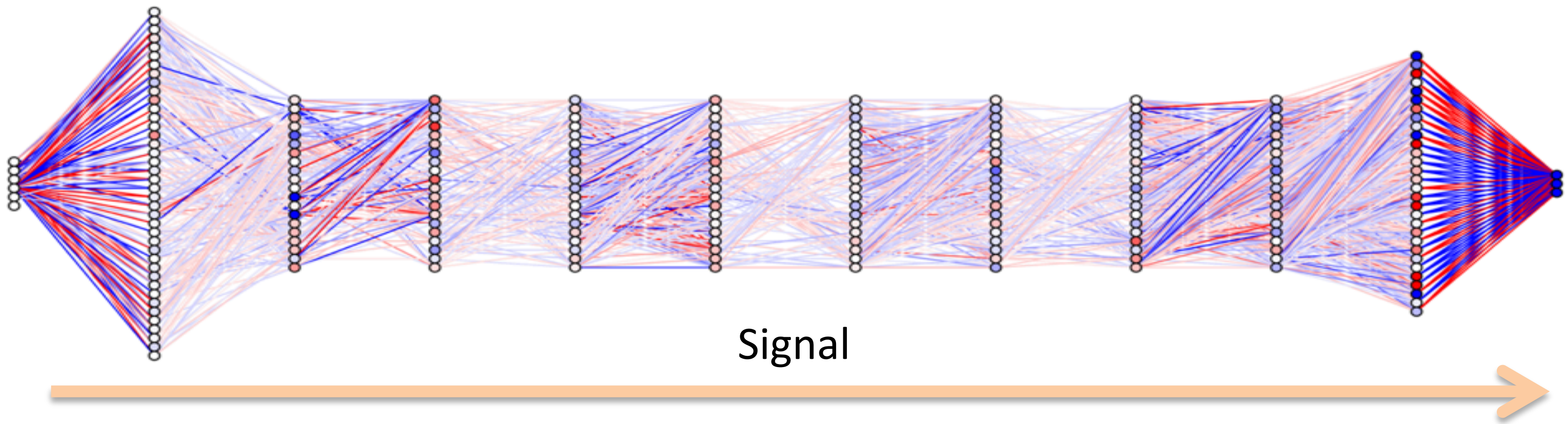
The SuperVision network

Image classification with deep convolutional neural networks

http://image-net.org/challenges/LSVRC/2012/supervision.pdf

- 7 hidden “weight” layers
- 650K neurons
- **60M** parameters
- 630M connections



The SuperVision network

Image classification with deep convolutional neural networks

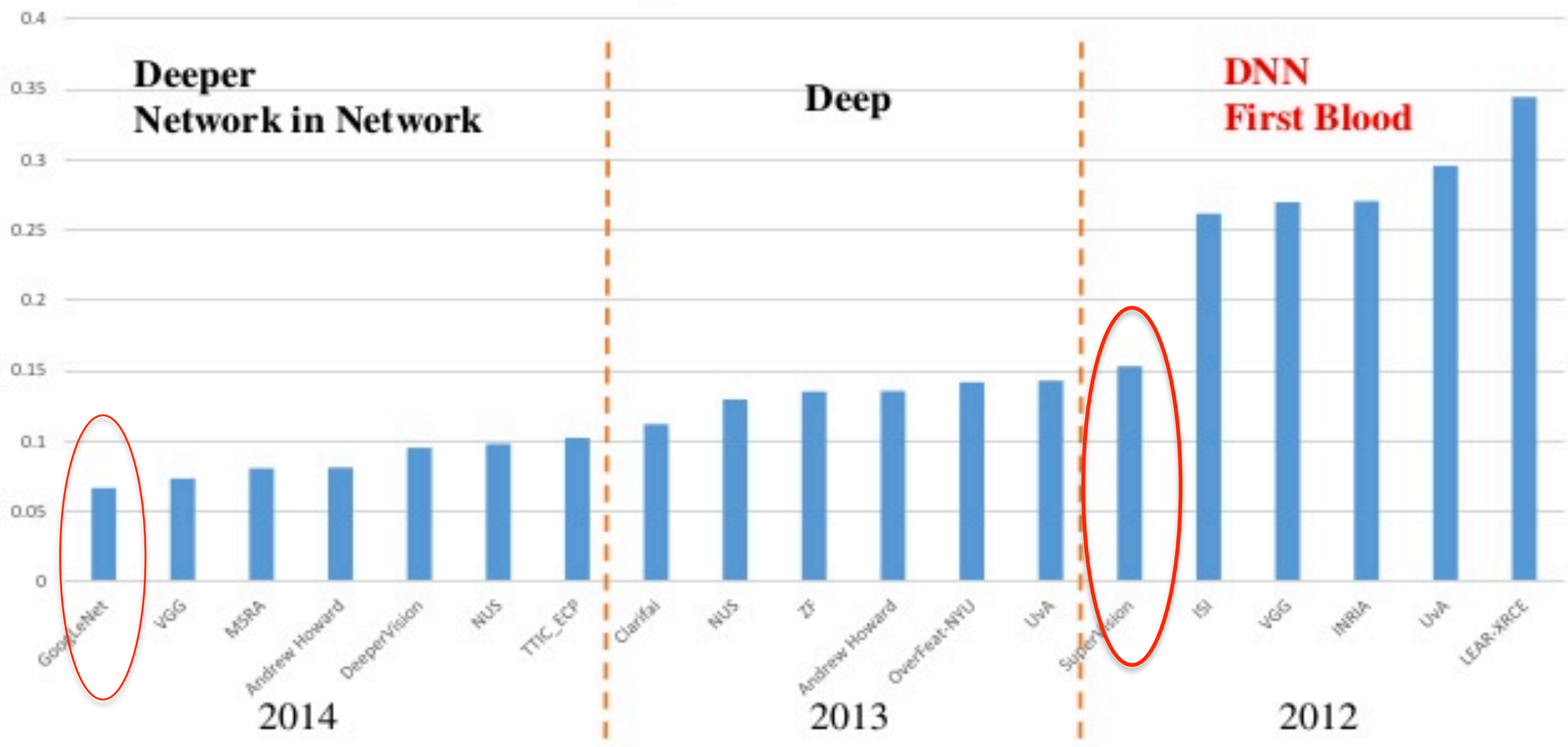
http://image-net.org/challenges/LSVRC/2012/supervision.pdf

- 7 hidden "weight" layers
- 650K neurons
- 60M parameters
- 630M connections

- Rectified Linear Units (ReLU)
- Overlapping pooling
- Dropout trick
- Randomly extracted 224x224 patches for more data

ILSVRC

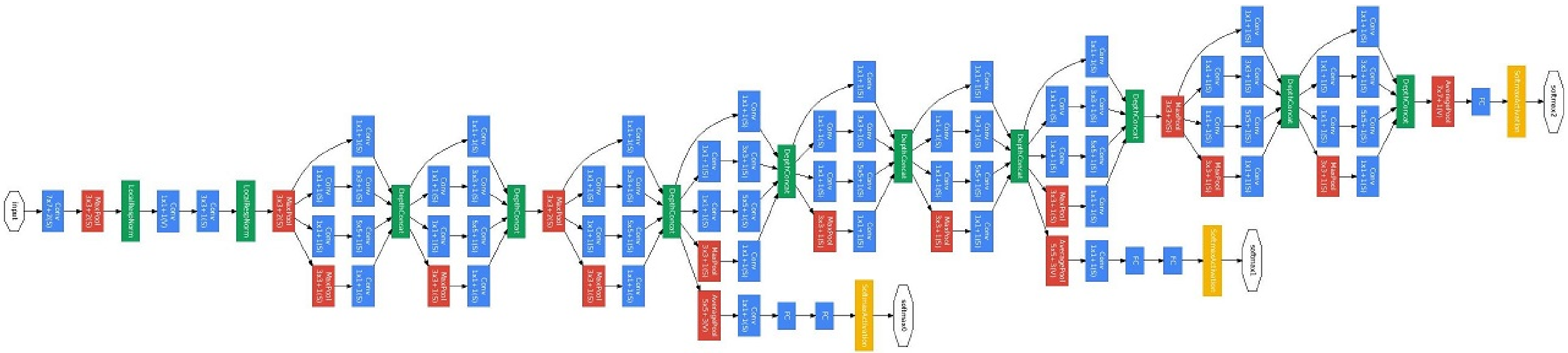
ImageNet classification error throughout years and groups



Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/

GoogleNet

- Un **mécano** de réseaux de neurones



Illustration

Système développé par Google et U. de Stanford

- Reconnaissance de visages
  - Sous conditions de lumière diverses
  - Sous tout angle

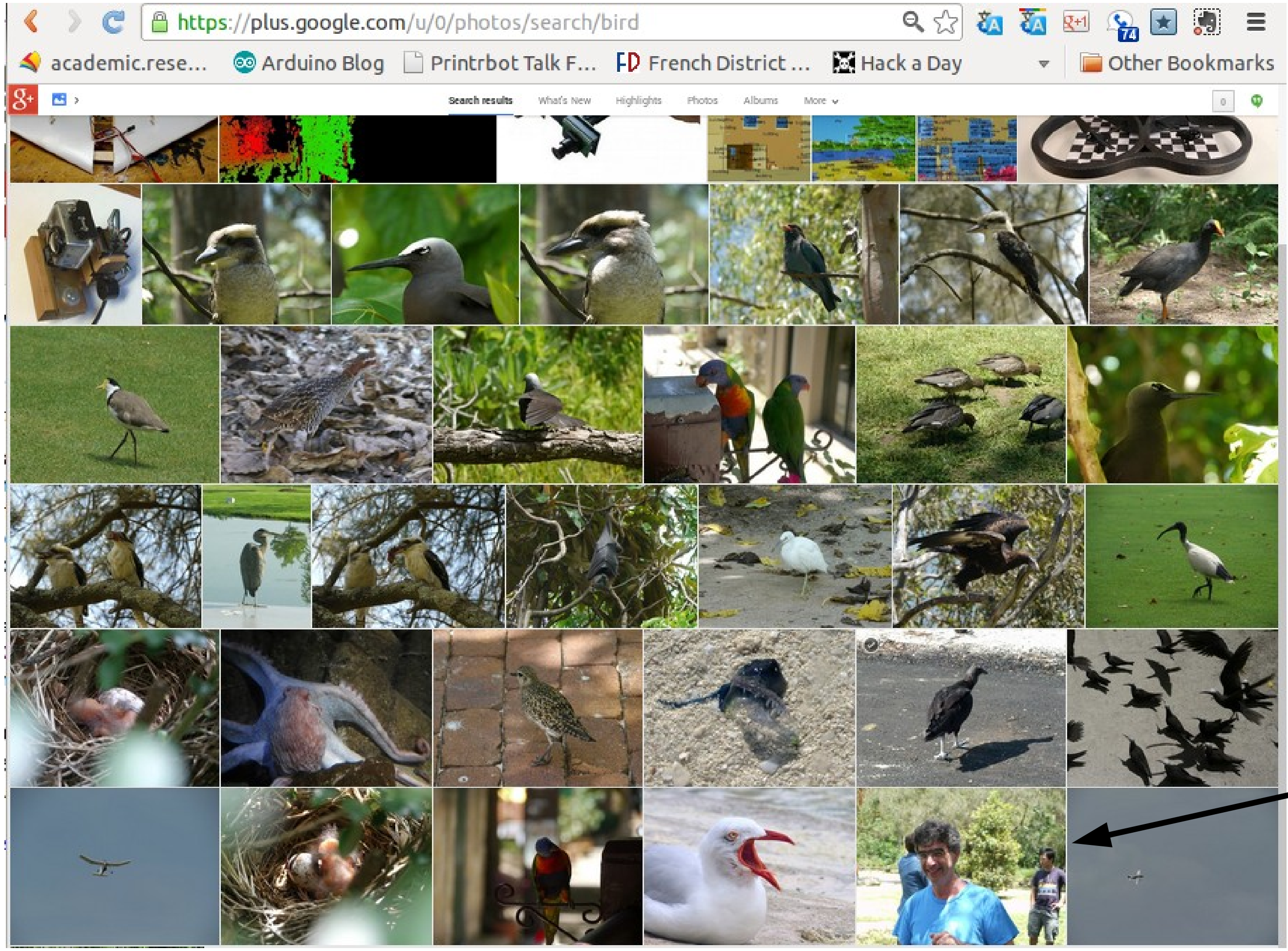
- Apprentissage non supervisé
  - 9 couches ; 10^9 connexions
  - 10 millions d'images
  - 3 jours de calcul sur 16 000 processeurs
- Amélioration des performances de 70% / état de l'art

Object recognition

[Krizhevsky, Sutskever and Hinton (2012)]

Object retrieval. ConvNet-Based Google+ Photo Tagger

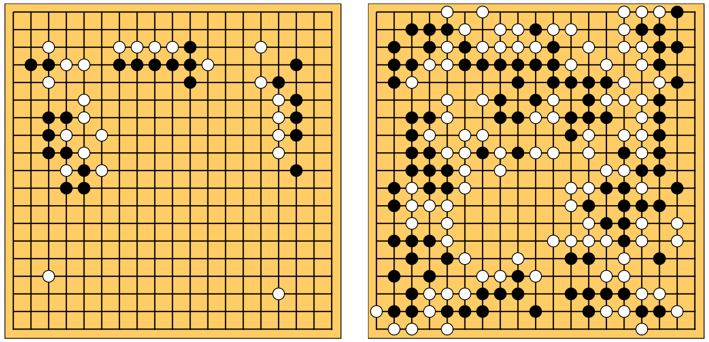
Searched my personal collection for "bird"

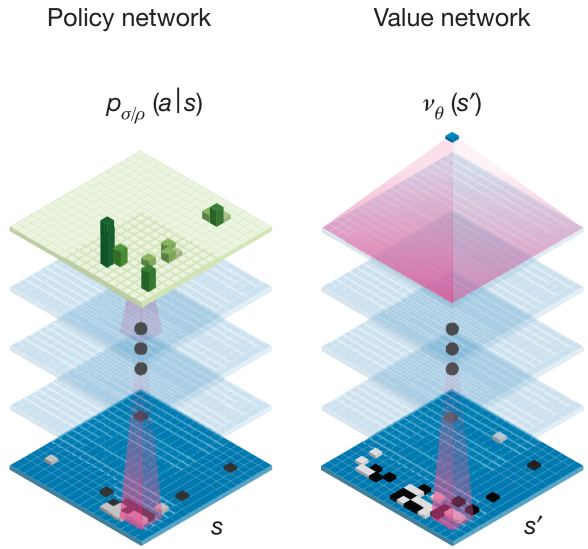


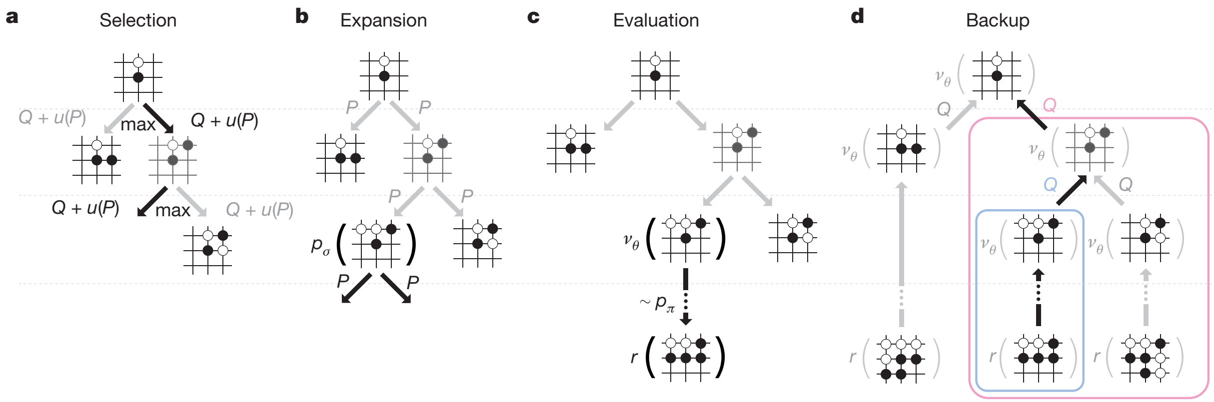
Samy
Bengio
???

Game playing with Reinforcement Learning

- E.g. AlphaGo





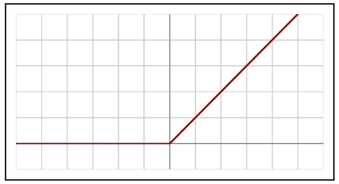


Beaucoup de « recettes de cuisine »

Les grandes idées nouvelles

- La **rétro-propagation classique ne marche pas** avec un grand nombre de couches (trop dilué)

Unité ReLU



Drop Out

- **Risque de sur-apprentissage** avec un nombre gigantesque de paramètres

Auto-encoder : couches apprises en non-supervisé

Réseaux à convolutions :

imposer une structure (avec motifs répétitifs) au réseau

Le « Drop Out »

- Classiquement :
  - Les poids sont initialisés aléatoirement et difficiles à ajuster

- **Principe** :
  - **Débrancher** des **neurones aléatoirement** lors de l'apprentissage (tirage aléatoire à chaque nouvel exemple)
  - Paramètre par défaut : 0.5

Le « Drop Out »

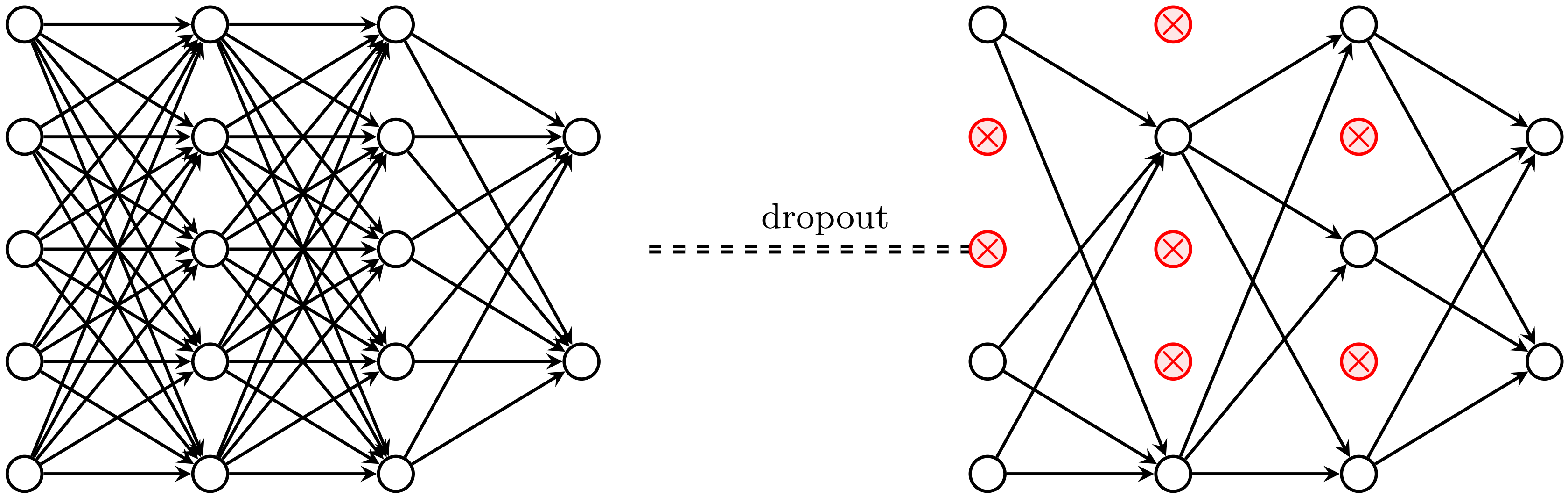


FIGURE 9: Schéma descriptif du dropout

…

Techniques d’optimisation

- **Très grande** activité de recherche

Du nouveau dans le hardware

Technologie : du hardware spécifique

- **GPU** *(Graphics Processing Unit)* historiquement utilisé comme carte graphique pour les jeux vidéo
- Hardware spécialisé dans le **calcul matriciel** hautement **parallélisé**
- Des algorithmes très contraints : des **milliers « threads »** qui doivent exécuter **la même opération** simultanément

Des implémentations modernes des réseaux de neurones permettent de tirer partie des GPU *(ex: Torch 7, Cuda conv-net, Theano, TensorFlow …)*

Un « bolide » délicat à piloter

Requiert

1. beaucoup de **données** (en général)
   - Des millions d'images
   - Des dizaines de milliers de documents
2. du **savoir-faire** (des data scientists)
   - Nombreuses « **astuces** » d'ingénierie
   - Utilisation de réseaux déjà appris (**transfert**)
   - L'état de l'art **progresse très vite**
3. des machines adaptées
   - Puissance **calcul** : clusters et/ou cartes graphiques
   - **Mémoire** centrale importante (≥ 128 Go)

Enseigné dans certaines écoles et universités

Les comportements **étranges**

Sait-on pourquoi ça marche ... Quand ça marche

Quelque chose de troublant

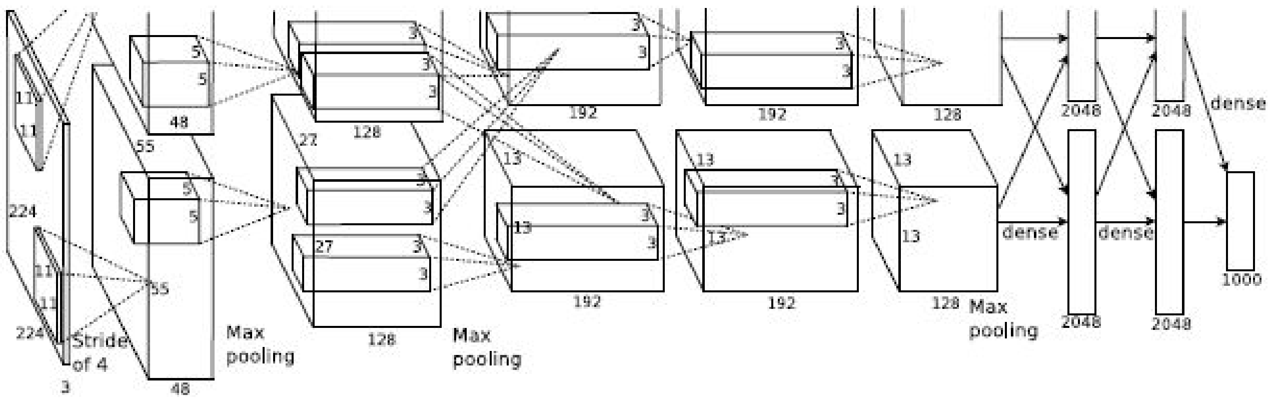
– C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals (ICLR, **May 2017**).
"Understanding deep learning requires rethinking generalization"

Quelque chose de troublant

– C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals (ICLR, **May 2017**).
"Understanding deep learning requires rethinking generalization"

Extensive experiments on the classification of images

– The AlexNet (**> 1,000,000 parameters**) + 2 other architectures



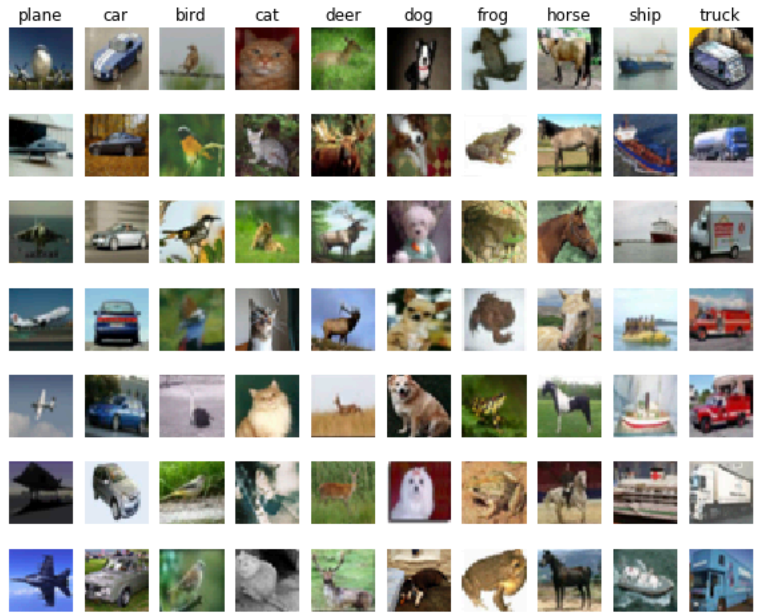
– The **CIFAR-10 data set**:

- **60,000** images categorized in **10 classes** (50,000 for training and 10,000 for testing)
- Images: 32x32 pixels in 3 color channels

Quelque chose de troublant

Experiments

1. **Original dataset** without modification
   - Results ?
     - **Training** accuracy = 100%   ;   **Test** accuracy = 89%
     - Speed of convergence ~ 5,000 steps



Quelque chose de troublant

Experiments

1. **Original dataset** without modification
   - Results ?
     - **Training** accuracy = 100%   ;   **Test** accuracy = 89%
     - Speed of convergence ~ 5,000 steps

Expected behavior if the **capacity** of the hypothesis space is **limited**

i.e. the system **cannot** fit any (arbitrary) training data

$$\forall h \in \mathcal{H}, \forall \delta \leq 1 : \quad P^m \left[R(h) \; \leq \; \widehat{R}(h) \; + 2\,\widehat{Rad}_m(\mathcal{H}) + \; 3\sqrt{\frac{\ln(2/\delta)}{m}} \right] \; > \; 1 - \delta$$

Troubling findings

Experiments

1. **Original dataset** without modification
   - Results ?
     - **Training** accuracy = 100% ; **Test** accuracy = 89%
     - Speed of convergence ~ 5,000 steps

2. Random **labels**

   !!!

     - **Training** accuracy = 100% !!?? ; **Test** accuracy = 9.8%
     - Speed of convergence = similar behavior (~ 10,000 steps)

Troubling findings

Experiments

1. **Original dataset** without modification
   - Results ?
     - **Training** accuracy = 100% ; **Test** accuracy = 89%
     - Speed of convergence ~ 5,000 steps

2. Random **labels**
   - **Training** accuracy = 100% !!?? ; **Test** accuracy = 9.8%
   - Speed of convergence = similar behavior (~ 10,000 steps)

3. Random **pixels**
   - **Training** accuracy = 100% !!?? ; **Test** accuracy ~ 10%
   - Speed of convergence = similar behavior (~ 10,000 steps)

Now, we are in trouble!!

Troubling findings

- Deep NNs can accommodate ANY training set

Can grow without limit!!

$$\forall h \in \mathcal{H}, \forall \delta \leq 1 : \quad P^m \left[R(h) \;\leq\; \widehat{R}(h) \;+\; 2\, \widehat{Rad}_m(\mathcal{H}) \;+\; 3 \sqrt{\frac{\ln(2/\delta)}{m}} \right] \;>\; 1 - \delta$$

But then,

why are deep NNs ***so good on image classification tasks?***

Plan

Ce que l’on sait faire.

Sait-on d’ailleurs vraiment le faire ?

Ce qui interroge.

Ce qui reste à faire.

Un peu de recul :

Que sait-on faire

et **où sont** les **limites** ?

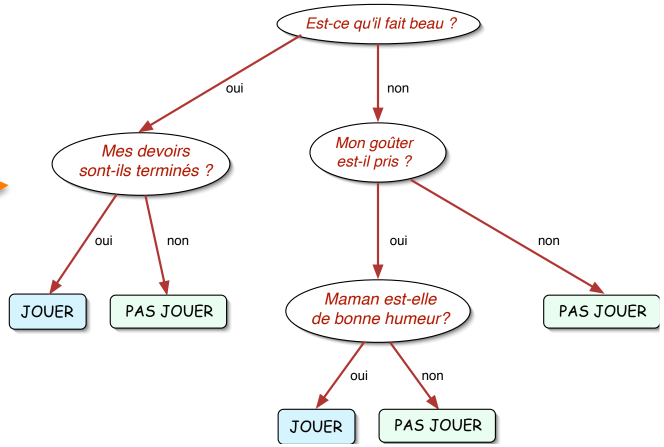
Ce que l’on sait faire

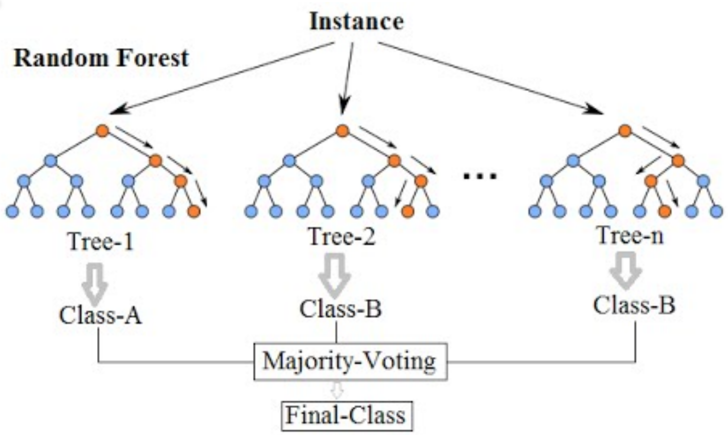
Ce que l'**on sait faire**

- Apprentissage prédictif
  - En environnement **stationnaire**
  - À partir de (très) **nombreux exemples**
  - Classification / régression
- Apprentissage descriptif
  - Problème de la **validation**
- Apprentissage de recommandation
- Apprentissage de contrôle / commande (app. par renforcement)

Nombreuses méthodes d'apprentissage

- Réseaux de neurones
- Arbres de décision
- Méthodes d'ensemble
- Apprentissage bayésien
- Chaînes de Markov et HMM
- Outils de fouille de données
- …

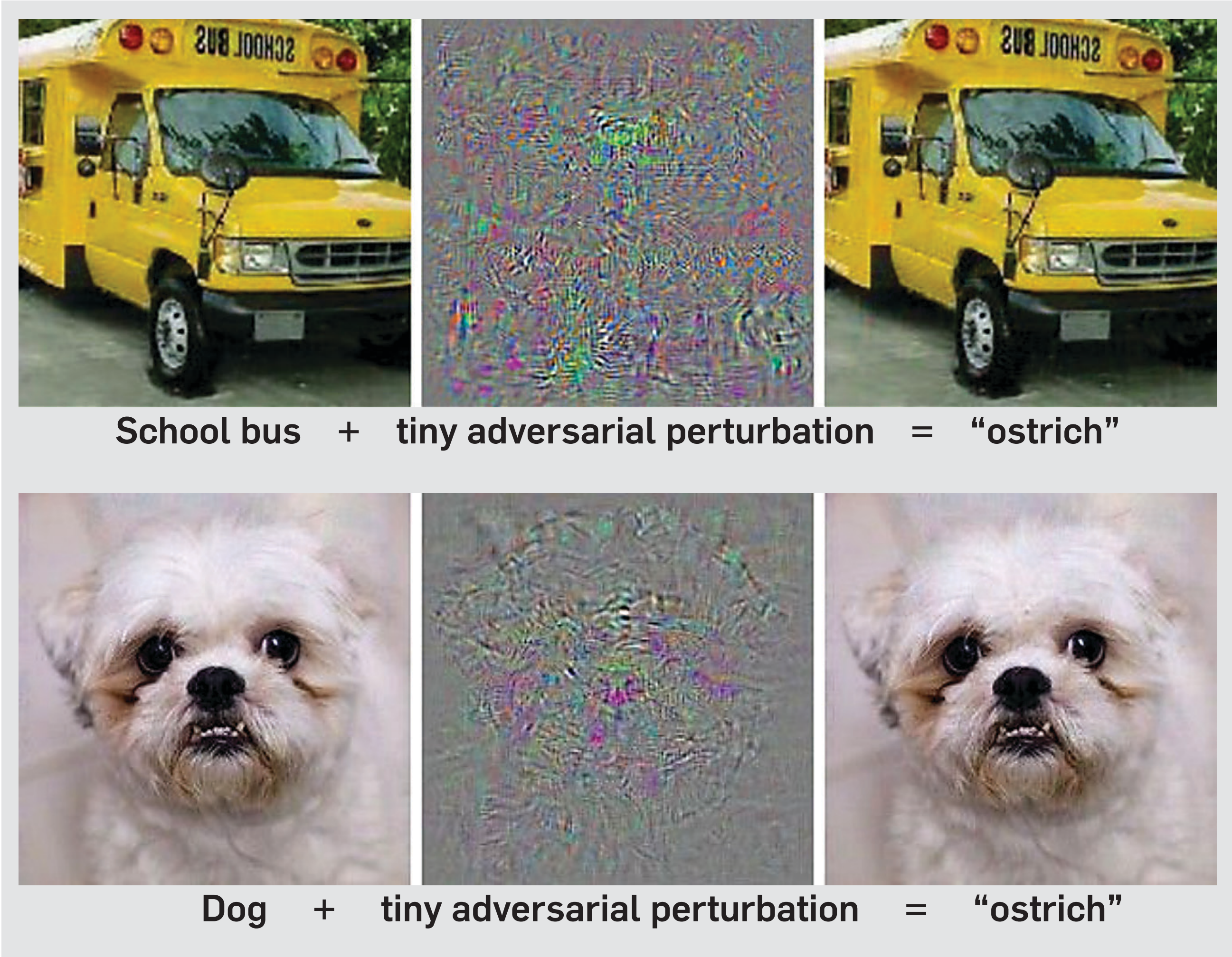




Ce que l'on sait faire

Quoique !?

Adversarial learning



Adversarial input can fool a machine-learning algorithm into misperceiving images.

Explanations and deep neural networks

Optical illusions: how to explain them?

Boxer: 0.40 Tiger Cat: 0.18

(a) Original image

Airliner: 0.9999

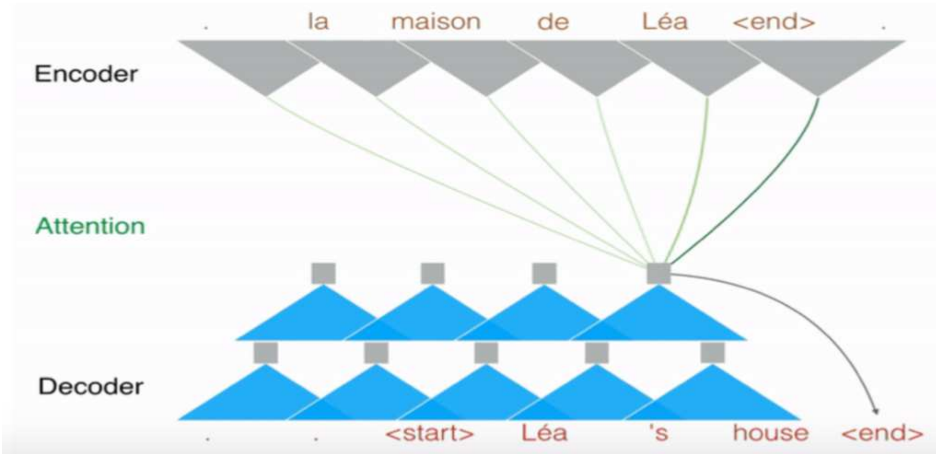
(b) Adversarial image

!!??

[Selvaraju et al. (2017) « *Grad-CAM: Visual explanations from deep networks via gradient-based localization* »]

Machine translation

- Still far from perfect, but …



From Hofstädter (2018)

Annotation d'images

- 

Figure 2.11: "A group of young people playing a game of frisbee"—that caption was written by a computer with no understanding of people, games or frisbees.

Exemple en médecine

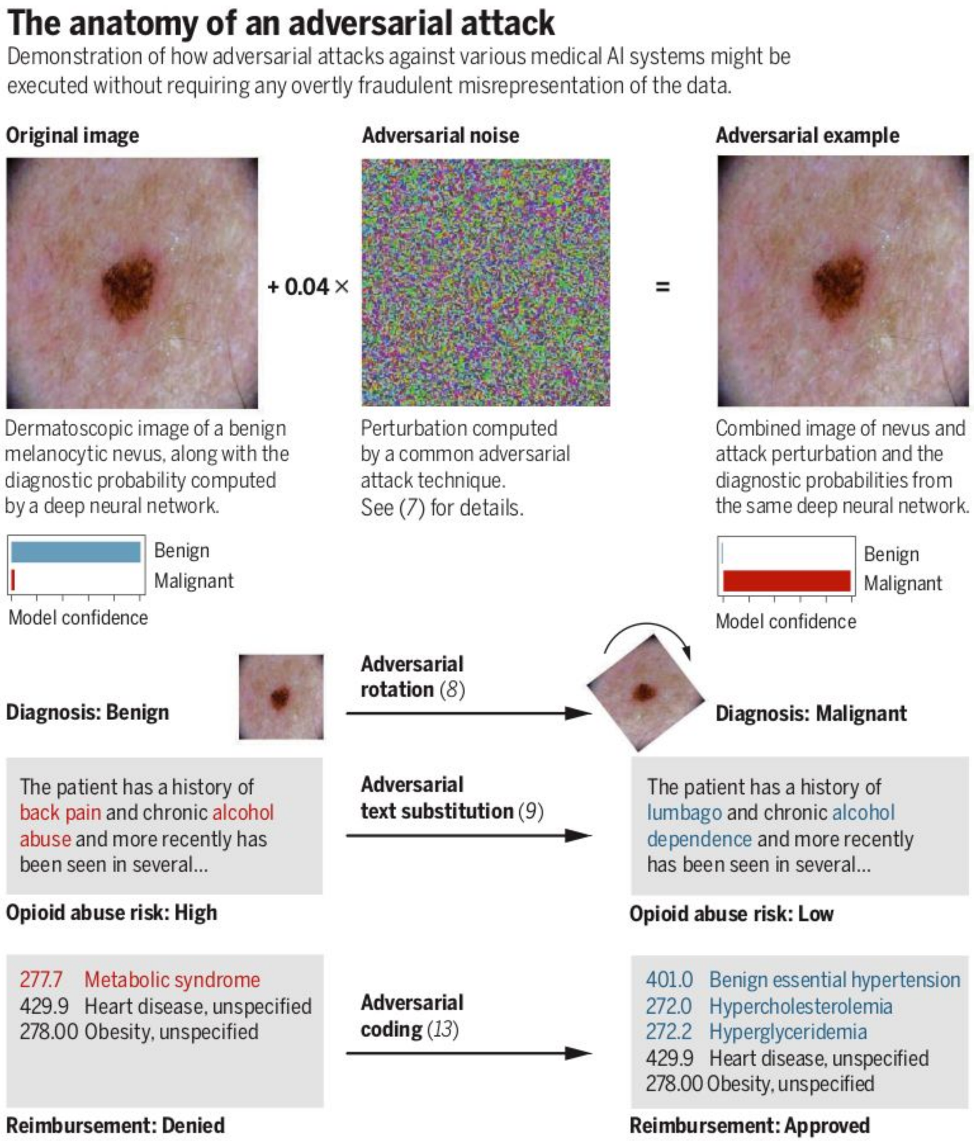
MACHINE LEARNING

Science

Adversarial attacks on medical machine learning

Emerging vulnerabilities demand new conversations

22 March 2019



Voiture dans une piscine

- … ou pas de voiture … ?

Is this less of a car
because the context is wrong?

[Léon Bottou (ICML-2015, invited talk) « *Two big challenges in Machine Learning* »]

L'IA comprend-t-elle ?

https://www.youtube.com/watch?v=QPSgM13hTK8&t=117

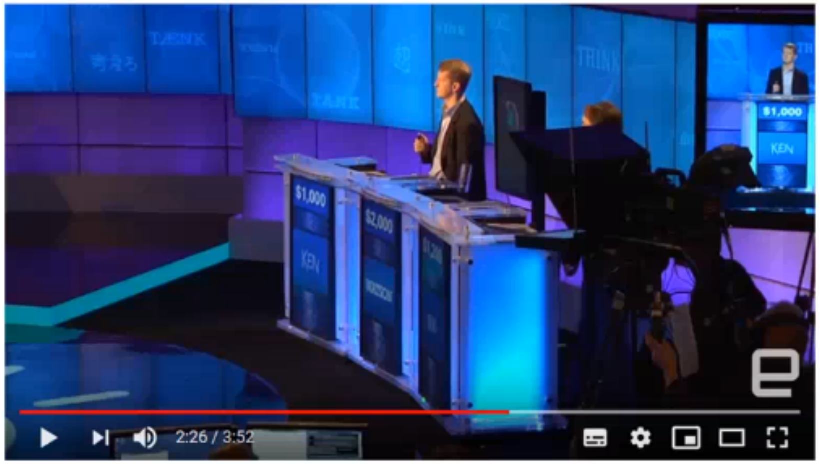
WATSON et le jeu Jeopardy! (2011)

Jeopardy! In the category U.S. cities:

– "Its largest airport was named for a World War II hero; its second largest, for a World War battle."

➢ What is *Toronto*?

New-York!!



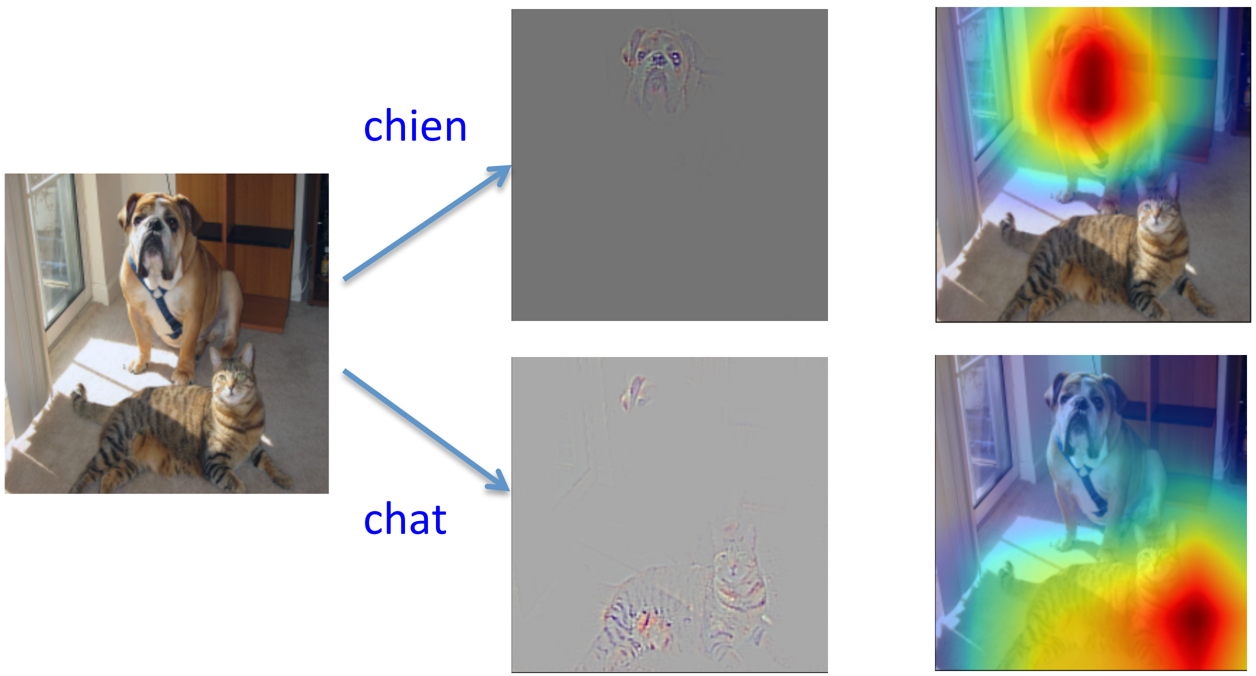
IBM's Watson Supercomputer Destroys Humans in Jeopardy | Engadget

Sait-on expliquer une conclusion ?

Explication et réseaux de neurones profonds

Identification de classes d'objets dans une image

- Ici deux classes : « **chien** » et « **chat tigré** »



[Selvaraju et al. (2017) « *Grad-CAM: Visual explanations from deep networks via gradient-based localization* »]

Les assistants « intelligents » : quelle assistance ?

Décident

- L'attribution d'un **crédit**
- La **sélection** dans les filières (e.g. Parcours Sup)

Suggèrent un diagnostic

- **Médical**
- **Légal**

- Quelle transparence ?
  - Les « **raisons** » de la décision
  - Peut-on les **orienter**, les **remettre en cause** ?

Question supplémentaire

- Comment **garder une trace des « raisons »** d'une décision quand le système **apprend** en permanence et évolue donc ?

AgroParisTech

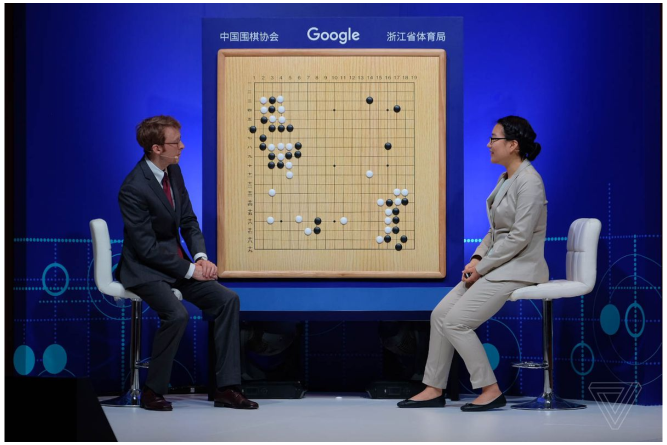
Le cas AlphaGo

- Un joueur « extraterrestre »
- Un jeu stupéfiant
- Révolutionne la manière de jouer
- Effervescence dans les écoles de go

Le cas AlphaGo : comprendre

Fan Hui, Gu Li, Zhou Ruyang (très forts joueurs de Go) se reconvertissent dans l'analyse des parties jouées par AlphaGo

- Sorte d'exégèse. Explications a posteriori
- Nécessaire pour
  - La communication
  - L'enseignement

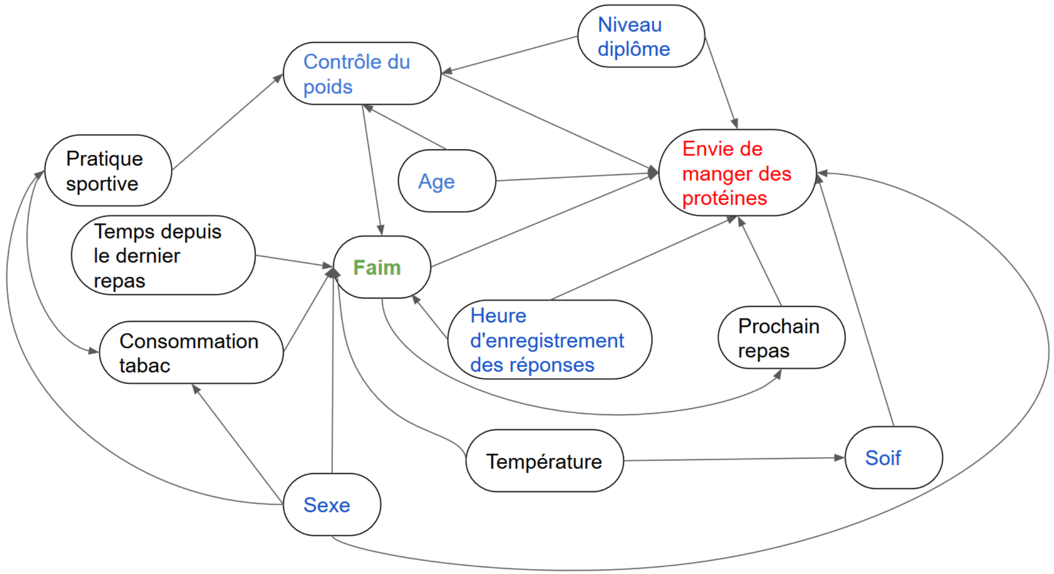


Et même AlphaGo peut se tromper

La recherche de relations causales

- Qu'est-ce qui cause l'appétence pour des plats protéinés ?

  - La **faim** ?
  - **L'heure** dans la journée ?
  - Le **genre** ?
  - L'aspect **visuel** ?
  - L'aspect **olfactif** ?
  - La richesse en **protéines** des **repas précédents** ?
  - …



Conclusions

Le paradigme actuel

- Induire nécessite d'avoir des biais
- La **théorie**
  - Est entièrement focalisée sur le taux d'erreur
  - Présuppose un environnement stationnaire et des entrées/requêtes (i.i.d.)
  - Exige un **nombre de données d'apprentissage assez grand** par rapport à la **capacité de** $\mathcal{H}$
- Nous ne comprenons pas bien les réseaux de neurones profonds
- Corrélations $\neq$ structures, sémantique, causalité

Limites

- Apprentissage **passif** et **données et questions i.i.d.**
  - Agents situés : le monde n'est pas i.i.d.
- Requiert **beaucoup** d'exemples
  - Nous sommes beaucoup plus efficaces
  - « Producteurs de théories », théories que nous testons ensuite
- Pas adapté à la recherche de **causalités**
- Pas **intégré** avec un **raisonnement**

Ces machines **apprenantes** ne sont pas des machines **pensantes**

Mes paris pour l'avenir

Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**

2. Apprendre à partir de **multiples** **sources de données** hétérogènes

3. Apprendre par **analogie** et par **transfert**

4. Apprendre pour **construire** **des théories ?** (causalité et explications)

5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

Mes paris sur les directions à venir

1. Apprendre à partir de **multiples** **sources de données** hétérogènes

2. Apprendre à partir de **très peu d'exemples**

3. Apprendre par **analogie** et par **transfert**

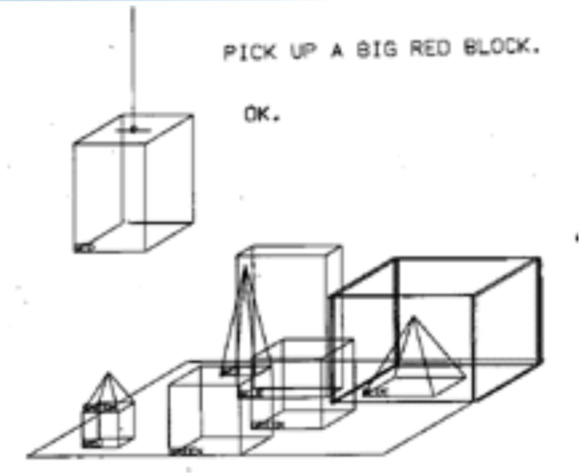
4. Apprendre pour **construire des théories ?** (causalité et explications)

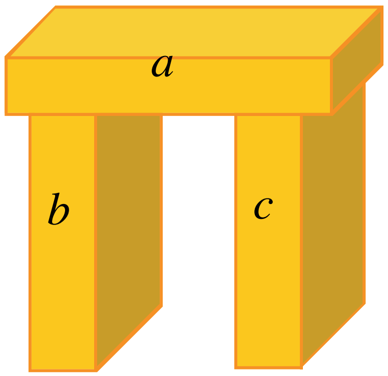
5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

ARCH [Winston, 1970]

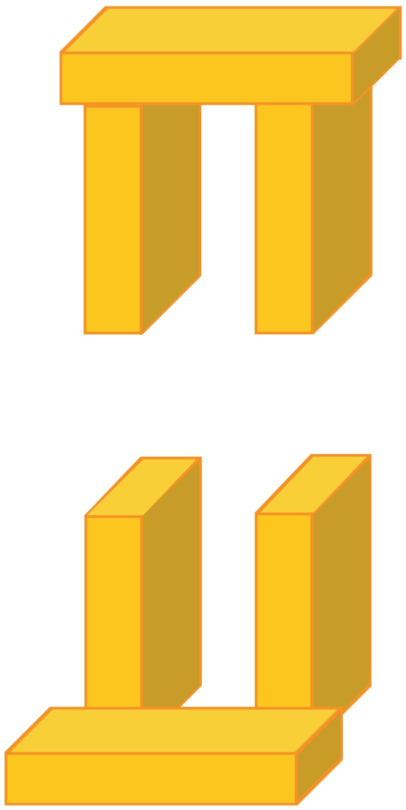
- **Apprentissage** de **concept** (e.g. arche) dans un monde de blocs



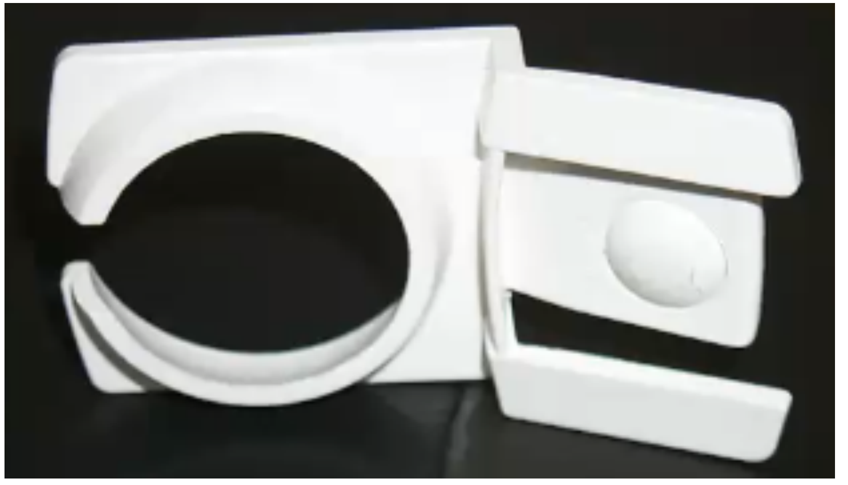


ARCH [Winston, 1970]

- Les exemples ne sont **pas choisis au hasard**

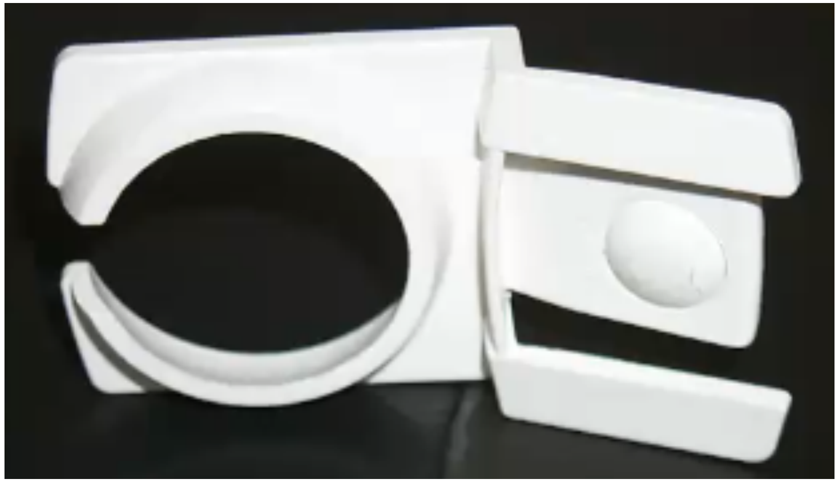


Apprentissage à partir d'un seul exemple



…

Apprentissage à partir d'un seul exemple



A child learns about four+ new words a day

Goulden, R., Nation, P. & Read, J. (1990). How large can a receptive vocabulary be? *Applied linguistics*, *11* (4), 341-363.

- Conférence de Jean-Louis Dessalles du 5/11/2020

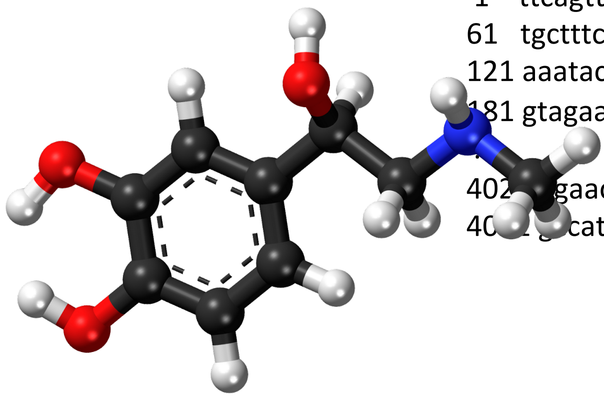
Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**
2. Apprendre à partir de **multiples** **sources de données** hétérogènes
3. Apprendre par **analogie** et par **transfert**
4. Apprendre pour **construire des théories ?** (causalité et explications)
5. L'**intégration** de multiples systèmes apprenants
6. Le « **teaching** data science »

Intégration de multiple sources de données

- Annotation de protéines

Protéine « sp|P00004|CYC_HORSE » is activated by …

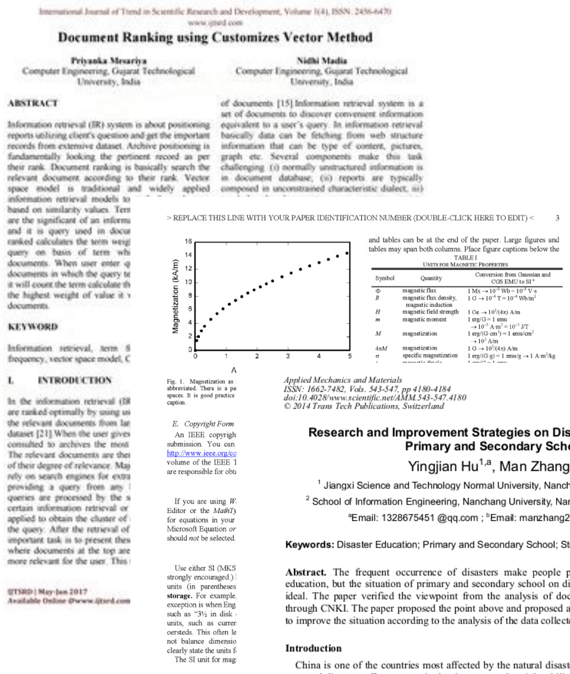


1 ttcagttgtg aatgaatgga cgtgccaaat agacgtgccg ccgccgctcg attcgcactt
61 tgctttcggt tttgccgtcg tttcacgcgt ttagttccgt tcggttcatt cccagttctt
121 aaataccgga cgtaaaaata cactctaacg gtcccgcgaa gaaaaagata aagacatctc
181 gtagaaatat taaaataaat tcctaaagtc gttggtttct cgttcacttt cgctgcctgc

402[illegible] gaacacgcc gaggctccat tcatagcacc acttcgtcgt cttaatcccc tccctcatcc
40[illegible] catggcgg tgcaaaaaat aaaaagaact c

Intégration de multiple sources de données

- GIEC
  - Documents scientifiques multiples
  - Tableaux
  - mesures


International Journal of Trend in Scientific Research and Development, Volume 1(4), ISSN: 2456-6470
www.ijtsrd.com

Document Ranking using Customizes Vector Method

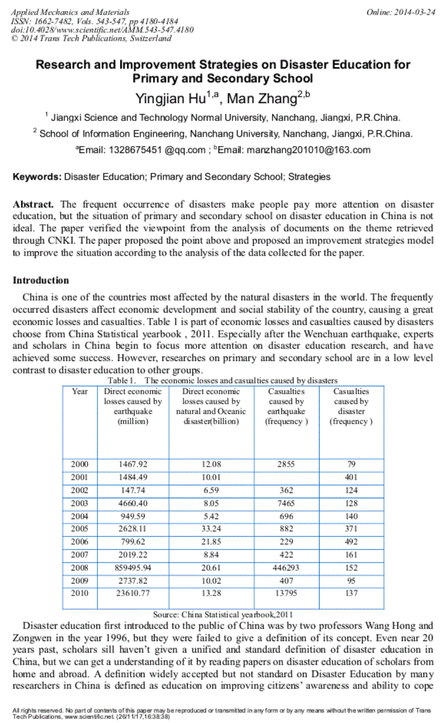
Priyanka Mesariya, Computer Engineering, Gujarat Technological University, India
Nidhi Madia, Computer Engineering, Gujarat Technological University, India

ABSTRACT

KEYWORD

I. INTRODUCTION

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <


Applied Mechanics and Materials *Online: 2014-03-24*
ISSN: 1662-7482, Vols. 543-547, pp 4180-4184
doi:10.4028/www.scientific.net/AMM.543-547.4180
© 2014 Trans Tech Publications, Switzerland

Research and Improvement Strategies on Disaster Education for Primary and Secondary School

Yingjian Hu[1,a], Man Zhang[2,b]

[1] Jiangxi Science and Technology Normal University, Nanchang, Jiangxi, P.R.China.
[2] School of Information Engineering, Nanchang University, Nanchang, Jiangxi, P.R.China.
[a]Email: 1328675451 @qq.com ; [b]Email: manzhang201010@163.com

Keywords: Disaster Education; Primary and Secondary School; Strategies

Abstract. The frequent occurrence of disasters make people pay more attention on disaster education, but the situation of primary and secondary school on disaster education in China is not ideal. The paper verified the viewpoint from the analysis of documents on the theme retrieved through CNKI. The paper proposed the point above and proposed an improvement strategies model to improve the situation according to the analysis of the data collected for the paper.

Introduction

China is one of the countries most affected by the natural disasters in the world. The frequently occurred disasters affect economic development and social stability of the country, causing a great economic losses and casualties. Table 1 is part of economic losses and casualties caused by disasters choose from China Statistical yearbook , 2011. Especially after the Wenchuan earthquake, experts and scholars in China begin to focus more attention on disaster education research, and have achieved some success. However, researches on primary and secondary school are in a low level contrast to disaster education to other groups.

Table 1. The economic losses and casualties caused by disasters

Year	Direct economic losses caused by earthquake (million)	Direct economic losses caused by natural and Oceanic disaster(billion)	Casualties caused by earthquake (frequency)	Casualties caused by disaster (frequency)
2000	1467.92	12.08	2855	79
2001	1484.49	10.01		401
2002	147.74	6.59	362	124
2003	4660.40	8.05	7465	128
2004	949.59	5.42	696	140
2005	2628.11	33.24	882	371
2006	799.62	21.85	229	492
2007	2019.22	8.84	422	161
2008	859495.94	20.61	446293	152
2009	2737.82	10.02	407	95
2010	23610.77	13.28	13795	137

Source: China Statistical yearbook,2011

Disaster education first introduced to the public of China was by two professors Wang Hong and Zongwen in the year 1996, but they were failed to give a definition of its concept. Even near 20 years past, scholars sill haven't given a unified and standard definition of disaster education in China, but we can get a understanding of it by reading papers on disaster education of scholars from home and abroad. A definition widely accepted but not standard on Disaster Education by many researchers in China is defined as education on improving citizens' awareness and ability to cope

All rights reserved. No part of contents of this paper may be reproduced or transmitted in any form or by any means without the written permission of Trans Tech Publications, www.scientific.net. (26/11/17,16:38:38)

Moore's Law has, for nigh half a century, reliably predicted the growth in efficiency of processors: Moore's Law states that the number of transistors that can be placed on a given surface area doubles every two years [Intel Corporation, 2005]. As a consequence, the number of transistors – and consequently, the computing power – of processors has grown exponentially until recently. However, this growth can no longer be sustained due to a combination of several factors. The most important cause are quantum mechanical effects which raise the electrical resistance of the transistors and thus cause heat dissipation problems which result in energy loss [Feynman, 1985; Tanenbaum, 1990].

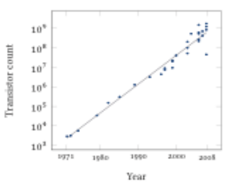


Figure 1: Moore's Law illustrated by the number of transistors of typical processors for each era. Note that the y axis is logarithmic.

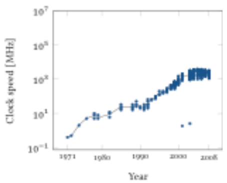


Figure 2: Clock speed (in MHz) of Intel processors over the years and their mean values for each year.

On the other hand, we're dealing with ever increasing amounts of data that our programs have to process. Figure 3 illustrates this using the example of the number of

	MaxEnt			MaxEnt + GE			Unsup GE		
	P	R	F	P	R	F	P	R	F
BKG	.38	.19	.25	.49	.48	**.48**	.49	.44	.46
PROB	0	0	0	.38	.23	.29	.28	.38	**.32**
METH	0	0	0	.29	.50	**.37**	.08	.56	.14
RES	0	0	0	.68	.51	**.58**	.08	.51	.14
CON	.69	.96	.80	.81	.84	**.82**	.74	.69	.71
CN	.35	.06	.10	.39	.29	**.33**	.40	.13	.20
DIFF	0	0	0	.21	.30	**.25**	.12	.13	.12
FUT	0	0	0	.24	.44	.31	.26	.61	**.36**

Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**

2. Apprendre à partir de **multiples sources de données** hétérogènes

3. Apprendre par **analogie** et par **transfert**

4. Apprendre pour **construire des théories ?** (causalité et explications)

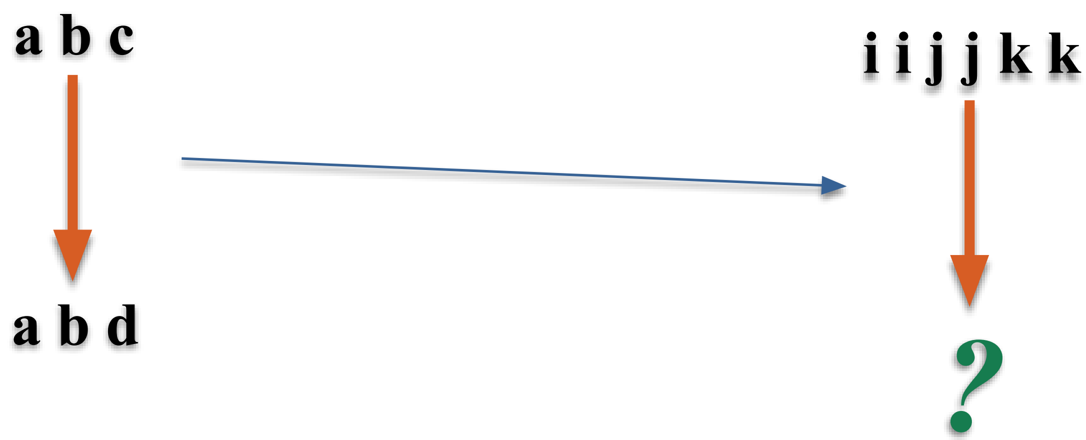
5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

Que savons-nous de l'apprentissage

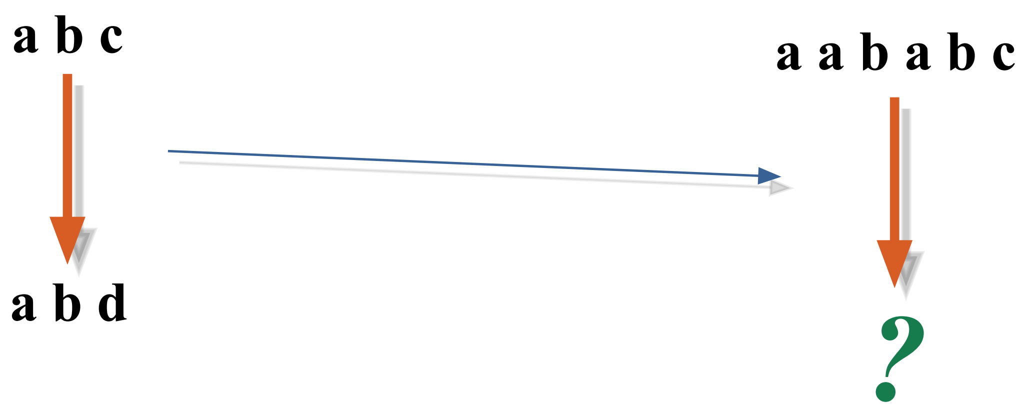
en environnement non stationnaire ?

Transfert et analogie



- a b d
- i i j j k d
- i i j j k l
- i i j j k k
- ?

Transfer and analogy



Why should 'a a b a b c d' be any better than 'a b d'?

Transfer learning

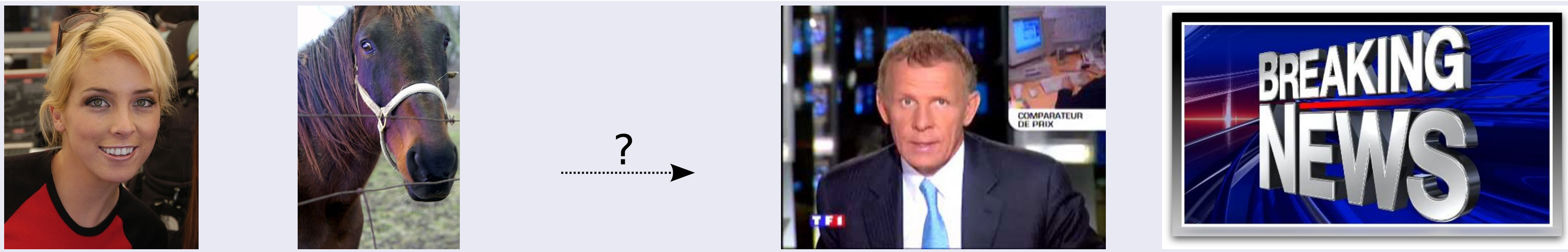
Definition [Pan, TL-IJCAI'13 tutorial]

– Ability of a system to **recognize** and **apply** knowledge and skills learned in **previous domains/tasks** to **novel domains/tasks**

Example

– We have **labeled images** (person / no person) from a **web corpus**

– Novel task: **is there a person** in unlabeled images from a **video corpus**?



Person   no Person   Is there a Person?

Web corpus   Video corpus

Transfert learning: questions

- What can be **the basis** of transfer learning?

  How to translate formally :

  *"the target domain **is like** the source domain"*?

Not i.i.d. anymore

- What **determine a good transfer**?
  - A "good source"?
  - A high "similarity" between source and target?

- What **formal guarantees** can we have on the transferred hypothesis?

Apprentissage en environnement **non** stationnaire

- La distribution en **utilisation** n'est pas la même qu'en apprentissage
  - L'échantillon d'apprentissage n'est **pas représentatif**

E.g.:

- Apprendre à discriminer des évènements rares
- Apprentissage actif
- Environnement changeant

→ La **théorie statistique** de l'apprentissage ne fonctionne plus

- Les garanties théoriques sont trop éloignées de l'usage

Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**

2. Apprendre à partir de **multiples sources de données** hétérogènes

3. Apprendre par **analogie** et par **transfert**

4. Apprendre pour **construire des théories ?** (causalité et explications)

5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

We start to pay attention to new demands

1. The need for explanations
   - Structures
   - **Causal** reasoning
   - No more only error rate

Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**

2. Apprendre à partir de **multiples sources de données** hétérogènes

3. Apprendre par **analogie** et par **transfert**

4. Apprendre pour **construire des théories ?** (causalité et explications)

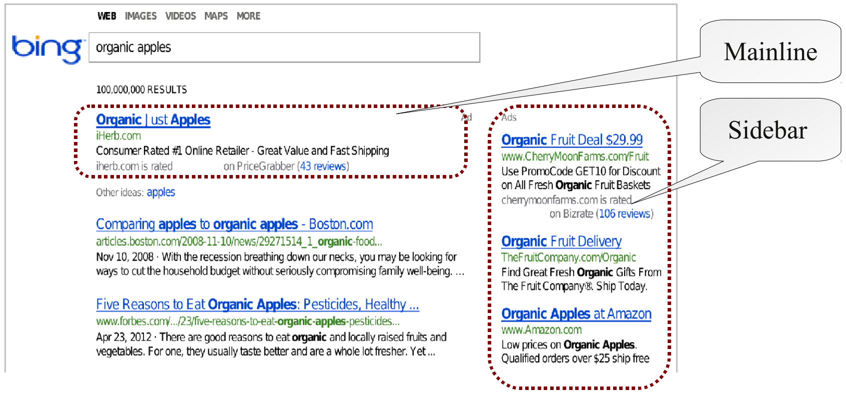
5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

Interactions between learning modules

Adaptive advertising placement system

- Two **sub-systems**
  - One placing **advertising links**
  - The other one choosing the **adds**



- **Mutually influencing** each other
  - Each one is based on click data
  - Which also **depends on the intervention of the other system**
  - And other **uncontrolled factors** (price, user requests, …)

[L. Bottou et al. «*Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising* », JMLR, 14, (2013), 3207-3260]

Mes paris sur les directions à venir

1. Apprendre à partir de **très peu d'exemples**

2. Apprendre à partir de **multiples sources de données** hétérogènes

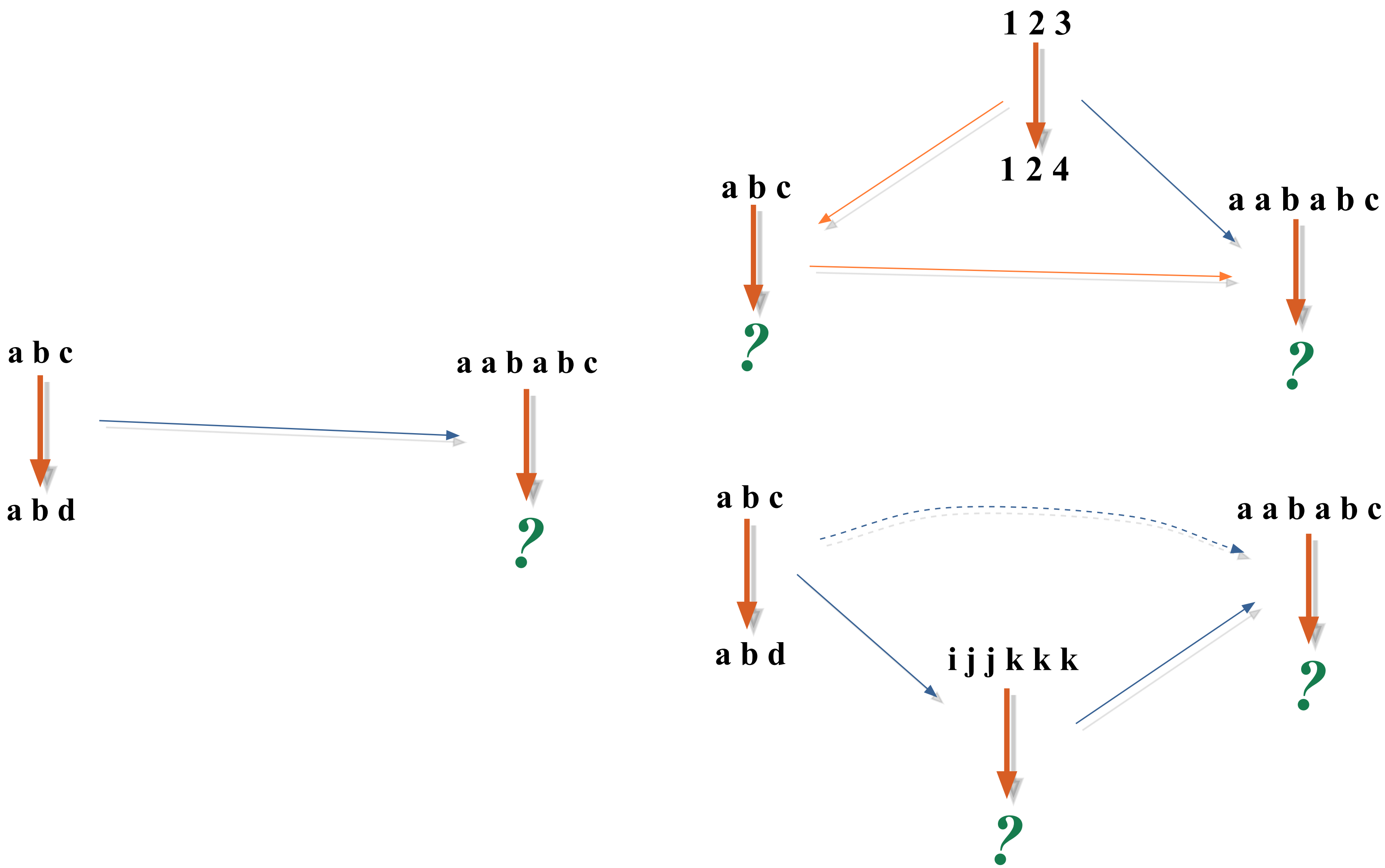
3. Apprendre par **analogie** et par **transfert**

4. Apprendre pour **construire des théories ?** (causalité et explications)

5. L'**intégration** de multiples systèmes apprenants

6. Le « **teaching** data science »

Transfer and sequence effects



- t

Long-life learning

- Learning organized in **a sequence of tasks**
  - Very **far from the i.i.d. scenario**

→ Learning will be affected by the **history of the system**

- We need a theory of the **dynamics of learning**
  1. Which **sequence effects** can we expect?
  2. How to **best organize the curriculum** of a learning system?

Un pari

Aller vers des systèmes capables d'enseigner

1. **Expliquer** un cas

2. **Synthétiser**

3. Organiser un **curriculum**

➢ Vers une **évaluation** des systèmes **par la performance de leurs élèves** ?

Conclusions: “new” scenarios

- **Limited** data sources
  - We often learn from (very) few examples

- The past **history** of learning affects learning: Education
  - Sequence effects

- We learn in order to **build “theories”**
  - All the time: small and large theories

For instance, what would you like to ask?

Suppléments

En pratique

En pratique

1. Obtenir les données
2. Importance des **prétraitements**
3. Importance de la disponibilité des **experts métier**
4. Bien penser le **recueil des données**
5. Les questions **juridiques**

Obtenir les données

Souvent **difficile** !!!

- Les données ne sont **pas encore disponibles**
- Le donneur d'ordre n'est **pas détenteur des données**
  - Pas le même service / département
- Les données sont **protégées par des droits**
- Une partie des données **reste à recueillir**

Les prétraitements

- **90%** du temps d'un projet
- **Recueil** des données
- Mise dans un **format adéquat**
- **Nettoyage**
  - **Bruit** dans les données
  - Données **manquantes**
  - Données **aberrantes**
  - **Doublons**
  - **Normalisation** des mesures
  - **Discrétisation** de valeurs continues
  - **Rendre continues** des valeurs discrètes
- Élimination des **attributs rougeondants** / calcul de **nouveaux attributs**
- **Précision / incertitude**
- Intégration de plusieurs **sources de données (hétérogènes)**
- …

Choix d'un bon critère de performance

Disponibilité des experts métier

Essentiel !!!

- **Comprendre** le problème
- Établir un **vocabulaire commun**
- **Évaluer** les résultats
- Orienter / **ré-orienter**
- **S'approprier** les résultats / assurer la suite

Bien penser le recueil des données

Essentiel !!!

- Exemple : **I**nternet **d**es **O**bjets (IoT)
  - Objets **faciles** et **agréables** à utiliser
  - **Mais**
    - Ne recueille pas les données nécessaires
    - Développement « agile »
      - ✓ Changements de formats
      - ✓ Changements des mesures recueillies

2 ans de perdus

Tout reprendre à zéro

Les questions juridiques

Essentiel !!!

- Données **personnelles**
- **Obtenir l'autorisation**
  - CNIL
  - RGPD
    - À partir du **25 mai 2018**, le Règlement Général Européen sur la Protection des Données (RGPD) affectera toutes les organisations traitant les données personnelles identifiables (DPI) de résidents européens.