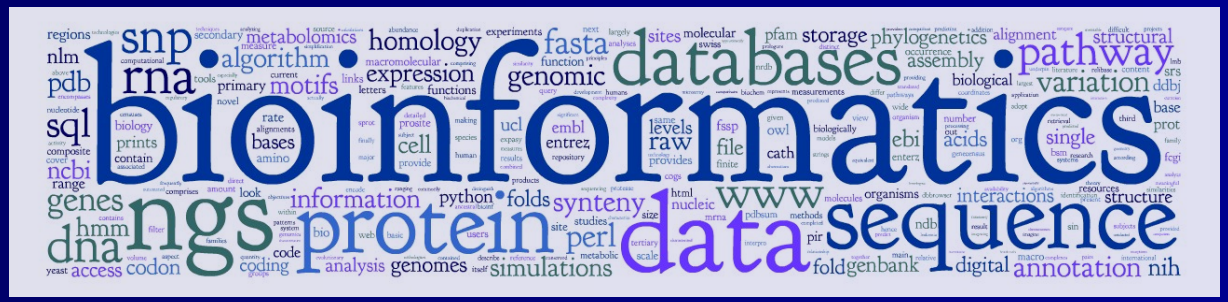
SUPERVISED CLASSIFICATION and REGRESSION:
APPLICATIONS
PATTERN RECOGNITION

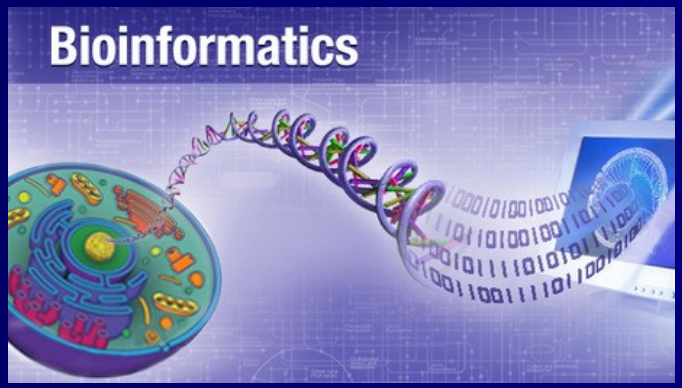


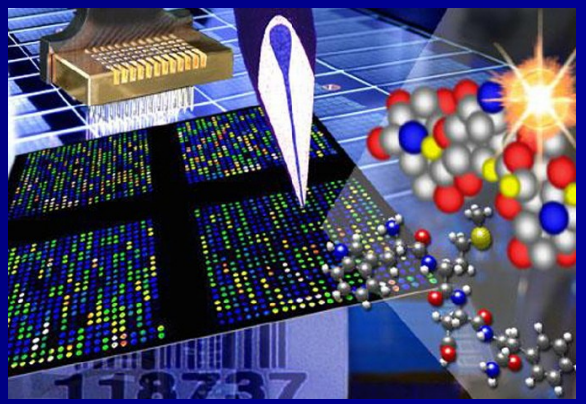




BIOINFORMATICS
DIAGNOSIS AND PROGNOSIS OF DISEASES
BIOMARKER DISCOVERY

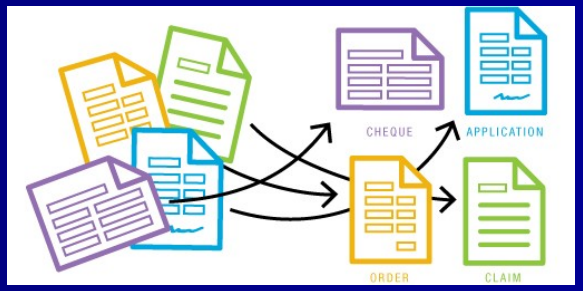






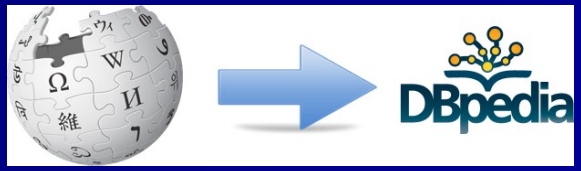
DOCUMENT CLASSIFICATION

- "Natural Language Processing" (NLP)



- Topic - category
- Level of difficulty
- Author's genre

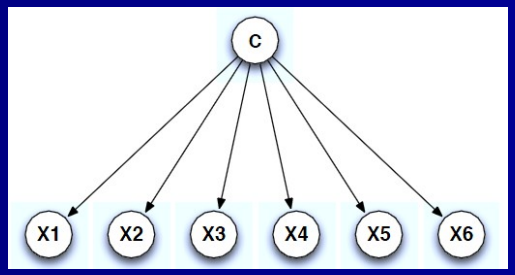
- .....

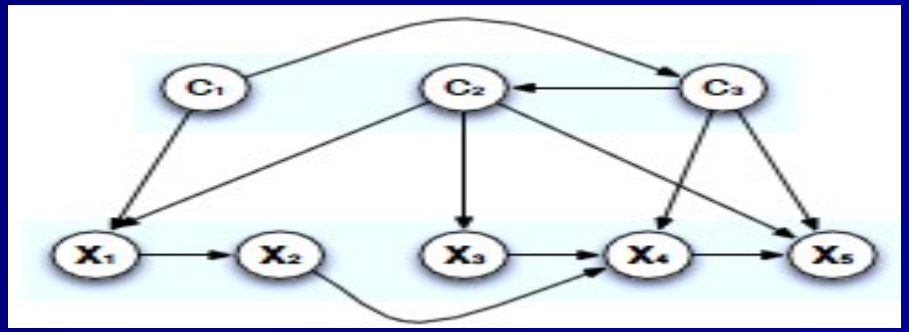


BEYOND SINGLE CLASS VARIABLE...
MULTIDIMENSIONAL CLASSIFICATION

- Several class variables to be jointly predicted
- Learn relationships between class variables
- New term: Joint accuracy

X_1	X_2	...	X_n	C_1	C_2	...	C_m
$x_1^{(1)}$	$x_2^{(1)}$	...	$x_n^{(1)}$	$c_1^{(1)}$	$c_2^{(1)}$	...	$c_m^{(1)}$
$x_1^{(2)}$	$x_2^{(2)}$	...	$x_n^{(2)}$	$c_1^{(2)}$	$c_2^{(2)}$	...	$c_m^{(2)}$
...	...	...	...	...	...	...	...
$x_1^{(N)}$	$x_2^{(N)}$	...	$x_n^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	...	$c_m^{(N)}$





MULTIDIMENSIONAL CLASSIFICATION APPLICATIONS

Neurocomputing 92 (2012) 98–115

Contents lists available at SciVerse ScienceDirect

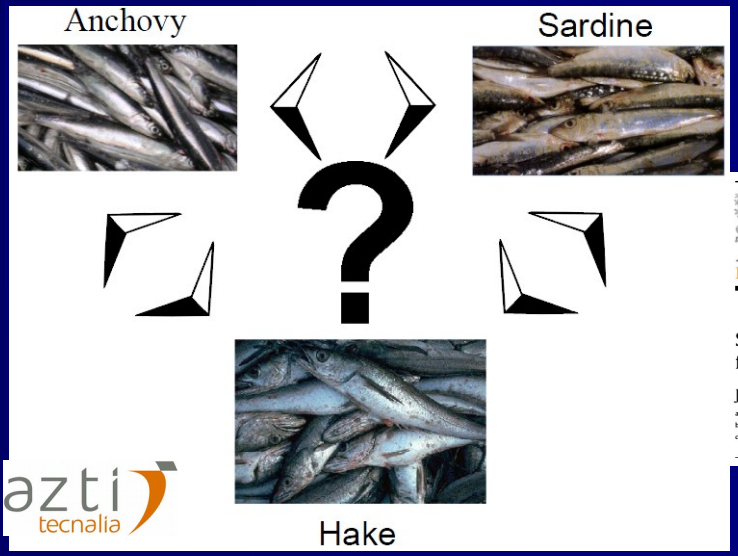
Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers

Jonathan Ortigosa-Hernández [a,*], Juan Diego Rodríguez [a], Leandro Alzate [b], Manuel Lucania [b], Iñaki Inza [a], Jose A. Lozano [a]

[a] Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, Computer Science Faculty, The University of the Basque Country UPV/EHU, San Sebastián, Spain
[b] Socialware®, Bilbao, Spain



Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

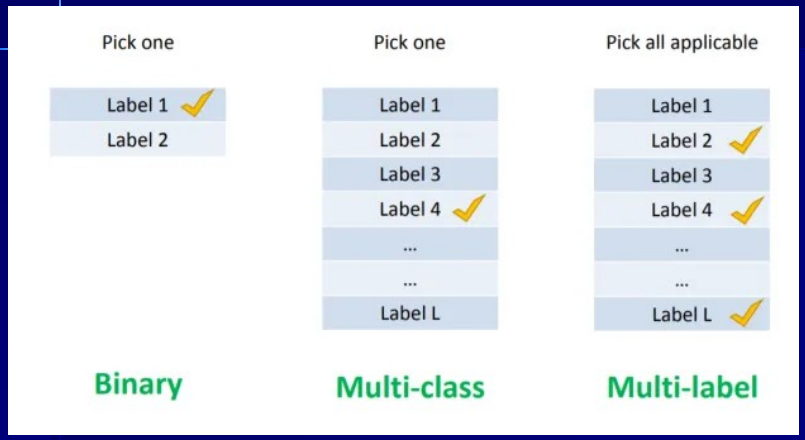
journal homepage: www.elsevier.com/locate/envsoft

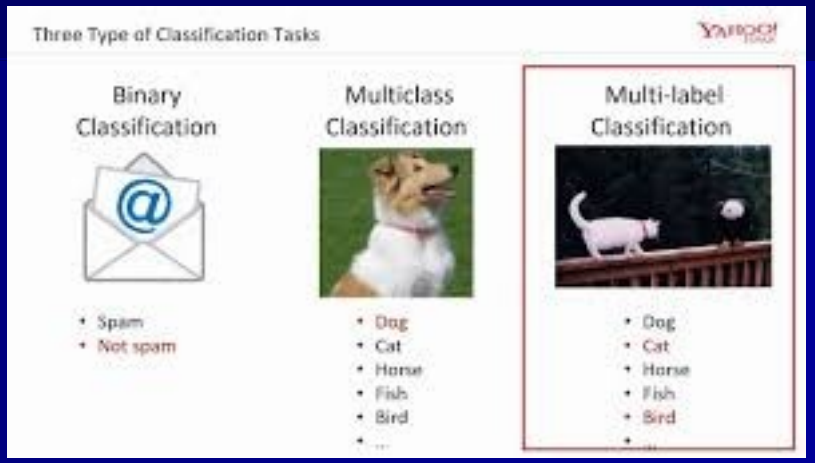
Supervised pre-processing approaches in multiple class variables classification for fish recruitment forecasting

Jose A. Fernandes [a,b,*], Jose A. Lozano [b], Iñaki Inza [b], Xabier Irigoien [a,c], Aritz Pérez [b], Juan D. Rodríguez [b]

[a] AZTI-Tecnalia, Marine Research Division, Herrera Kaia z/g, E-20110 Pasaia (Gipuzkoa), Spain
[b] University of the Basque Country, Department of Computer Science and AI, Intelligent Systems Group (ISG), Paseo Manuel de Lardizabal, 1, E-20018 Donostia – San Sebastián, Spain
[c] King Abdullah University of Science and Technology (KAUST), Chemical and Life Sciences and Engineering, Red Sea Research Center, Thuwal 23955-6900, Saudi Arabia

MULTILABEL CLASSIFICATION





Multilabel Text Classification for Automated Tag Suggestion

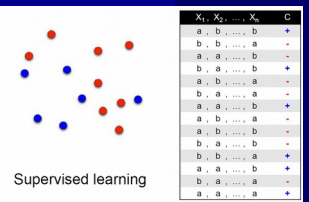
Ioannis Katakis, Grigorios Tsoumakas, and Ioannis Vlahavas

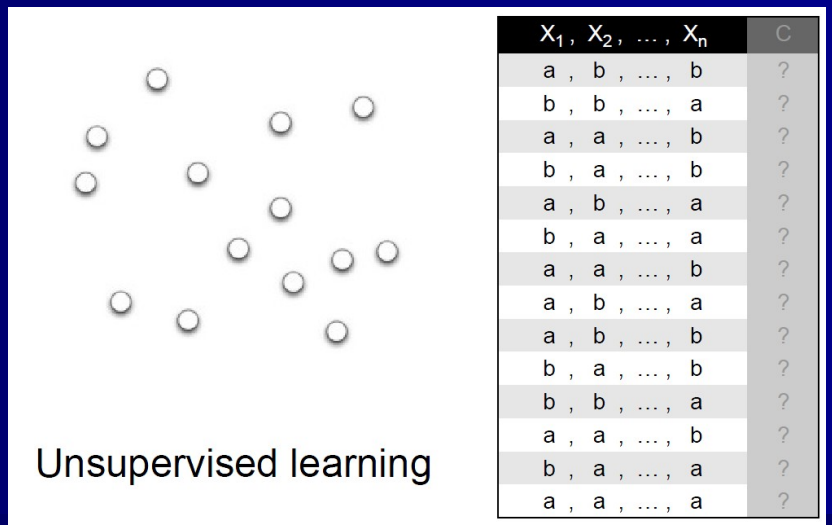
Department of Informatics,
Aristotle University of Thessaloniki,
54124 Thessaloniki, Greece
{katak,greg,vlahavas}@csd.auth.gr

UNSUPERVISED CLASSIFICATION
CLUSTERING

- Given a collection of records-samples (*training set*)
  - Each record contains a set of *attributes-features-~~predictors~~*
  - *No "target feature" (class)* which supervises the learning process

- Groups of cases:
  - Large intra-group homogeneity
  - Large inter-groups heterogeneity

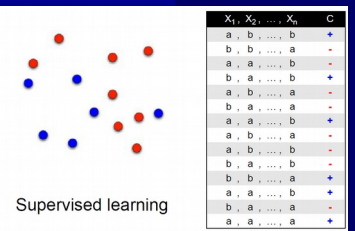




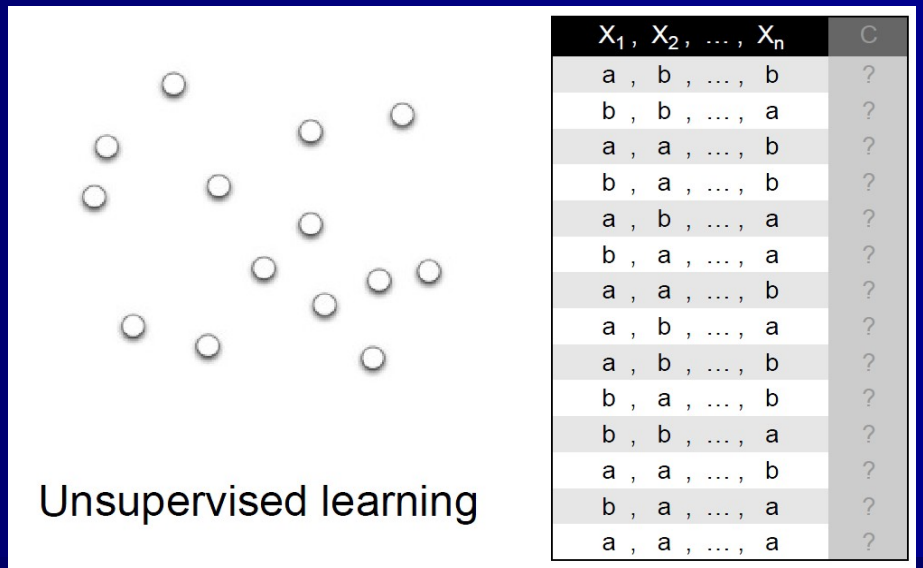
X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?

UNSUPERVISED CLASSIFICATION CLUSTERING

- Difficult evaluation → "no accuracy" concept
- Number of groups → difficult decision
- "Distance" function: e.g. Euclidean (ordinal), overlap (nominal: a=a~dist=0, a≠b~dist=1, a≠c~dist=1, b≠c~dist=1)

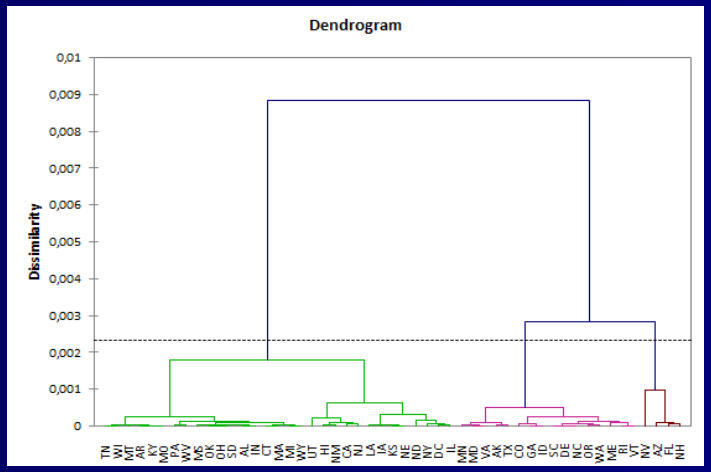


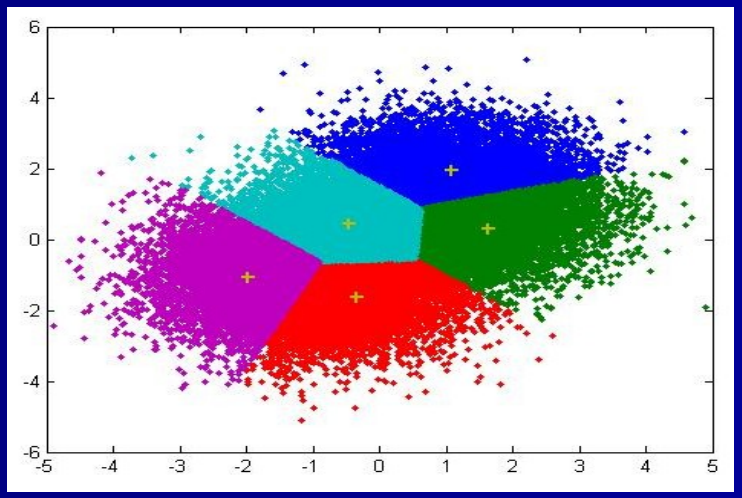
X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-
a , a , … , b	+
a , b , … , b	-
b , a , … , b	-
b , b , … , a	+
a , a , … , b	+
b , a , … , a	-
a , a , … , a	+

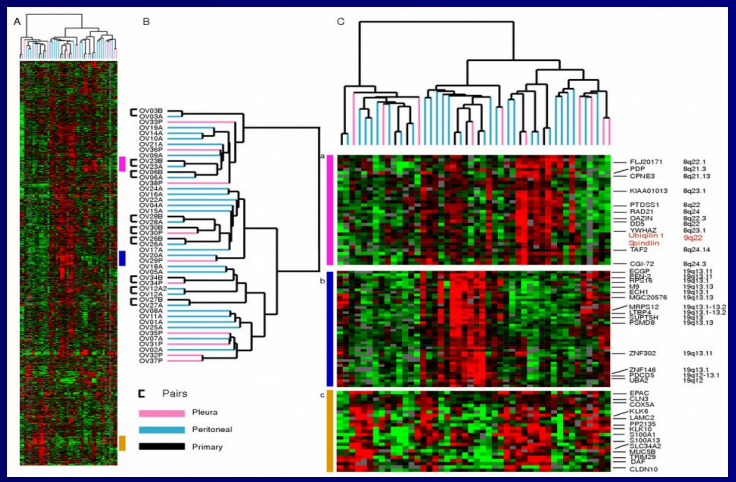


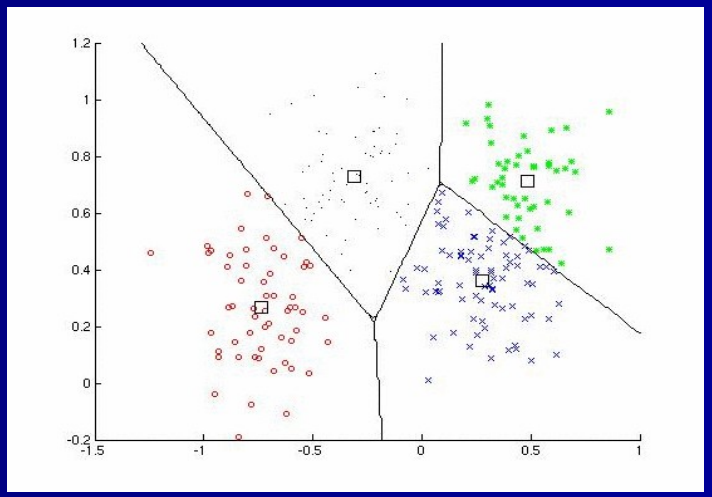
X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?

CLUSTERING: MODELS









CLUSTERING: APPLICATIONS
CUSTOMER SEGMENTATION

- Identify micro-markets and develop policies for each
- Targeted marketing
- Similar customers are grouped in the same cluster

CLUSTERING: APPLICATIONS
CLUSTERING DOCUMENTS

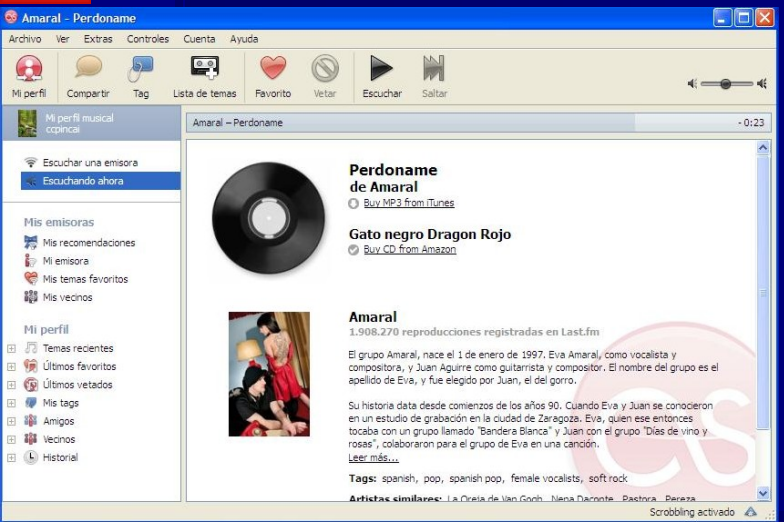
Co-clustering documents and words using Bipartite Isoperimetric Graph Partitioning

Manjeet Rege Ming Dong Farshad Fotouhi
Machine Vision and Pattern Recognition Lab Database and Multimedia Systems Group
Department of Computer Science, Wayne State University
Detroit, MI 48202, USA

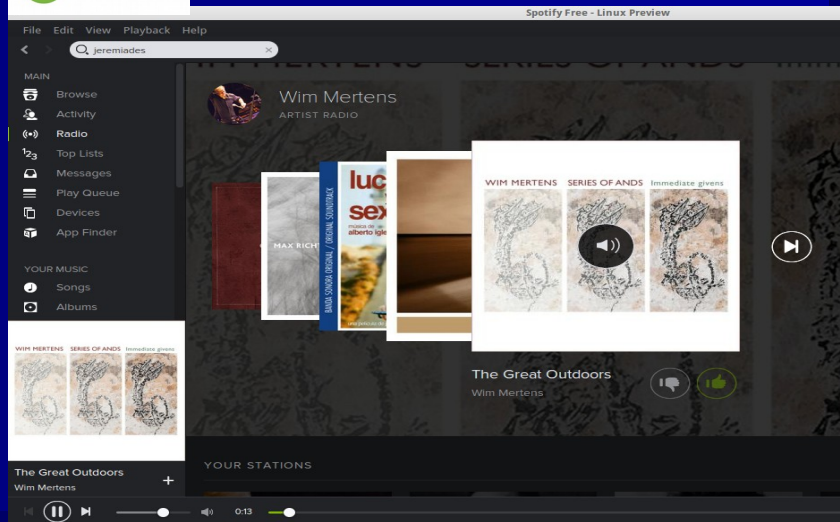
COLLABORATIVE FILTERING RECOMMENDER SYSTEMS

Movie	Alice (1)	Bob (2)	Carol (3)	Dave (4)
Love at last	5	5	0	6
Romance forever	5	?	?	0
Cute puppies of love	?	4	0	?
Nonstop car chases	0	0	5	4
Swords vs. karate	0	0	5	?

last.fm



Spotify

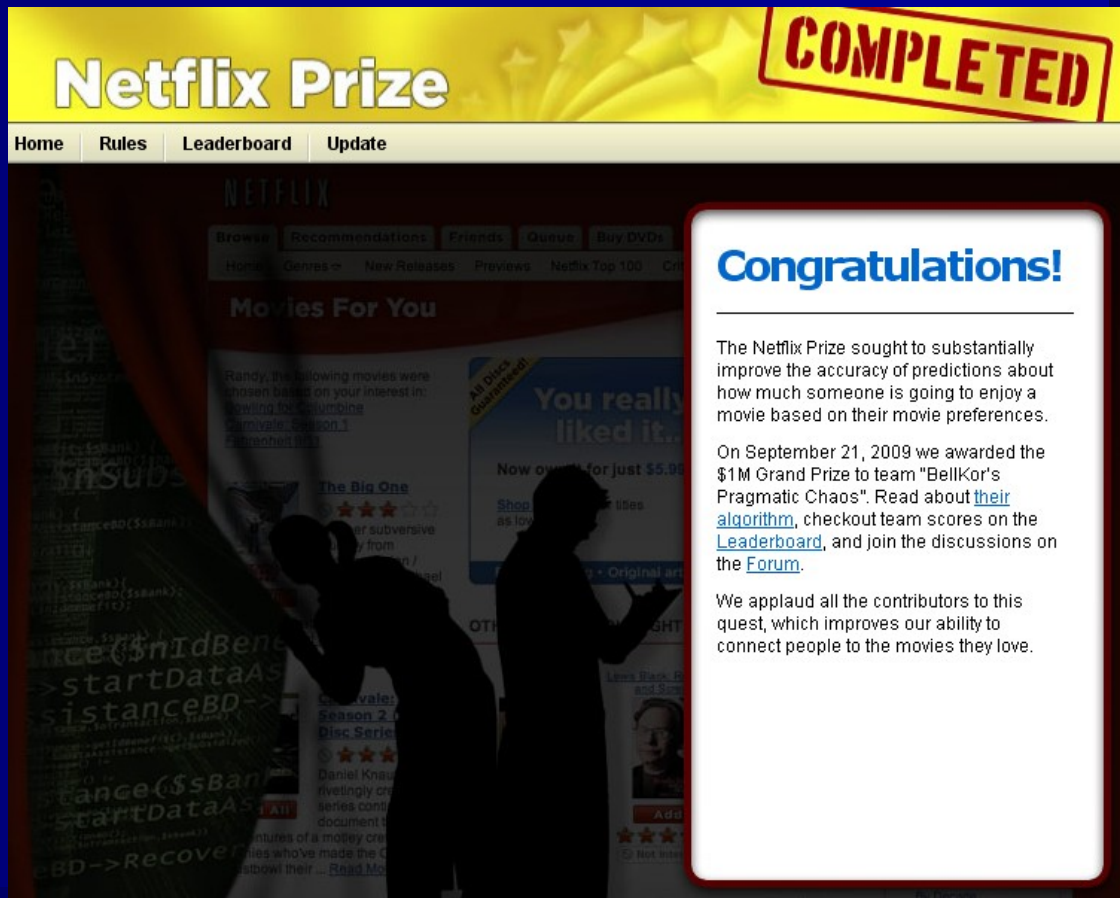


FILM RECOMMENDATION SYSTEMS

NETFLIX PRIZE

NETFLIX

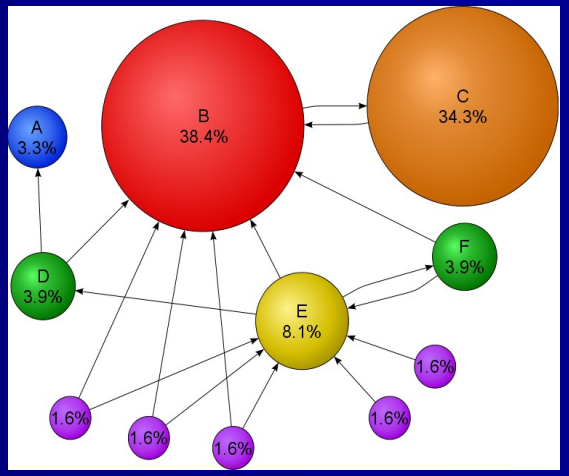
Watch TV shows & movies anytime, anywhere. For one low monthly price.



TEXT CLUSTERING

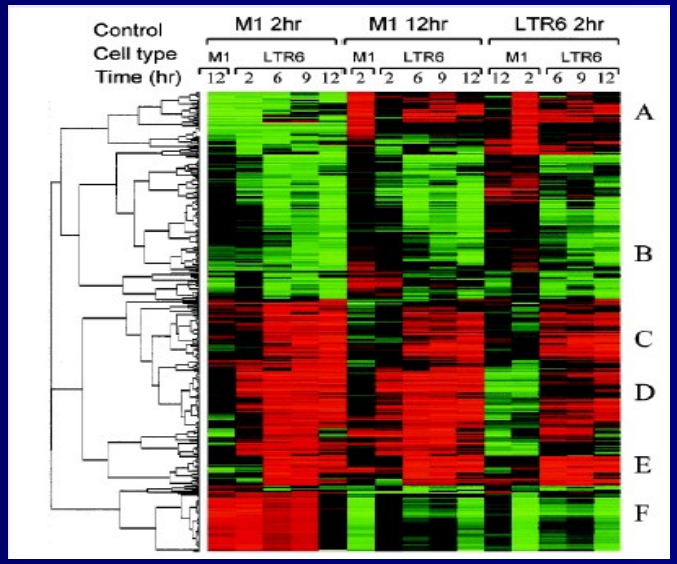
- Text clustering: documents that are similar → Google
  - PageRank
  - Diversity in search results

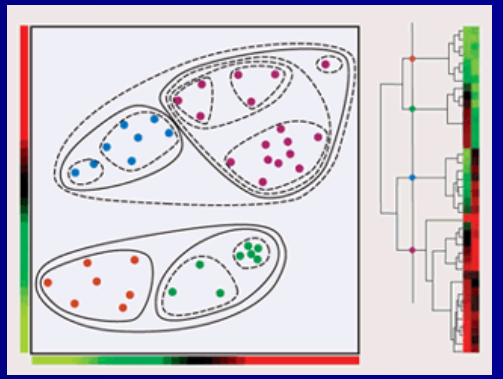
- Google, a key example of efficient data organization-"clustering"



BICLUSTERING GENE EXPRESSION

- Find genes with similar expression profiles ~ a way to infer the function of genes whose function is unknown
- Biclustering... a classic concept in fashion again:
  - Finding a subgroup of samples with a similar pattern in a subgroup of variables (not in all the variables)





BICLUSTERING WORDS IN DOCUMENTS

International Conference on Artificial Intelligence and Soft Computing
ICAISC 2016: Artificial Intelligence and Soft Computing pp 102-113 | Cite as

Text Mining with Hybrid Biclustering Algorithms

Authors | Authors and affiliations

Patryk Orzechowski ✉, Krzysztof Boryczko

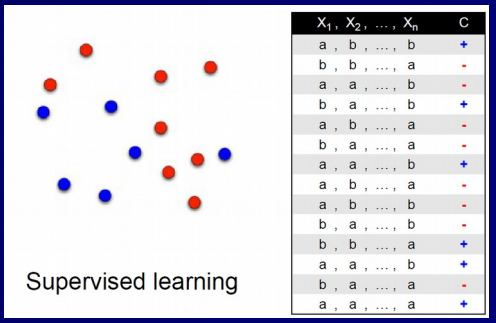
Conference paper
First Online: 29 May 2016

Part of the Lecture Notes in Computer Science book series (LNCS, volume 9693)

Abstract

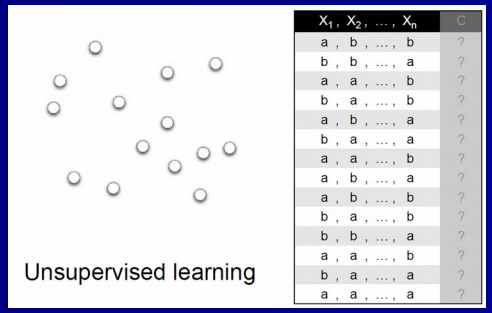
Text data mining is the process of extracting valuable information from a dataset consisting of text documents. Popular clustering algorithms do not allow detection of the same words appearing in multiple documents. Instead, they discover general similarity of such documents. This article presents the application of a hybrid biclustering algorithm for text mining documents collected from Twitter and symbolic analysis of knowledge spreadsheets. The proposed method automatically reveals words appearing together in multiple texts. The proposed approach is compared to some of the most recognized clustering algorithms and shows the advantage of biclustering over clustering in text mining. Finally, the method is confronted with other biclustering methods in the task of classification.

IS THERE SOMEONE IN THE MIDDLE?



Supervised learning

$X_1, X_2, \ldots, X_n$	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-
a , a , … , b	+
a , b , … , a	-
a , b , … , b	-
b , a , … , b	-
b , b , … , a	+
a , a , … , b	+
b , a , … , a	-
a , a , … , a	+



Unsupervised learning

$X_1, X_2, \ldots, X_n$	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?

Pattern Recognition Letters 69 (2016) 49–55

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Weak supervision and other non-standard classification problems: A taxonomy☆

Jerónimo Hernández-González*, Iñaki Inza, Jose A. Lozano

Intelligent Systems Group, University of the Basque Country UPV/EHU, P. Manuel Lardizabal 1, 20018 Donostia, Spain

ARTICLE INFO

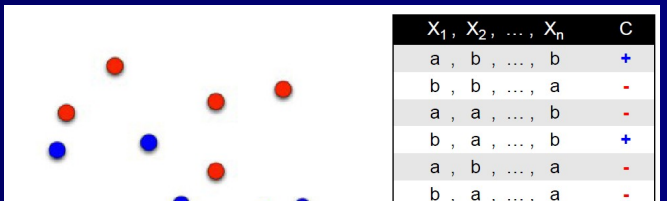
Article history:
Received 10 May 2015
Available online 24 October 2015

Keywords:
Weakly supervised classification
Partially supervised classification
Degrees of supervision

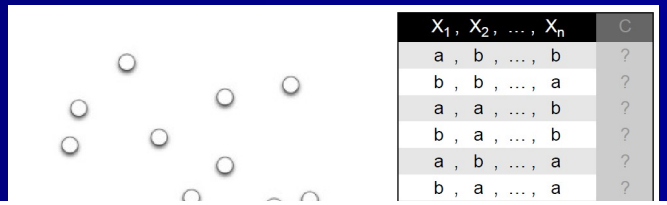
ABSTRACT

In recent years, different researchers in the machine learning community have presented new classification frameworks which go beyond the standard supervised classification in different aspects. Specifically, a wide spectrum of novel frameworks that use partially labeled data in the construction of classifiers has been studied. With the objective of drawing up a description of the state-of-the-art, three identifying characteristics of these novel frameworks have been considered: (1) the relationship between instances and labels of a problem, which may be beyond the *one-instance one-label* standard, (2) the possible provision of partial class information for the training examples, and (3) the possible provision of partial class information also for the examples in the prediction stage. These three ideas have been formulated as axes of a comprehensive taxonomy that organizes the state-of-the-art. The proposed organization allows us both to understand similarities/differences among the different classification problems already presented in the literature as well as to discover unexplored frameworks that might be seen as further challenges and research opportunities. A representative set of state-of-the-art problems has been used to illustrate the novel taxonomy and support the discussion.

IS THERE SOMEONE IN THE MIDDLE?

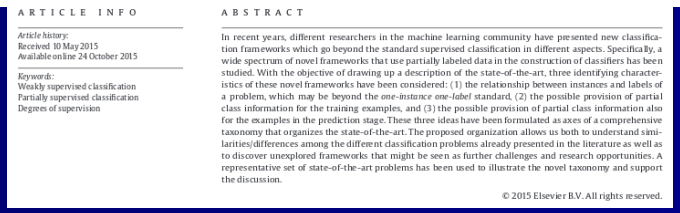


X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-



X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?

- Hidden big data. Large quantities of useful data are in fact useless because they are untagged, file-based, and unstructured. The 2012 IDC study on big data [117] explained that, in 2012, 23% (643 exabytes) of the digital universe would be useful for big data if tagged and analyzed. However, at that time only 3% of the potentially useful data was tagged, and even less was analyzed. The figures have probably gotten worse in recent years. The Open Data and Semantic Web movements have emerged, in part, to make us aware and improve on this situation. No comments



ARTICLE INFO

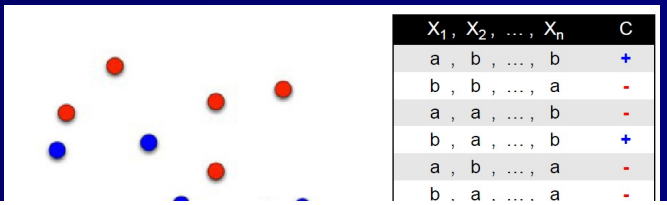
Article history:
Received 10 May 2015
Available online 24 October 2015

Keywords:
Weakly supervised classification
Partially supervised classification
Degrees of supervision

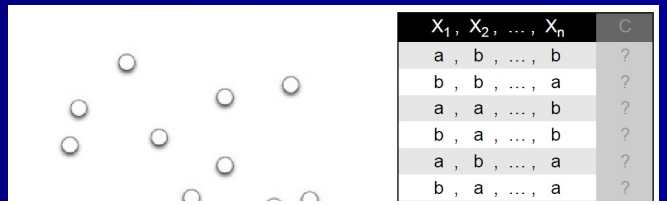
ABSTRACT

In recent years, different researchers in the machine learning community have presented new classification frameworks which go beyond the standard supervised classification in different aspects. Specifically, a wide spectrum of novel frameworks that use partially labeled data in the construction of classifiers has been studied. With the objective of drawing up a description of the state-of-the-art, three identifying characteristics of these novel frameworks have been considered: (1) the relationship between instances and labels of a problem, which may be beyond the *one-instance one-label* standard, (2) the possible provision of partial class information for the training examples, and (3) the possible provision of partial class information also for the examples in the prediction stage. These three ideas have been formulated as axes of a comprehensive taxonomy that organizes the state-of-the-art. The proposed organization allows us both to understand similarities/differences among the different classification problems already presented in the literature as well as to discover unexplored frameworks that might be seen as further challenges and research opportunities. A representative set of state-of-the-art problems has been used to illustrate the novel taxonomy and support the discussion.

IS THERE SOMEONE IN THE MIDDLE?

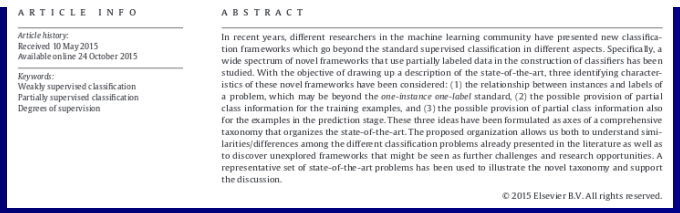


X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-



X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?

- Hidden big data. Large quantities of useful data are in fact useless because they are untagged, file-based, and unstructured. The 2012 IDC study on big data [117] explained that, in 2012, 23% (643 exabytes) of the digital universe would be useful for big data if tagged and analyzed. However, at that time only 3% of the potentially useful data was tagged, and even less was analyzed. The figures have probably gotten worse in recent years. The Open Data and Semantic Web movements have emerged, in part, to make us aware and improve on this situation. No comments



ARTICLE INFO

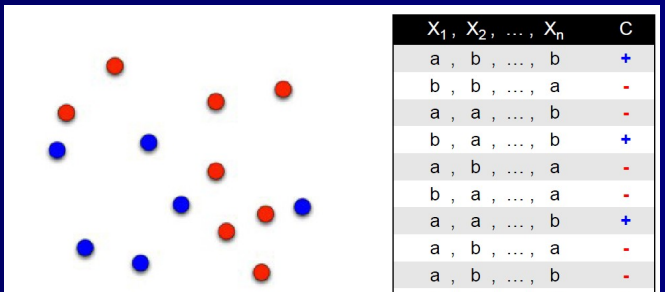
Article history:
Received 10 May 2015
Available online 24 October 2015

Keywords:
Weakly supervised classification
Partially supervised classification
Degrees of supervision

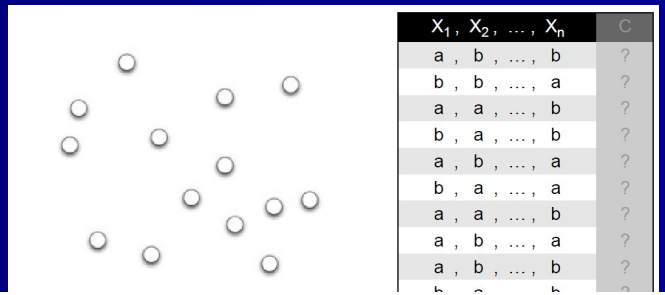
ABSTRACT

In recent years, different researchers in the machine learning community have presented new classification frameworks which go beyond the standard supervised classification in different aspects. Specifically, a wide spectrum of novel frameworks that use partially labeled data in the construction of classifiers has been studied. With the objective of drawing up a description of the state-of-the-art, three identifying characteristics of these novel frameworks have been considered: (1) the relationship between instances and labels of a problem, which may be beyond the *one-instance one-label* standard, (2) the possible provision of partial class information for the training examples, and (3) the possible provision of partial class information also for the examples in the prediction stage. These three ideas have been formulated as axes of a comprehensive taxonomy that organizes the state-of-the-art. The proposed organization allows us both to understand similarities/differences among the different classification problems already presented in the literature as well as to discover unexplored frameworks that might be seen as further challenges and research opportunities. A representative set of state-of-the-art problems has been used to illustrate the novel taxonomy and support the discussion.

IS THERE SOMEONE IN THE MIDDLE?



X_1 , X_2 , … , X_n	C
a , b , … , b	+
b , b , … , a	-
a , a , … , b	-
b , a , … , b	+
a , b , … , a	-
b , a , … , a	-
a , a , … , b	+
a , b , … , a	-
a , b , … , b	-



X_1 , X_2 , … , X_n	C
a , b , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?

Labelomania

We are witnessing a data labeling market explosion: labeling platforms have hit prime time. S&P Global released an October 11 report entitled *Avoiding Garbage in Machine Learning* in which it termed unlabeled data "garbage data" to highlight the importance of labeling in AI. The Economist

Weak supervision and other non-standard classification problems: A taxonomy☆

Jerónimo Hernández-González*, Iñaki Inza, Jose A. Lozano

Intelligent Systems Group, University of the Basque Country UPV/EHU, P. Manuel Lardizabal 1, 20018 Donostia, Spain

ARTICLE INFO

Article history:
Received 10 May 2015
Available online 24 October 2015

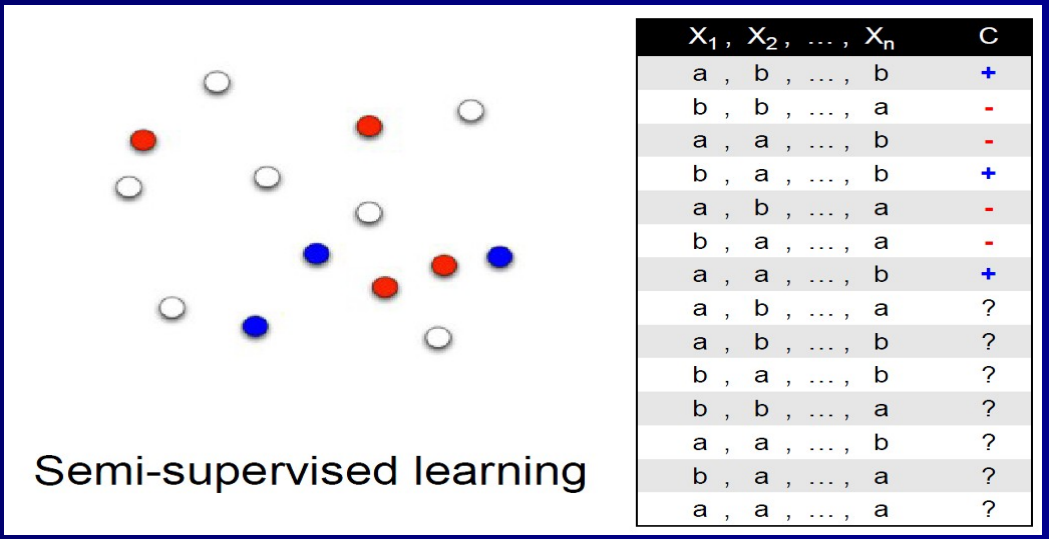
Keywords:
Weakly supervised classification
Partially supervised classification
Degrees of supervision

ABSTRACT

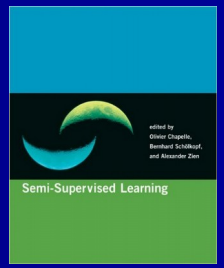
In recent years, different researchers in the machine learning community have presented new classification frameworks which go beyond the standard supervised classification in different aspects. Specifically, a wide spectrum of novel frameworks that use partially labeled data in the construction of classifiers has been studied. With the objective of drawing up a description of the state-of-the-art, three identifying characteristics of these novel frameworks have been considered: (1) the relationship between instances and labels of a problem, which may be beyond the *one-instance one-label* standard, (2) the possible provision of partial class information for the training examples, and (3) the possible provision of partial class information also for the examples in the prediction stage. These three ideas have been formulated as axes of a comprehensive taxonomy that organizes the state-of-the-art. The proposed organization allows us both to understand similarities/differences among the different classification problems already presented in the literature as well as to discover unexplored frameworks that might be seen as further challenges and research opportunities. A representative set of state-of-the-art problems has been used to illustrate the novel taxonomy and support the discussion.

SEMI SUPERVISED CLASSIFICATION

- Most of the samples do not show a class value. Why?
  - Categorization: human-time consuming task
  - No knowledge to categorize the samples
- Objective: learn a supervised model
- Can a learning process which takes advantage of unlabeled samples, construct a better supervised classification model?

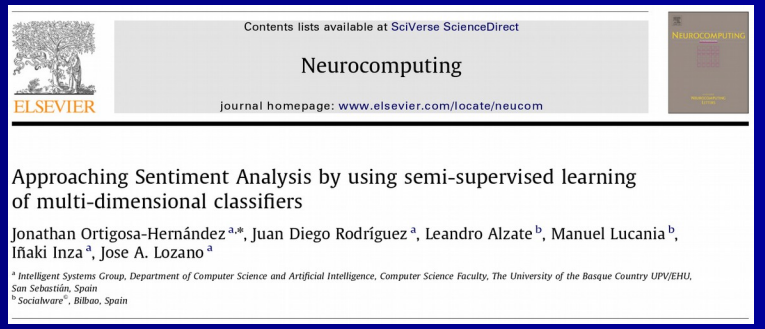


X_1 , X_2 , ... , X_n	C
a , b , ... , b	+
b , b , ... , a	-
a , a , ... , b	-
b , a , ... , b	+
a , b , ... , a	-
b , a , ... , a	-
a , a , ... , b	+
a , b , ... , a	?
a , b , ... , b	?
b , a , ... , b	?
b , b , ... , a	?
a , a , ... , b	?
b , a , ... , a	?
a , a , ... , a	?



SENTIMENT ANALYSIS

- Companies: reputation
- Opinions about its products:
  - social networks
  - blogs
  - forums...
- Automatically classify the written opinion: {+, -, neutral}
- NLP: “Natural Language Processing”



Contents lists available at SciVerse ScienceDirect

Neurocomputing

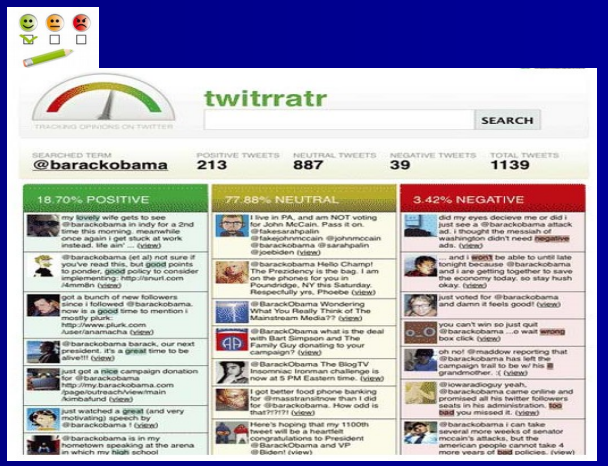
journal homepage: www.elsevier.com/locate/neucom

Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers

Jonathan Ortigosa-Hernández [a,*], Juan Diego Rodríguez [a], Leandro Alzate [b], Manuel Lucania [b], Iñaki Inza [a], Jose A. Lozano [a]

[a] Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, Computer Science Faculty, The University of the Basque Country UPV/EHU, San Sebastián, Spain

[b] Socialware®, Bilbao, Spain



AMBIGUOUS TRAINING DATA

- Training samples annotated by expert
- "Positive" + "Negative"

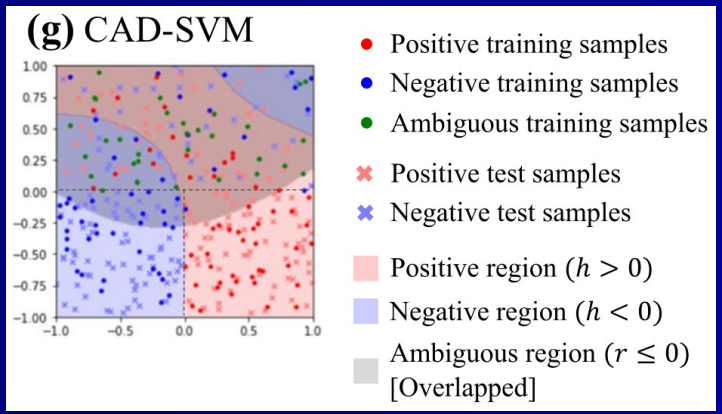
- But → "Ambiguous" samples
- Difficult to label by expert

- "Ambiguous" → "Unlabelled" + Semi-Supervised learning ????
- Not semi-supervised scenario !!
- Unlabeled instances → not uniformly distributed !!

Machine Learning (2020) 109:2369–2388
https://doi.org/10.1007/s10994-020-05915-2

Binary classification with ambiguous training data

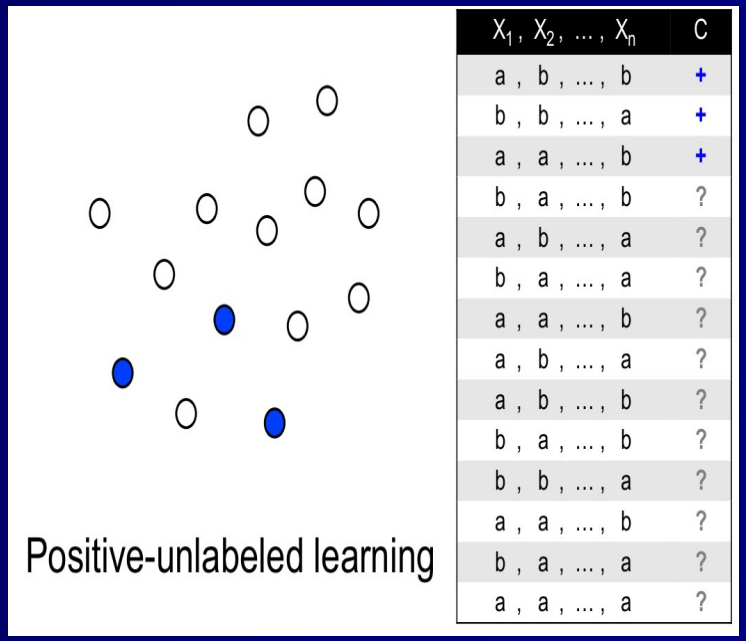
Naoya Otani[1] · Yosuke Otsubo[1] · Tetsuya Koike[1] · Masashi Sugiyama[2,3]



POSITIVE UNLABELED LEARNING

- More difficult than semi-supervised classification: positive-unlabaled
- Prediction: "+" or "-"

- Application: prediction of genes related to cancer



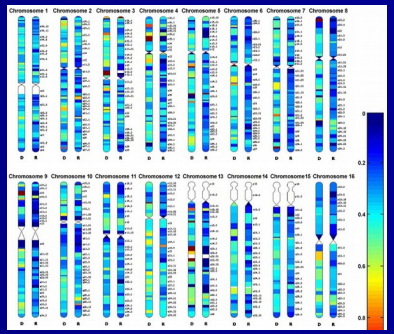
$X_1, X_2, \ldots, X_n$	C
a , b , … , b	+
b , b , … , a	+
a , a , … , b	+
b , a , … , b	?
a , b , … , a	?
b , a , … , a	?
a , a , … , b	?
a , b , … , a	?
a , b , … , b	?
b , a , … , b	?
b , b , … , a	?
a , a , … , b	?
b , a , … , a	?
a , a , … , a	?

Published online 18 August 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 18* **e115**
doi:10.1093/nar/gkn482

Prioritization of candidate cancer genes—an aid to oncogenomic studies

Simon J. Furney[1], Borja Calvo[2], Pedro Larrañaga[3], Jose A. Lozano[2] and Nuria Lopez-Bigas[1,*]

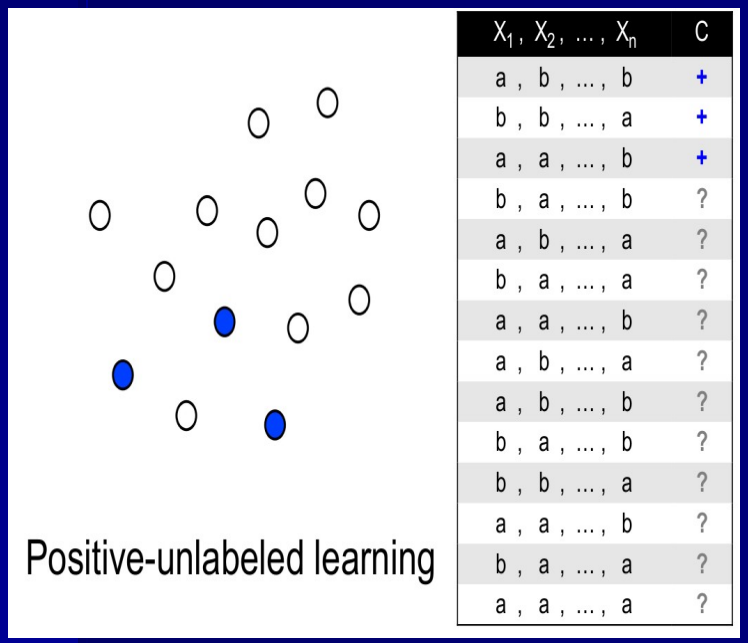
[1]Research Unit on Biomedical Informatics, Experimental and Health Science Department, Universitat Pompeu Fabra, Barcelona 08080, [2]Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country, Donostia-San Sebastián 20018 and [3]Department of Artificial Intelligence, Technical University of Madrid, Boadilla del Monte 28660, Spain



POSITIVE UNLABELED LEARNING

- More difficult than semi-supervised classification: positive-unlabaled
- Prediction: "+" or "-"

- Application: prediction of genes related to cancer



$X_1, X_2, \ldots, X_n$	C
a , b , ... , b	+
b , b , ... , a	+
a , a , ... , b	+
b , a , ... , b	?
a , b , ... , a	?
b , a , ... , a	?
a , a , ... , b	?
a , b , ... , a	?
a , b , ... , b	?
b , a , ... , b	?
b , b , ... , a	?
a , a , ... , b	?
b , a , ... , a	?
a , a , ... , a	?

Text Classification and Co-training from Positive and Unlabeled Examples

François Denis FDENIS@CMI.UNIV-MRS.FR
Anne Laurent ALAURENT@CMI.UNIV-MRS.FR
Équipe BDAA, LIF – UMR 6166, Centre de Mathématiques et d'Informatique (CMI), Université de Provence, 39, rue F. Joliot Curie, 13453 Marseille CEDEX 13, FRANCE

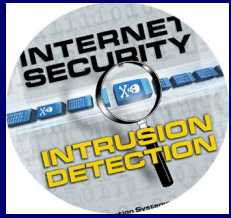
Rémi Gilleron GILLERON@UNIV-LILLE3.FR
Marc Tommasi TOMMASI@UNIV-LILLE3.FR
Équipe Grappa – EA 3588 and projet MOSTRARE – UR INRIA Futurs, Université de Lille 3, domaine universitaire du "Pont de bois", BP 149, 59653 Villeneuve d'Ascq CEDEX, FRANCE

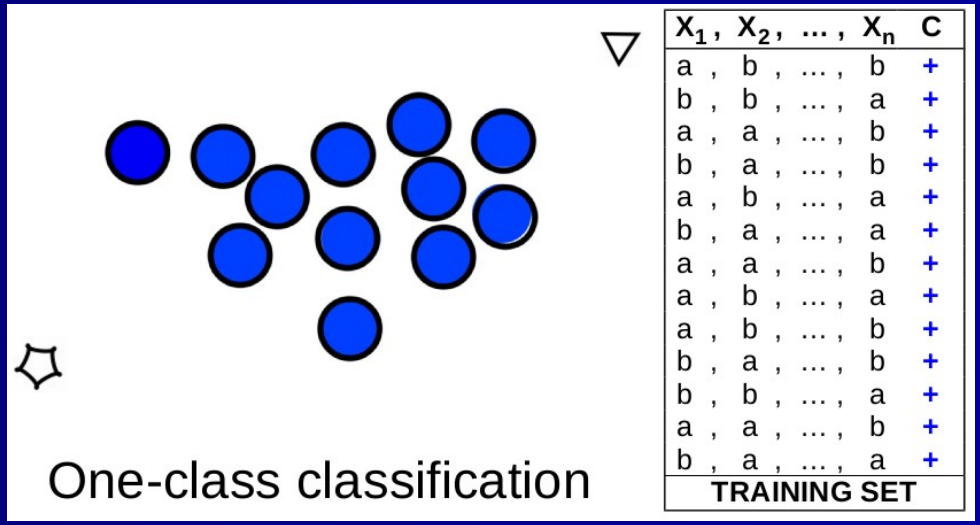
ONE CLASS CLASSIFICATION – OUTLIER DETECTION –

- One category: forms a representative sample
- Only "normal behaviour" samples in training time

- Training phase: model the "normal" behaviour
- Prediction phase: detect "deviations" from the "normal" model

- Model the "dominant" class and "isolate" outliers in "operation phase"





X_1	X_2	...	X_n	C
a	b	...	b	+
b	b	...	a	+
a	a	...	b	+
b	a	...	b	+
a	b	...	a	+
b	a	...	a	+
a	a	...	b	+
a	b	...	a	+
a	b	...	b	+
b	a	...	b	+
b	b	...	a	+
a	a	...	b	+
b	a	...	a	+

TRAINING SET

ONE CLASS CLASSIFICATION – OUTLIER DETECTION –

- One category: forms a representative sample
- Only "normal behaviour" samples in training time

- Training phase: model the "normal" behaviour
- Prediction phase: detect "deviations" from the "normal" model

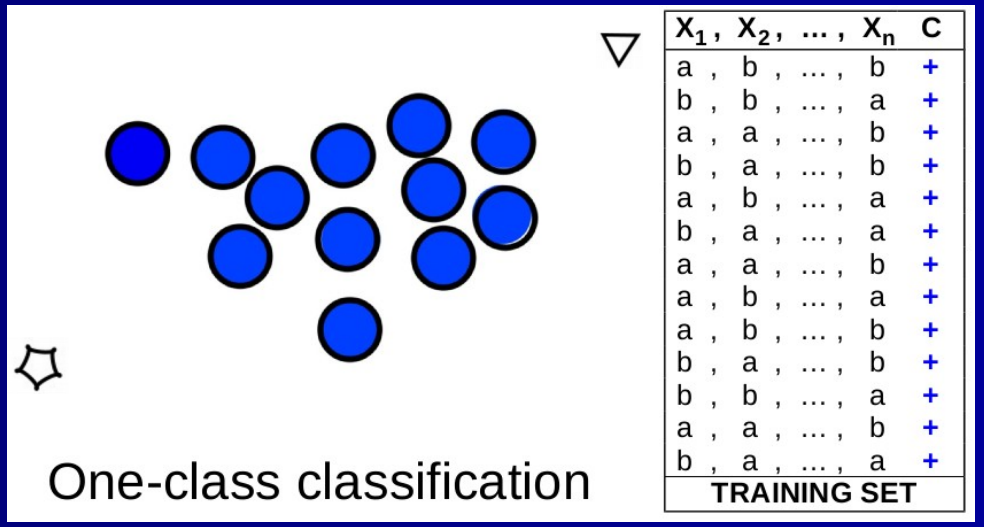
- Model the "dominant" class and "isolate" outliers in "operation phase"

ARTICLE

One-Class Classification of Text Streams with Concept Drift

Authors: Yang Zhang, Xue Li, Maria Orlowska Authors Info & Affiliations

Publication: ICDMW '08: Proceedings of the 2008 IEEE International Conference on Data Mining Workshops • December 2008 • Pages 116–125 • https://doi.org/10.1109/ICDMW.2008.54



X_1	X_2	...	X_n	C
a	b	...	b	+
b	b	...	a	+
a	a	...	b	+
b	a	...	b	+
a	b	...	a	+
b	a	...	a	+
a	a	...	b	+
a	b	...	a	+
a	b	...	b	+
b	a	...	b	+
b	b	...	a	+
a	a	...	b	+
b	a	...	a	+

TRAINING SET

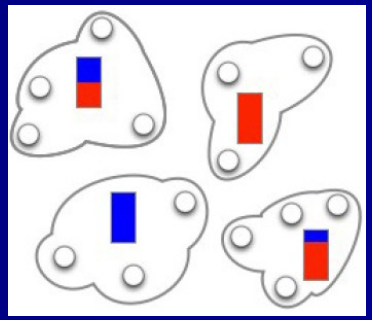
LEARNING with LABEL PROPORTIONS

$X_1, X_2, \ldots, X_n$	C
a , b , … , b	
b , b , … , a	0.5
a , a , … , b	0.5
b , a , … , b	
a , b , … , a	
b , a , … , a	0.0 1.0
a , b , … , a	
a , a , … , b	
a , b , … , b	0.25
b , a , … , b	0.75
b , a , … , a	
a , a , … , b	
b , b , … , a	1.0 0.0
a , a , … , a	

Supervised Learning by Training on Aggregate Outputs

David R. Musicant, Robert Atlas, Janara M. Christensen, Jamie F. Olson, Jeffrey M. Rzeszotarski, Emma R. D. Turetsky

Abstract—Supervised learning is a classic data mining problem where one wishes to be able to predict an output value associated with a particular input vector. We present a new twist on this classic problem where, instead of having the training set contain an individual output value for each input vector, the output values in the training set are only given in aggregate over a number of input vectors. This new problem arose from a particular need in learning on mass spectrometry data, but could easily apply to situations when data has been aggregated in order to maintain privacy. We provide a formal description of this new problem for both classification and regression. We then examine how k-nearest neighbor, neural networks, support vector machines, and decision trees can be adapted for this problem.



LABEL PROPORTIONS APPLICATIONS

Embryo selection in Assisted Reproductive Technologies (ART)

Two steps:

- **Transfer**: step in which one or several embryos are placed into the uterus of the patient.
- **Implantation**: step in which pregnancy is established (by one or several embryos).

Application	MILlp problem
Transfered embryos	Dataset
Implanted or not	Class labels
ART process	Bag
Number of children	Label proportions

Article

SMMR STATISTICAL METHODS IN MEDICAL RESEARCH

Fitting the data from embryo implantation prediction: Learning from label proportions

Statistical Methods in Medical Research
0(0) 1–11

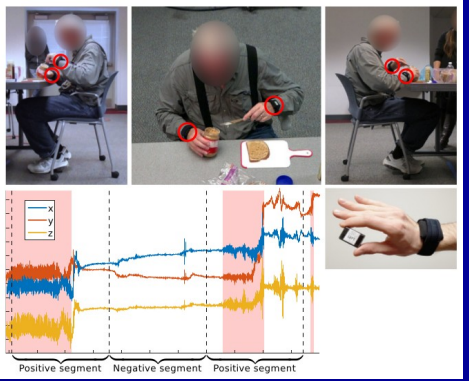
DOI: 10.1177/0962280216651098
smm.sagepub.com
SAGE

Jerónimo Hernández-González,[1] Iñaki Inza,[1] Lorena Crisol-Ortíz,[2] María A Guembe,[2] María J Iñarra[2] and Jose A Lozano[1,3]

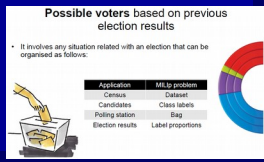
Weakly-supervised Learning for Parkinson's Disease Tremor Detection

Ada Zhang[1], Alexander Cebulla[2], Stanislav Panev[1], Jessica Hodgins[1], and Fernando De la Torre[1]
[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[2] ETH Zurich, Switzerland

Abstract— **Continuous, automated monitoring of Parkinsons Disease (PD) symptoms would provide clinicians with more information to understand their patients' disease progression and adjust treatment protocols, thereby improving PD care. Collecting precisely labeled data for Parkinson's symptoms, such as tremor, is difficult. Therefore, algorithms for monitoring should only require weakly-labeled training data. In this paper, we evaluate five standard weakly-supervised algorithms and propose a "stratified" version of three of the algorithms, which take advantage of knowing the approximate amount of tremor within each segment. In particular, we analyze PD tremor detection performance as training segments increase in length from 30 seconds to 10 minutes, and labels thereby become less precise. As segment length increases to 10 minutes, standard algorithms are not able to discriminate tremor from non-tremor. However, our stratified algorithms, which can make use of more nuanced labels, show little decrease in performance as segment length increases.**

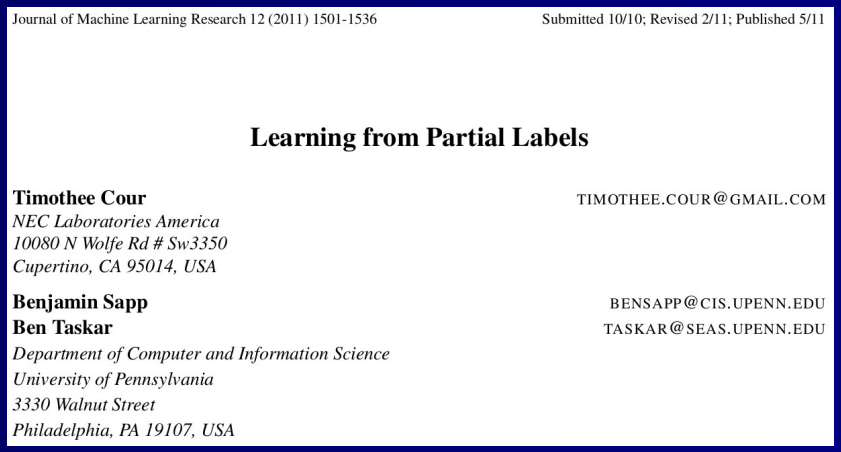


I. Introduction



CLASSIFICATION WITH PARTIAL LABELS

Each instance comes annotated with several class labels but only one of them is valid.



Journal of Machine Learning Research 12 (2011) 1501-1536 Submitted 10/10; Revised 2/11; Published 5/11

Learning from Partial Labels

Timothee Cour TIMOTHEE.COUR@GMAIL.COM
NEC Laboratories America
10080 N Wolfe Rd # Sw3350
Cupertino, CA 95014, USA

Benjamin Sapp BENSAPP@CIS.UPENN.EDU
Ben Taskar TASKAR@SEAS.UPENN.EDU
Department of Computer and Information Science
University of Pennsylvania
3330 Walnut Street
Philadelphia, PA 19107, USA

$X_1, X_2, \ldots, X_n$	C
a , b , … , b	a,b,c
b , b , … , a	a,c
a , a , … , b	d
b , a , … , b	b,c
a , b , … , a	a,d
b , a , … , a	a,b,d
a , a , … , b	b,c,d
a , b , … , a	c
a , a , … , b	b,c
b , a , … , a	b
a , a , … , a	a,b

PROBABILISTIC LABELS

LABEL DISTRIBUTIONS

X_1 , X_2 , … , X_n	c_1	c_2	c_3
a , b , … , b	0.3	0.3	0.4
b , b , … , a	0.4	0.2	0.4
a , a , … , b	0	1	0
a , b , … , a	0.7	0.2	0.1
b , a , … , a	0.5	0.5	0
a , a , … , b	0.3	0.1	0.6
a , b , … , a	0.4	0.2	0.4
a , b , … , b	0.7	0.2	0.4
b , a , … , b	0.9	0.1	0
b , b , … , a	0.6	0.2	0.2

Learning from data with uncertain labels by boosting credal classifiers

Benjamin Quost
HeuDiaSyC laboratory
deptartment of Computer Science
Compiègne University of Technology
Compiègne, France
quostben@hds.utc.fr

Thierry Denœux
HeuDiaSyC laboratory
deptartment of Computer Science
Compiègne University of Technology
Compiègne, France
tdenoeux@hds.utc.fr

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 7, JULY 2016

Label Distribution Learning

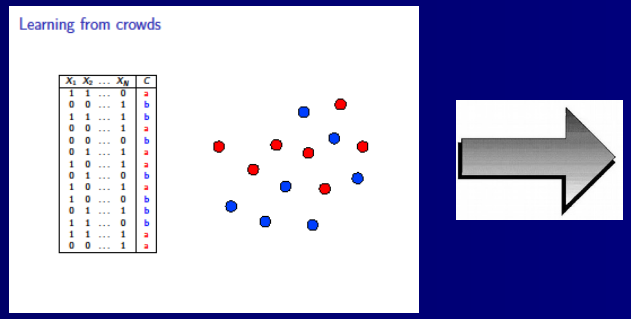
Xin Geng, *Member, IEEE*

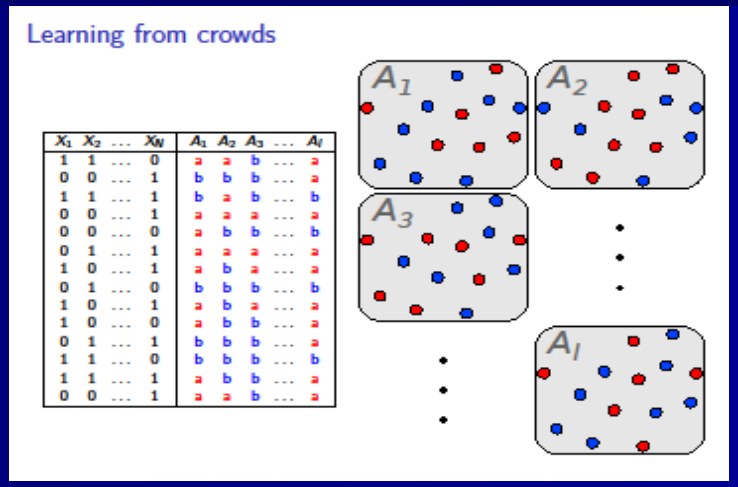
AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

LEARNING FROM CROWDS

- *real class for each object is not known: no "golden truth"*

- *domain experts (A_j) annotate their opinion about the label of each object*





Learning from crowds

Motivation

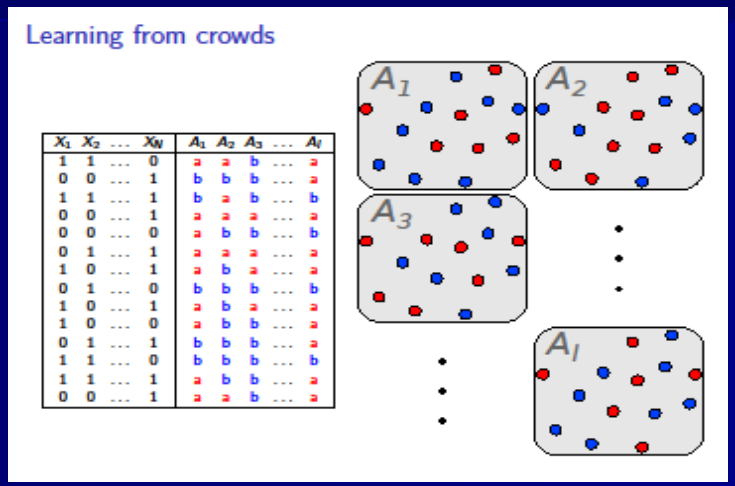
- Expensive/difficult expert labeling
- Recent availability of cheap (non-expert) labeling sources

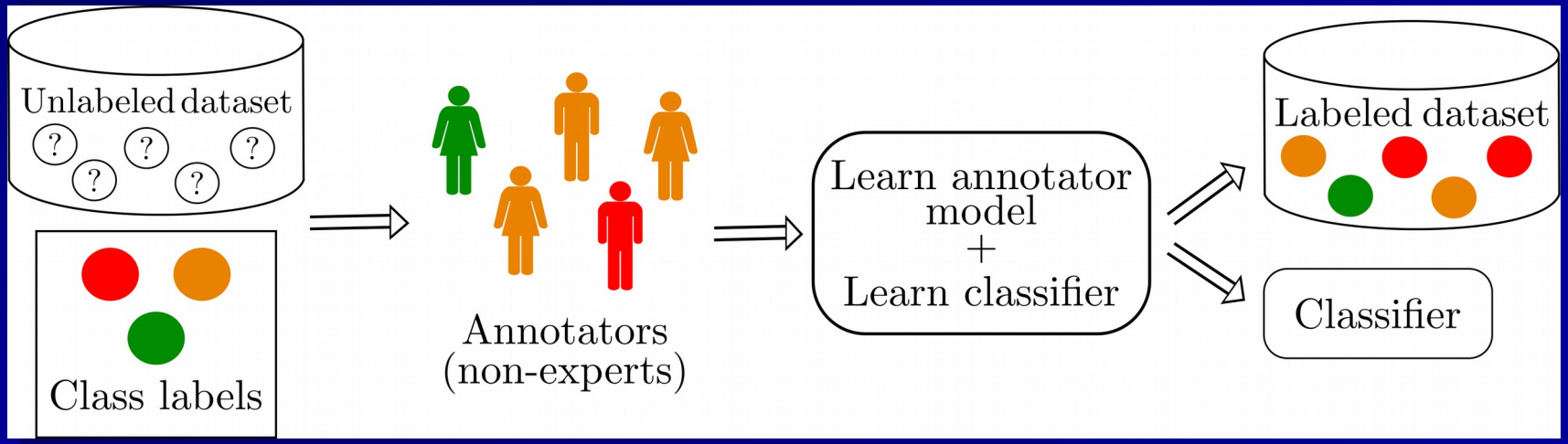
Data collection

- Social networks, games, etc.
- Specific platforms (e.g. Amazon Mechanical Turk)

AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

LEARNING FROM CROWDS





AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

LEARNING FROM CROWDS

Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines

Marta Sabou,* Kalina Bontcheva,† Leon Derczynski,† Arno Scharl*
*MODUL University Vienna
Am Kahlenberg 1, Vienna, Austria
{marta.sabou,arno.scharl}@modul.ac.at
†University of Sheffield
211 Portobello, Sheffield S1 4DP, UK
{K.Bontcheva,L.Derczynski}@dcs.shef.ac.uk

Abstract

Crowdsourcing is an emerging collaborative approach that can be used for the acquisition of annotated corpora and a wide range of other linguistic resources. Although the use of this approach is intensifying in all its key genres (paid-for crowdsourcing, games with a purpose, volunteering-based approaches), the community still lacks a set of best-practice guidelines similar to the annotation best practices for traditional, expert-based corpus acquisition. In this paper we focus on the use of crowdsourcing methods for corpus acquisition and propose a set of best practice guidelines based in our own experiences in this area and an overview of related literature. We also introduce GATE Crowd, a plugin of the GATE platform that relies on these guidelines and offers tool support for using crowdsourcing in a more principled and efficient manner.

AND WHEN ANNOTATIONS ARE NOT FULLY RELIABLE?...

LEARNING FROM CROWDS

A Crowd-Annotated Spanish Corpus for Humor Analysis

Santiago Castro, Luis Chiruzzo, Aiala Rosá, Diego Garat, Guillermo Moncecchi
Grupo de Procesamiento de Lenguaje Natural
Facultad de Ingeniería
Universidad de la República — Uruguay
{sacastro,luischir,aialar,dgarat,gmonce}@fing.edu.uy

Abstract

Computational Humor involves several tasks, such as humor recognition, humor generation, and humor scoring, for which it is useful to have human-curated data. In this work we present a corpus of 27,000 tweets written in Spanish and crowd-annotated by their humor value and funniness score, with about four annotations per tweet, tagged by 1,300 people over the Internet. It is equally divided between tweets coming from humorous and non-humorous accounts. The inter-annotator agreement Krippendorff's alpha value is 0.5710. The dataset is available for general use and can serve as a basis for humor detection and as a first step to tackle subjectivity.

other authors, such as Mihalcea and Strapparava (2005a,b); Sjöbergh and Araki (2007) have tackled humor recognition in English texts, building their own corpora by downloading *one-liners* (one-sentence jokes) from the Internet, since working with longer texts would involve additional work, such as determining humor scope.

The microblogging platform Twitter has been found particularly useful for building humor corpora due to its public availability and the fact that its short messages are suitable for jokes or humorous comments. Castro et al. (2016) built their corpus based on Twitter, selecting nine humorous accounts and nine non-humorous accounts about news, thoughts and curious facts. Reyes et al. (2013) built a corpus for detecting irony in tweets by searching for several hashtags (i.e., #irony, #humor, #education and #politics), which is also used in Barbieri and Saggion (2014) to train a classifier

SUPERVISION MODELS

Table 2
Collection of supervision models.

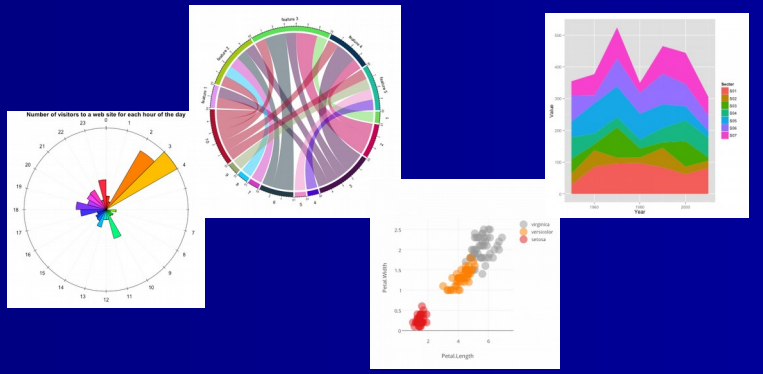
Model	References	Description
Full-supervision	[9,24,34,43]	For each example, complete class information is provided.
Unsupervision	[24]	No class information is provided with the examples.
Semi-supervision	[5]	Part of the examples are provided fully supervised. The rest are unsupervised.
Positive-unlabeled	[4,10,21,32]	Part of the examples are provided fully supervised, all of them with the same categorization. The rest are unsupervised.
Candidate labels	[7,13,16]	For each example, a set of class labels is provided. In this set, the class label(s) that compose the real categorization of the example are included.
Probabilistic labels	[18]	For each example, the probability of belonging to each class label is provided. This probability distribution is expected to assign high probability to the real label(s).
Incomplete	[3,33,42]	For each example, a subset of the labels that compose its real categorization is provided (SIML or MIML, Table 1).
Noisy labels	[2,44]	For each example, complete class information is provided, although its correctness is not guaranteed.
Crowd	[30,40]	For each example, many different non-expert annotators provide their (noisy) categorization.
Mutual label constraints	[19,20,31]	For each **group** of examples, an explicit relationship between their class labels is provided (e.g., all the examples have the same categorization).
Candidate labeling vectors	[22]	For each **group** of examples, a set of labeling vectors (including the real one) is provided. A labeling vector provides a class label for each examples of a group.
Label proportions	[15,25,28]	For each **group** of examples, the proportion of examples belonging to each class label is provided.

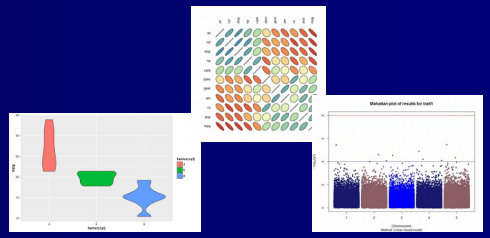
LEARNING SCENARIOS

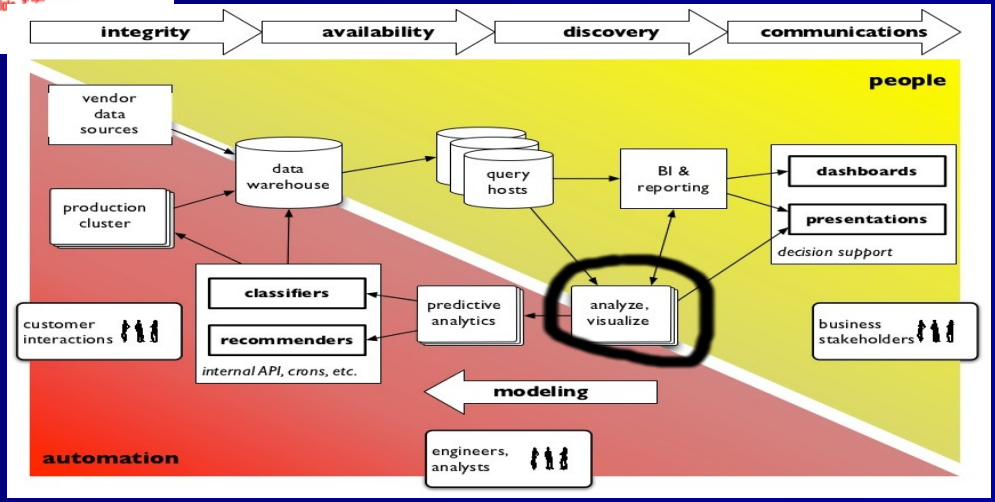
Table 3
Brief description of classification problems and characterization according to the three axes of the taxonomy.

Problem	Description	Application (e.g.)	IL rel.	SUPERVISION MODEL	
				Learning	Prediction
Standard problem [24]	Learning with full categorized examples	Handwritten digit recogn.	SISL	Full-supervision	Unsupervision
Semi-supervised [5]	Learning with categorized and uncategorized examples	Text classification	SISL	Semi-supervision	Unsupervision
Positive-unlabeled [4]	Learning with examples of a category and other uncategorized examples	Spam detection, Gene prediction	SISL	Positive-unlabeled	Unsupervision
Mislabeled data [2]	Learning with maybe wrong-categorized examples	Subjective labeler	SISL	Noisy Labels	Unsupervision
Ambiguous labels [44]					
Partial labels [7]	Learning and prediction with uncategorized examples that have a set of possible categorizations	Classifying photographs with captions	SISL	Candidate labels	Unsupervision / Candidate labels
Multiple labels [18]	Learning with uncategorized examples that, with some probability, belong to a certain categorization	Bioinformatics	SISL	Probabilistic labels	Unsupervision
Partial equivalence relations [19]	Learning with groups of examples of the same/different categorization	Computer vision	SISL	Mutual label constraints	Unsupervision
Full-class set [20]	Prediction for a group of examples, all of them with a different categorization	Automatic attendance recording	SISL	Full-supervision	Mutual label constraints
Label proportions [15]	Learning with groups of examples only knowing how many of them belong to each categorization	Embryo Selection, Polls prediction	SISL	Label proportions	Unsupervision
Aggregate outputs [25]					
Candidate labeling sets [22]	Learning with groups of examples and sets of possible categorizing vectors	Classifying photographs with captions	SISL	Candidate labeling vectors	Unsupervision
Learning from crowds [30,40]	Learning with examples categorized with many candidate noisy categorizations	Image annotation	SISL	Crowd	Unsupervision
Multi-label [34]	Learning with examples that belong to several categorizations at the same time	Film genre prediction	SIML	Full-supervision	Unsupervision
Semi-supervised multi-label [6]	Learning with examples categorized with multiple labels or uncategorized	Text categorization	SIML	Semi-supervision	Unsupervision
ML with weak label [33]	Learning with examples categorized with a subset of the real multiple labels	Image annotation	SIML	Incomplete	Unsupervision
ML incomplete class [3]					
Set classification [26]	Prediction for a group of examples, all of them with the same categorization	Face recognition with multiple photos	SIML	Full-supervision	Mutual label constraints
MIL [9]	Learning with multiple-instance examples that are positive if at least one of their instances is	Molecule activation prediction	MISL	Full-supervision	Unsupervision
G-MIL [39]	Learning with examples represented by several instances with generalized function for positives	Key-and-lock prediction problem	MISL	Full-supervision	Unsupervision
MISSL [29]	Learning with categorized and uncategorized multiple-instances examples	Content-based image retrieval	MISL	Semi-supervision	Unsupervision
MIML [43]	Learning with examples represented with several instances that belong to several categorizations	Classifying texts, images or videos	MIML	Full-supervision	Unsupervision
SSMIML [41]	Learning with multiple-instance examples categorized with multiple labels or uncategorized	Video annotation	MIML	Semi-supervision	Unsupervision
MIML with weak labels [42]	Learning with multiple-instance examples categorized with a subset of the real multiple labels	Image annotation	MIML	Incomplete	Unsupervision

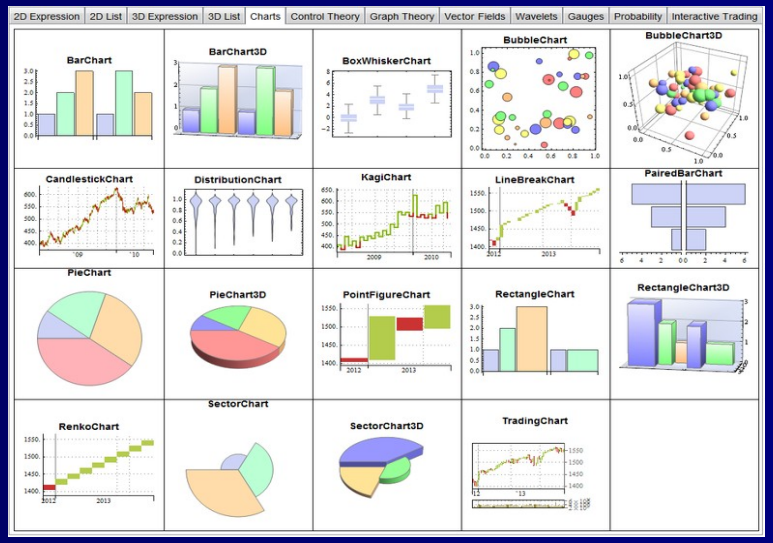
IS NEEDED TO LEARN "A CRYSTAL BALL"? OR JUST VISUALIZE MY DATA?







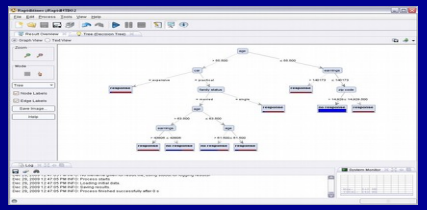
DATA VISUALIZATION
DATA EXPLORATION



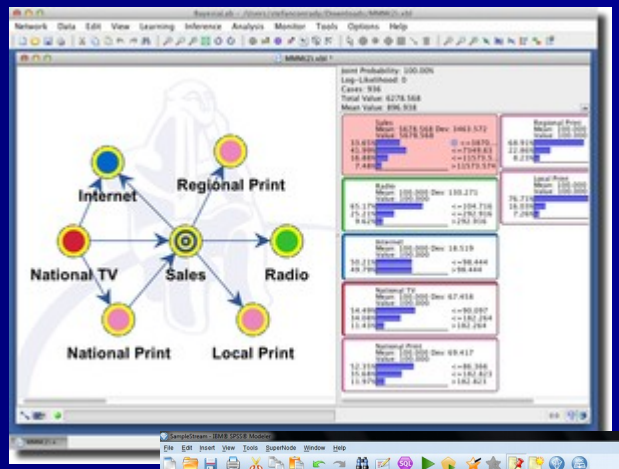
- Top 10 data visualization tools
- List of commercial and free visualization tools
- 11 steps for data exploration in R: (with codes)
- The R graph gallery: a collection of 200 graphs in R
- 7 simple visualizations you should know in R

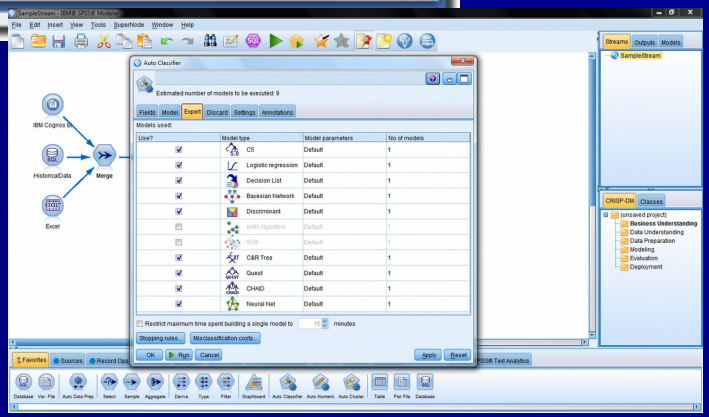
RESOURCES
SOFTWARE & DATASETS

kaggle

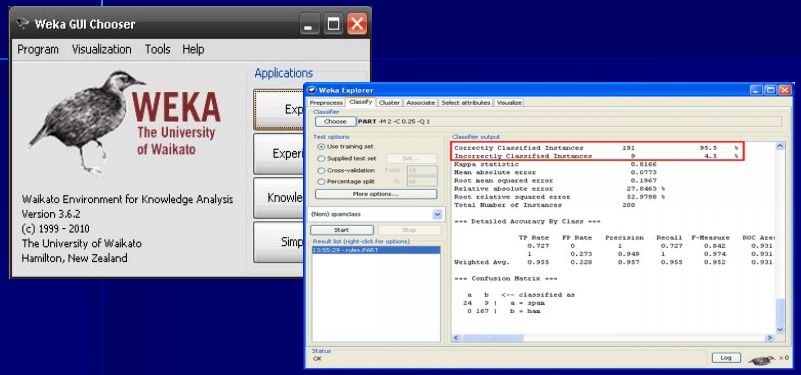


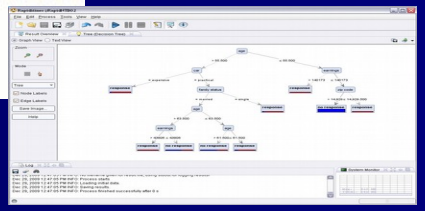
COMERCIAL SOFTWARE FOR DATA MINING





FREE SOFTWARE FOR DATA MINING





- Software suites for data mining, analytics and knowledge discovery
- 11 open source tools to make the most of machine learning
- Top 10 machine learning projects in GitHub
- 50 useful machine learning & prediction APIs
- Classification software: a list
- Top 15 frameworks for machine learning experts
- Bayesian networks and Bayesian classifier software