# Plan

1. Grands types d'apprentissage
2. Apprentissage prédictif par réseaux de neurones
3. Quelles garanties ?
4. Recette pour créer des algorithmes d'apprentissage
5. Les réseaux de neurones profonds
6. Ce que l'on sait faire et les défis à relever

Un peu de recul :

**Que sait-on** faire

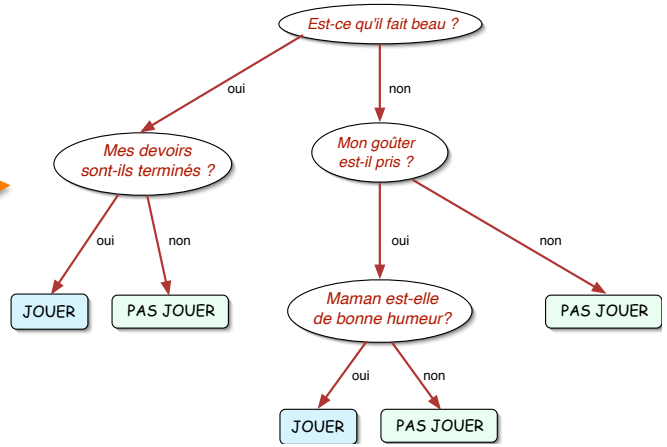
et **où sont** les **limites** ?

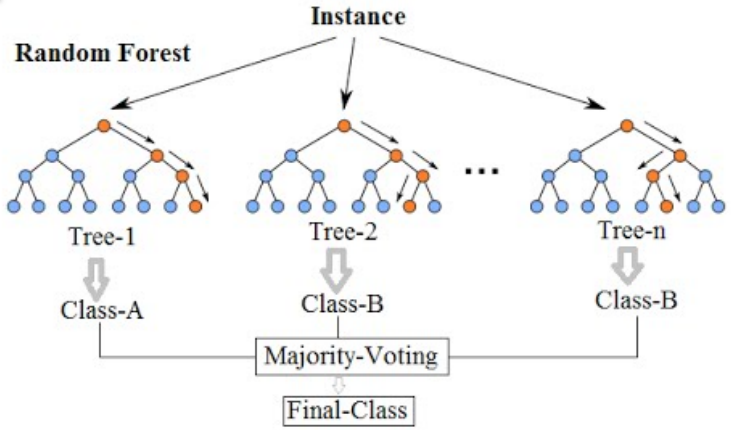
# Ce que l'**on sait faire**

- 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
- …



# Les méthodes et algorithmes

- Librairies / méthodes / algorithmes
  - Sont dans le **domaine public** !!!
    - Publications scientifiques
    - Forums
    - Conférences
    - Librairies (e.g. ScikitLearn)

- Des « **recettes** » privées
  - Réseaux de neurones profonds
  - Traitement d'images / télédétection
  - Connaissances métiers (e.g. alimentation)

# Les moyens calcul

- Important
- Mais **pas très honéreux**
  - **Station de travail** avec 8 cartes graphiques et 128 Go de mémoire centrale
  - **Cluster** de machines
  - **Utilisation de cloud computing**
- Mais … **évolue vite**

Et dépend de ce que l'on veut faire

# Les « data scientists »

- **Compétences** attendues
  1. Apprentissage artificiel / Statistiques
     - Bonne compréhension des questions et des hypothèses sur lesquelles reposent les méthodes
  2. Compétences en informatique
     - Algorithmique
     - Bases de données
     - Réseaux
  3. Capacités relationnelles

**En très forte demande**

100 000 en France
à l'horizon 2022 !!

- **Formations**
  - Quelques dizaines d'heures
  - **Master** ou équivalent
  - **Doctorat**

**Grand risque** de déconvenue
si pas les bons recrutements

Sait-on finalement **expliquer**

les capacités de généralisation ?

# Quelque chose de troublant

– C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals (ICLR, **May 2017**).

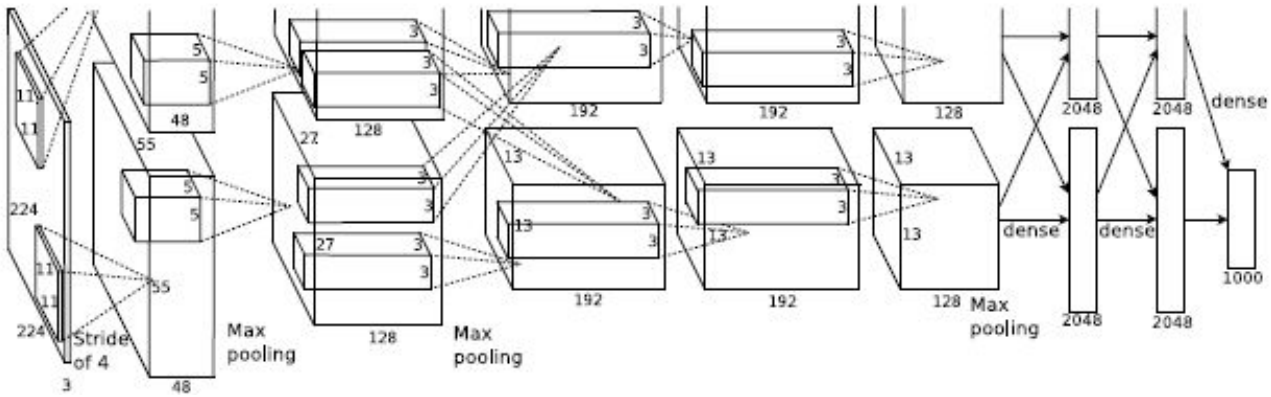
"Understanding deep learning requires rethinking generalization"

# 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**:

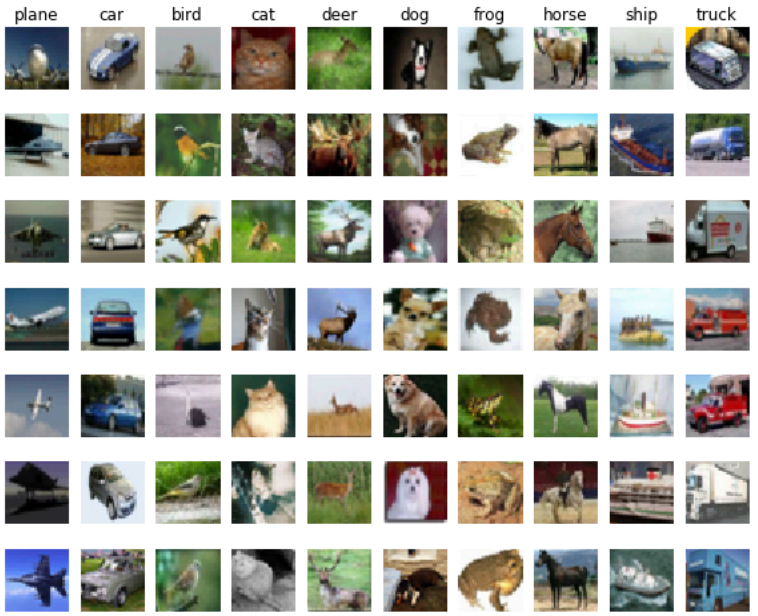
- **60,000** images categorized in **10 classes** (50,000 for training and 10,000 for testing)
- Images: 32x32 pixels in 3 color channels

# 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

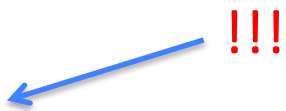
$$\forall h \in \mathcal{H}, \forall \delta \leq 1 : \quad P^m \left[ R(h) \leq \widehat{R}(h) + 2\,\widehat{Rad}_m(\mathcal{H}) + 3\sqrt{\frac{\ln(2/\delta)}{m}} \right] > 1 - \delta$$

# 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

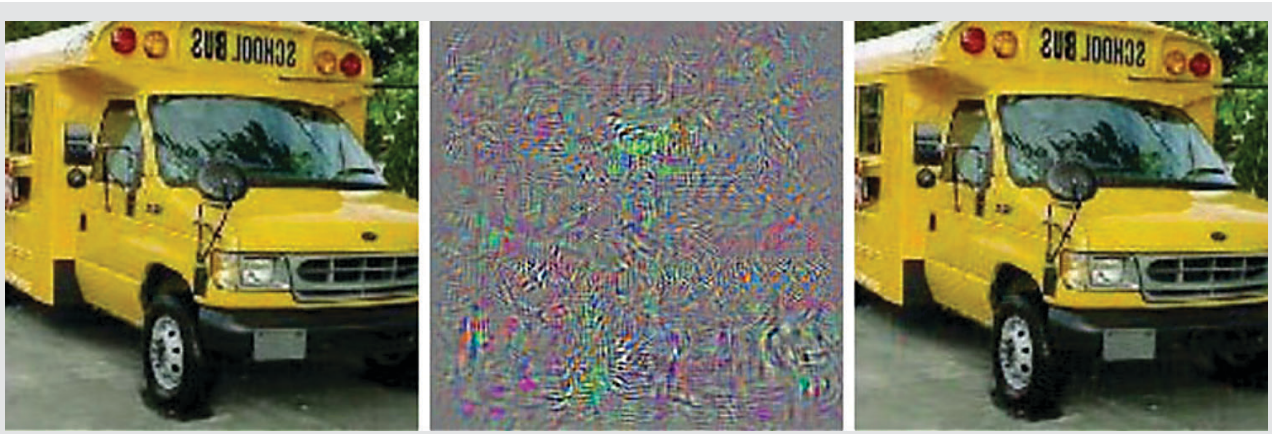
Can grow without limit!!

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

But then,

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

# Adversarial learning




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

# Explication et réseaux de neurones profonds

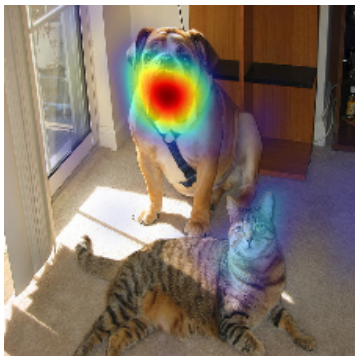
**Illusions d'optique** : quelle explication ?

Boxer: 0.40 Tiger Cat: 0.18

(a) Original image

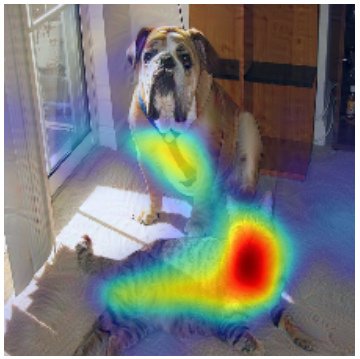
Airliner: 0.9999

(b) Adversarial image



Boxer: 1.1e-20

(c) Grad-CAM "Dog"



Tiger Cat: 6.5e-17

(d) Grad-CAM "Cat"

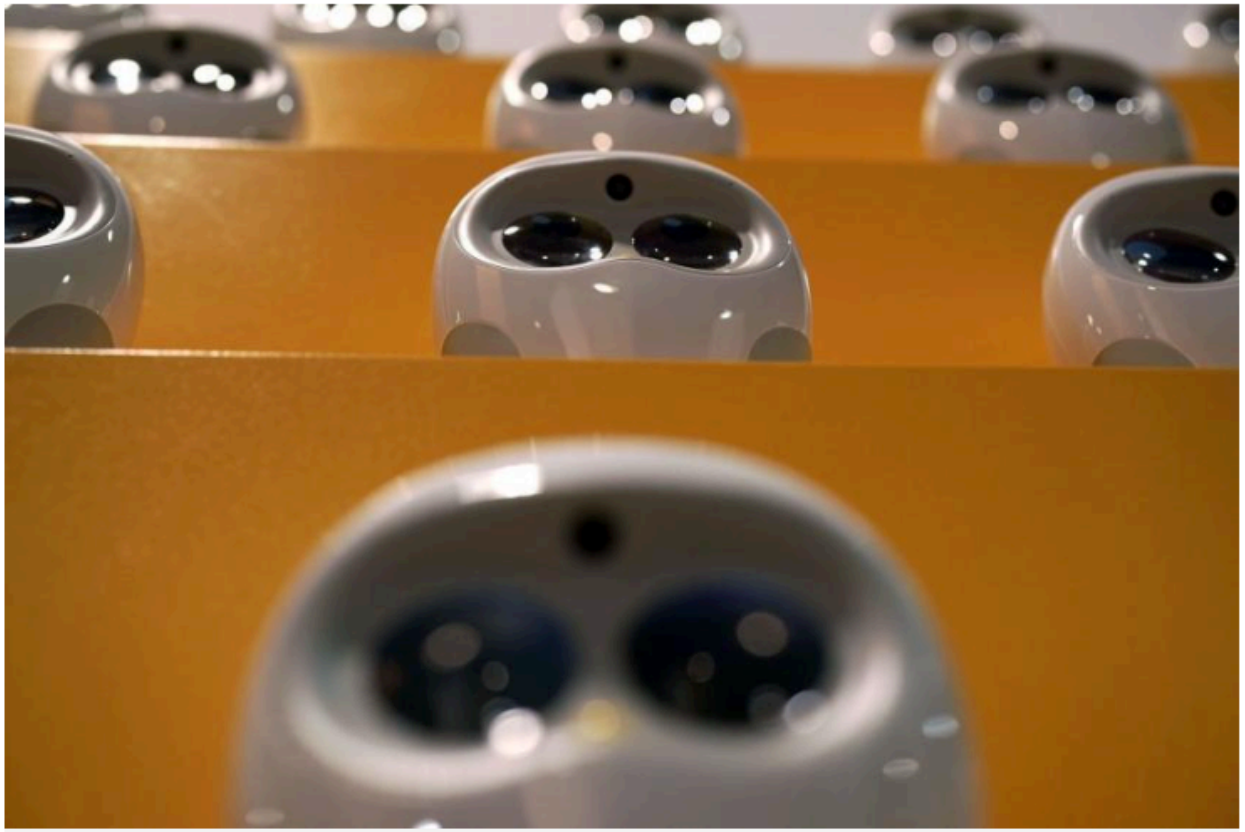
!!??

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

TECH & SCIENCE

# ROBOTS CAN NOW READ BETTER THAN HUMANS, PUTTING MILLIONS OF JOBS AT RISK

BY ANTHONY CUTHBERTSON ON 1/15/18 AT 8:00 AM

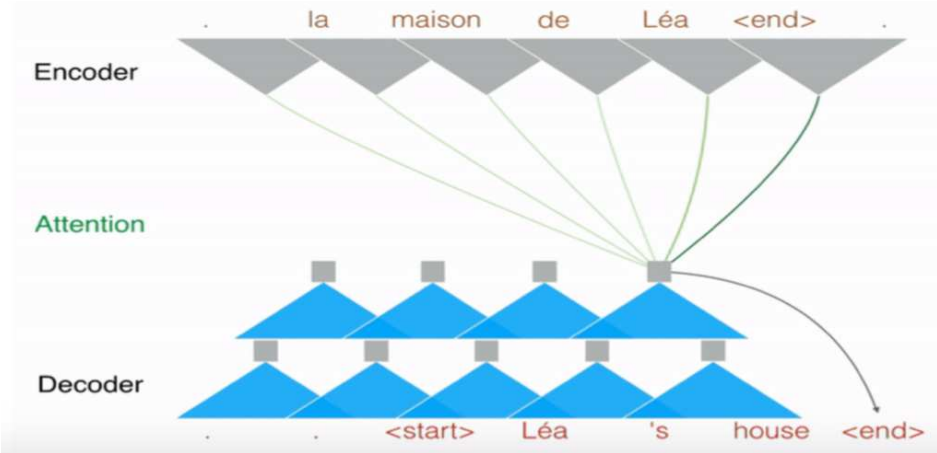


https://www.newsweek.com/robots-can-now-read-better-humans-putting-millions-jobs-risk-781393

- ex

# Machine translation

- Still far from perfect, but …




From Hofstädter (2018)

# Reading comprehension

**Paragraph**: Peyton Manning became the first quarter-back ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager.

**Question**: What is the name of the quarterback who was 38 in Super Bowl XXXIII?

- ex

# Reading comprehension

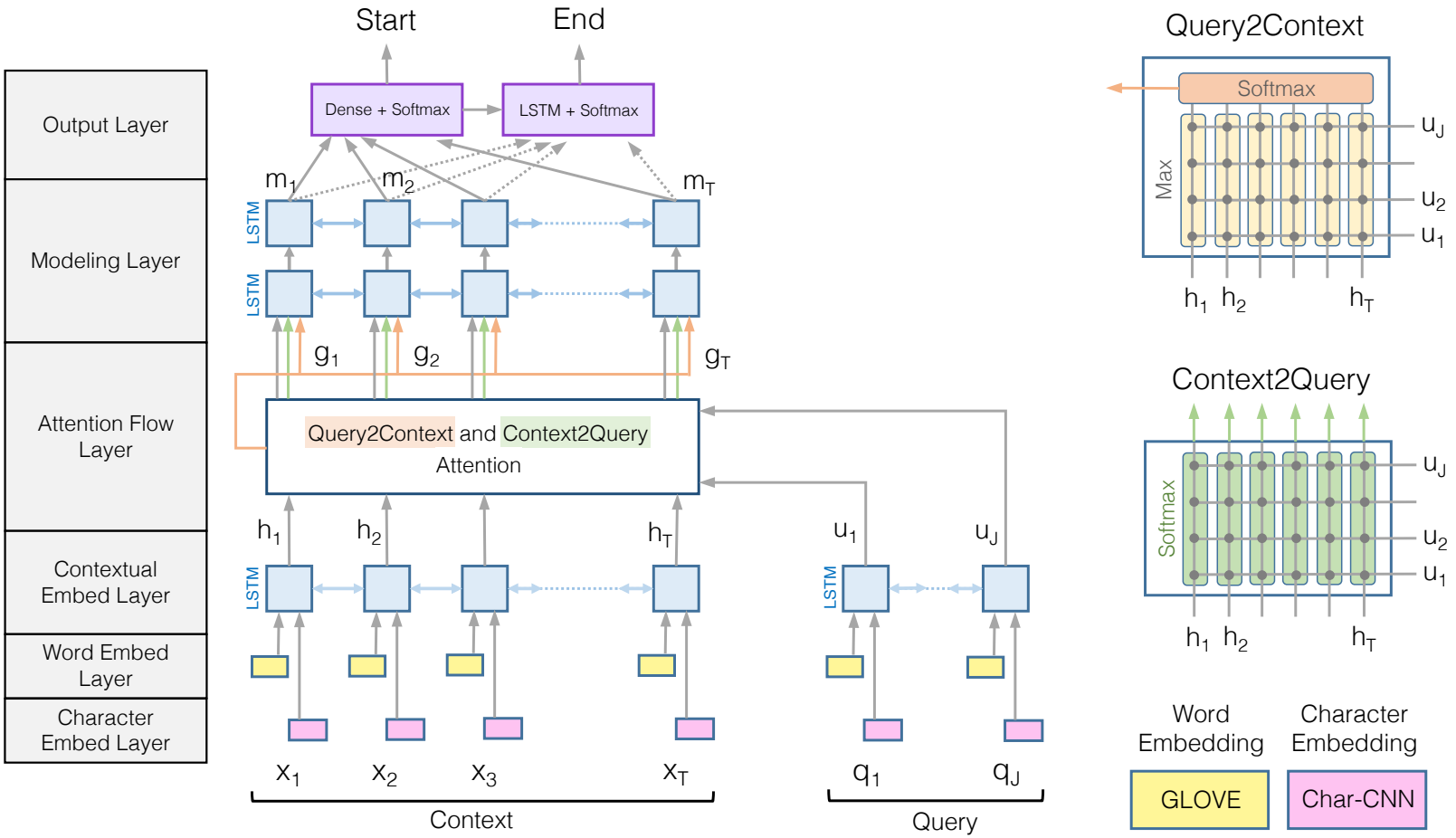
**Paragraph**: Peyton Manning became the first quarter-back ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by **John Elway**, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager.

**Question**: What is the name of the quarterback who was 38 in Super Bowl XXXIII?

**Answer**: John Elway

- ex

# Reading comprehension




Minjoon Seo et al. Bi-directional attention flow for machine comprehension. ICLR 2017

# Reading comprehension

## SQuAD1.1 Leaderboard

Since the release of SQuAD1.0, the community has made rapid progress, with the best models now rivaling human performance on the task. Here are the ExactMatch (EM) and F1 scores evaluated on the test set of v1.1.

| Rank | Model | EM | F1 |
|---|---|---|---|
| | Human Performance<br>*Stanford University*<br>(Rajpurkar et al. '16) | 82.304 | 91.221 |
| 1<br>Sep 09, 2018 | nlnet (ensemble)<br>*Microsoft Research Asia* | **85.356** | **91.202** |
| 2<br>Aug 28, 2018 | nlnet (ensemble)<br>*Microsoft Research Asia* | 85.104 | 91.055 |
| 3<br>Jul 08, 2018 | r-net (ensemble)<br>*Microsoft Research Asia* | 84.003 | 90.147 |

# Reading comprehension

**Paragraph**: **Peyton Manning** became the first quarter-back ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by **John Elway**, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager.

**Question**: **What is the name of** the quarterback who was 38 in Super Bowl XXXIII?

- ex

# Reading comprehension

**Paragraph**: **Peyton Manning** became the first quarter-back ever to lead two different teams to multiple Super Bowls. He is also the oldest **quarterback** ever to play in a Super Bowl at age 39. The past record was held by **John Elway**, who led the Broncos to victory in **Super Bowl XXXIII** at age **38** and is currently Denver's Executive Vice President of Football Operations and General Manager.

**Question**: **What is the name** of the **quarterback** who was **38** in **Super Bowl XXXIII**?

- ex

# Reading comprehension

**Paragraph**: Peyton Manning became the first quarter-back ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. **Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV**.

**Question**: What is the name of the quarterback who was 38 in Super Bowl XXXIII?

**Answer**: Jeff Dean 

- ex

# Sentiment classification

There is really but one thing to say about this sorry movie It should never have been made The first one one of my favourites An American Werewolf in London is a great movie with a good plot good actors and good FX But this one It stinks to heaven with a cry of helplessness

➜ negative

- ex

# Sentiment classification

There is really but one thing to say about **this** sorry movie It should never have been made The first one one of my favourites An American Werewolf in London is a great movie with a good plot good actors and good FX But this one It stinks to heaven with a cry of helplessness

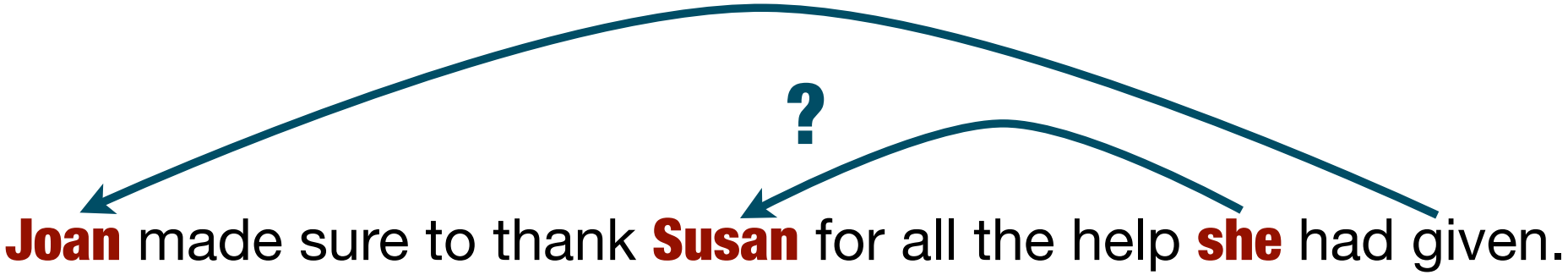
➜ negative

There is really but one thing to say about **that** sorry movie It should never have been made The first one one of my favourites An American Werewolf in London is a great movie with a good plot good actors and good FX But this one It stinks to heaven with a cry of helplessness

➜ positive ✗

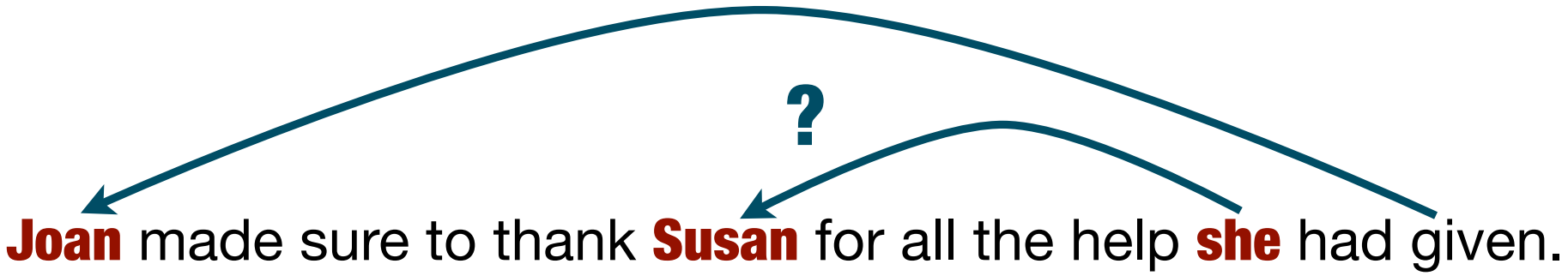
- ex

# Winograd schema challenge




- ex

# Winograd schema challenge



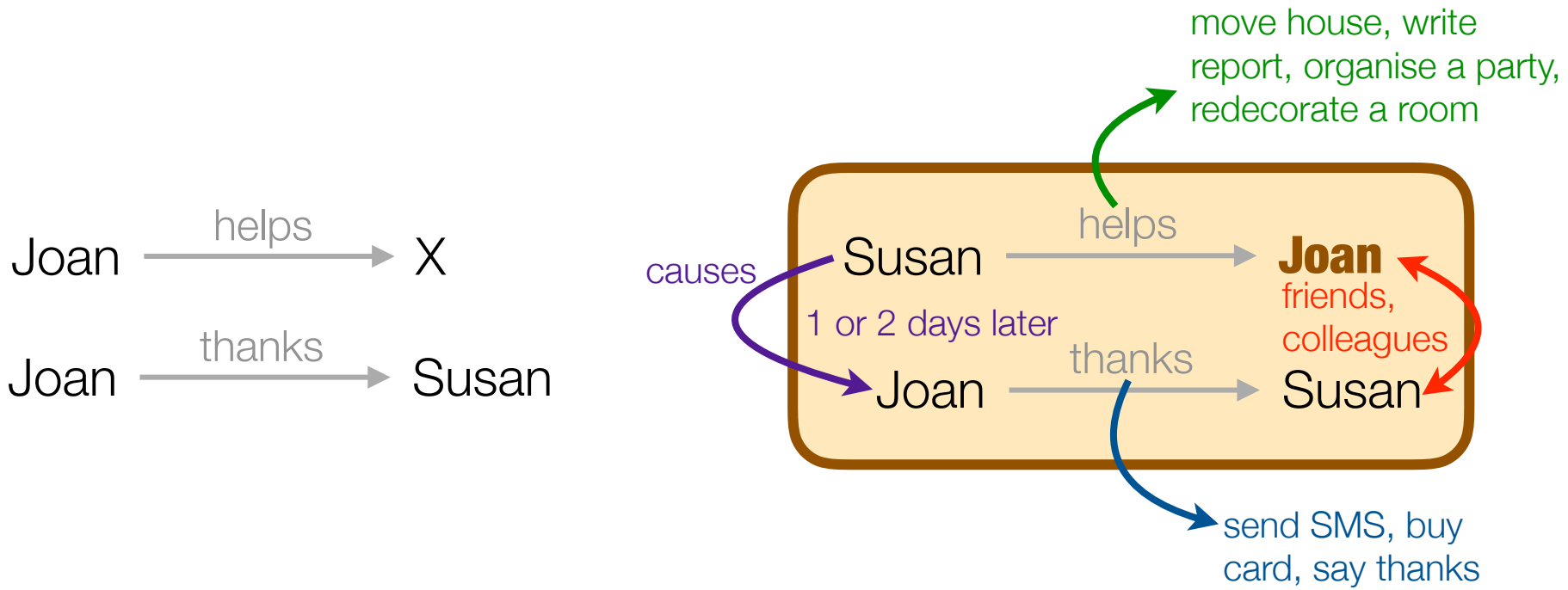

Random guessing: 50% correct

Deep neural network (DNN): 50.6% correct

- ex

# Winograd schema challenge

**Joan** made sure to thank **Susan** for all the help **she** had given.




**Pattern matching + commonsense knowledge**

- ex

# Sait-on expliquer une conclusion ?

# **Voiture** dans une piscine

- … ou pas de voiture … ?

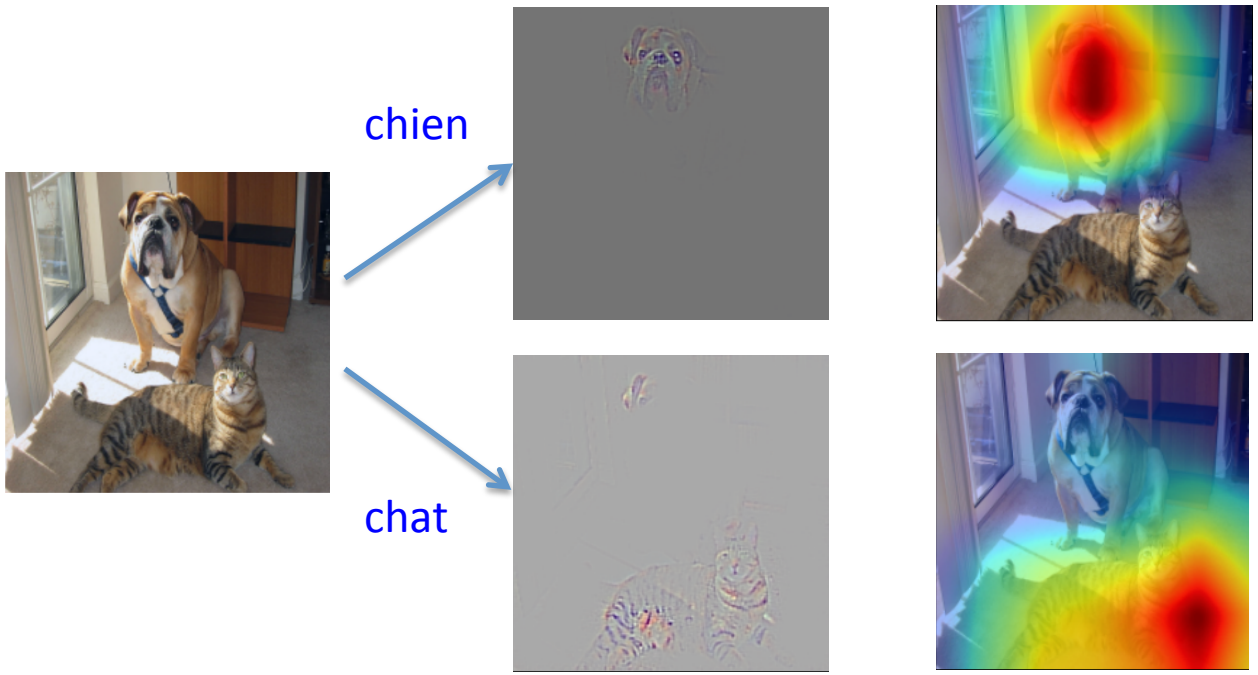

Is this less of a car
because the context is wrong?

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

# 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* »]

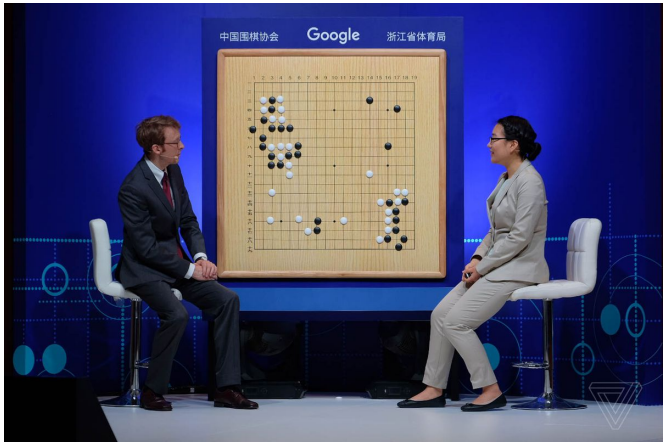
# Le cas AlphaGo

- 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

# Les explications des SEs

**MYCIN :** L'infection du malade a t-elle été contractée à l'hôpital ?

**Utilisateur :** pourquoi ?

**TEIRESIAS :** [pourquoi est-ce important de savoir cela ?]

Cela aidera à déterminer la catégorie à laquelle appartient l'organisme responsable de l'infection :

- il a déjà été établi que 1° l'organisme est à Gram négatif, 2° c'est un bâtonnet, 3° il se développe à l'air ou sans air,

- donc si l'infection par cet organisme n'a pas été contactée lors du séjour à l'hôpital

alors il est presque certain (0,8) que l'organisme est de la catégorie des enterobactoriaceae; de plus, il se peut (0,1 ; règle 37) que ce soit la pseudomonia aeruginosa.

**Utilisateur :** pourquoi n'avez-vous pas envisagé la possibilité que l'organisme soit un streptocoque ?

**TEIRESIAS :** la règle 33 aurait du être utilisée pour identifier l'organisme comme un streptocoque, mais la seconde clause "la morphologie est de type coque" était déjà connue comme fausse et la règle 33 n'a jamais été essayée.

# Les explications des SEs

- Quel est l'âge du patient ?
- ** Pourquoi
- Pour déterminer si on peut prescrire de la tétracycline
  - Si le patient a moins de 8 ans
  - Alors on ne peut pas prescrire de tétracycline
  - [Règle 122]
- ** Pourquoi ?
- …

Pourquoi ne faut-il pas prescrire de tétracycline à un enfant de moins de 8 ans ?

# Les explications des SEs

*Pourquoi ne faut-il pas prescrire de tétracycline à un enfant de moins de 8 ans* ?

## Connaissances justificatives

Dépôt de la drogue sur les **os en développement**

→ **Noircissement** définitif des dents

→ Coloration socialement **indésirable**

→ **Ne pas administrer** de tétracycline aux enfants de moins de 8 ans

Notion d'**effets secondaires** indésirables

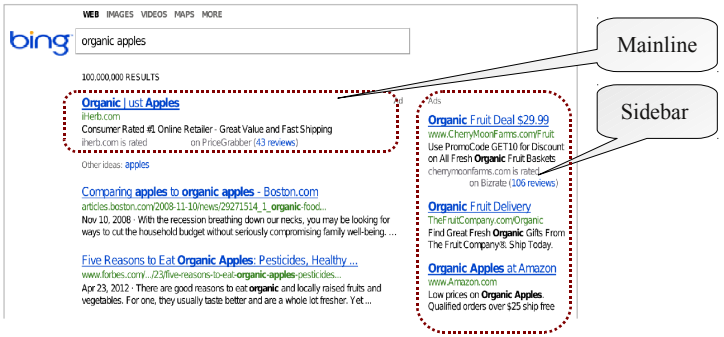
Relations de **causalité**

Possédons nous les bases

d'un génie logiciel des systèmes apprenants ?

# Système adaptatif de placement de publicité

- Deux **sous-systèmes**
  - L'un plaçant les **liens publicitaires**
  - L'autre choisissant les **publicités**

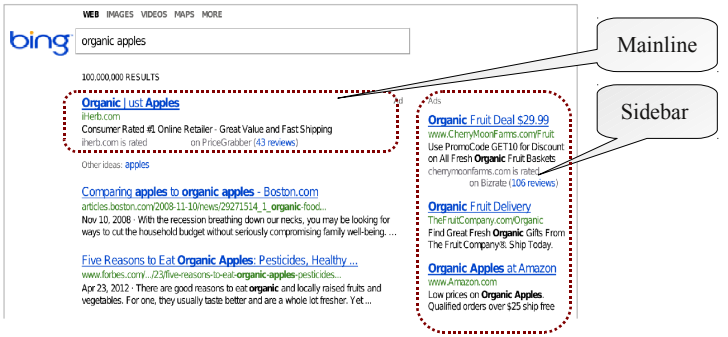



- Qui **s'influencent mutuellement**
  - Chacun s'appuie sur les données de clicks
  - Qui dépendent aussi de l'intervention de l'autre systèmes
  - Et d'autres facteurs non contrôlés (prix, requête de l'utilisateur, …)

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

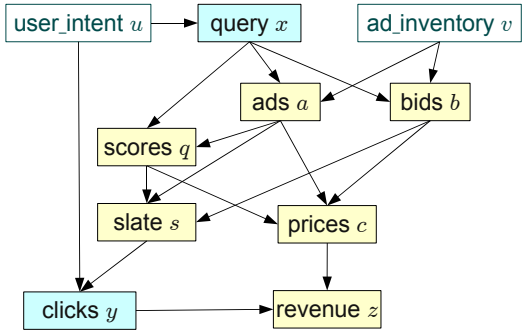
# Système adaptatif de placement de publicité

- Deux **sous-systèmes**
  - L'un plaçant les **liens publicitaires**
  - L'autre choisissant les **publicités**




- Qui **s'influencent mutuellement**
  - Chacun s'appuie sur les données de clicks
  - Qui dépendent aussi de l'intervention de l'autre systèmes
  - Et d'autres facteurs non contrôlés (prix, requête de l'utilisateur, …)

Importance de l'identification
du **graphe causal**




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

Que savons-nous de l'apprentissage

à partir de (très) peu d'exemples ?

# Les passages à l'échelle

## 1. Savoir traiter de (très) **gros** **volumes de données**

- Méthodes efficaces
  - Gradient stochastique
  - Apprentissage convexe
  - Optimisation du code
    - ✓ Accès mémoire
    - ✓ Complexité computationnelle
- Distribution des calculs
  - Cartes graphiques / cœurs
  - Clusters de machines
  - Cloud computing
    - ✓ Approches Map rougeuce

# Les passages à l'échelle

## 2. Savoir traiter de (très) **petits** **volumes de données**

**Compenser** le manque d'information dans les données

- Par de la **connaissance experte**
- **Enrichissement** des données
  - Ontologies
  - Web sémantique
  - Wikipedia and Co
- Question de la **validation des résultats**
  - Les experts

# Que savons-nous de l'apprentissage

# en environnement non stationnaire ?

# Transfer learning

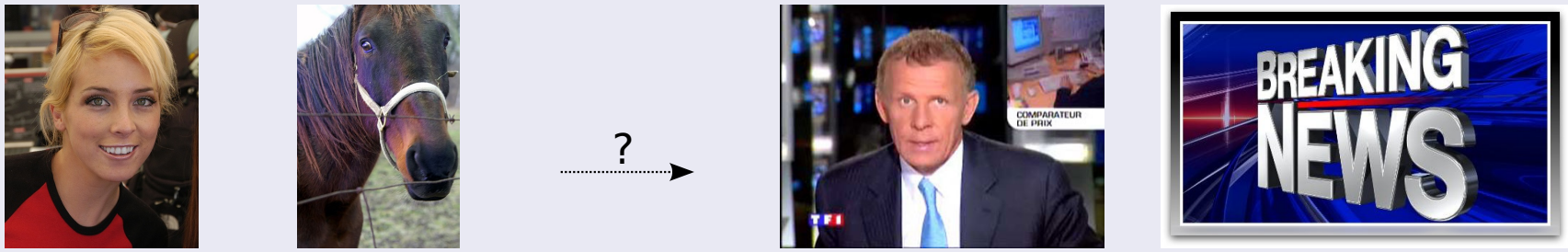
## Definition [Pan, TL-IJCAI'13 tutorial]

– 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?

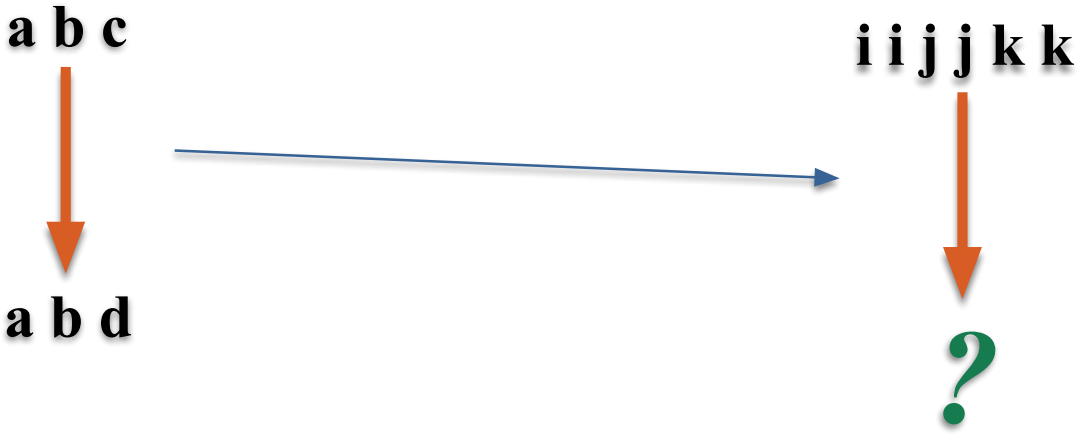
  How to translate formally :

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

Not i.i.d. anymore

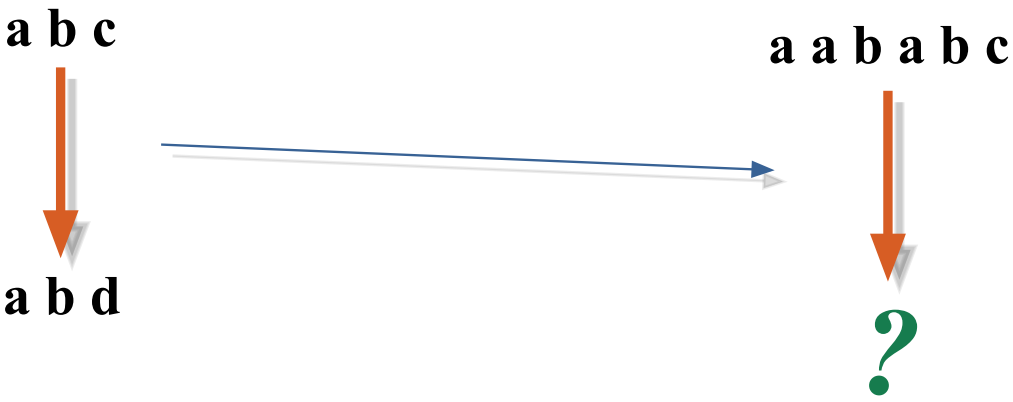
- What **determine a good transfer**?
  - A "good source"?
  - A high "similarity" between source and target?
- What **formal guarantees** can we have on the transferred hypothesis?

# Transfert et analogie



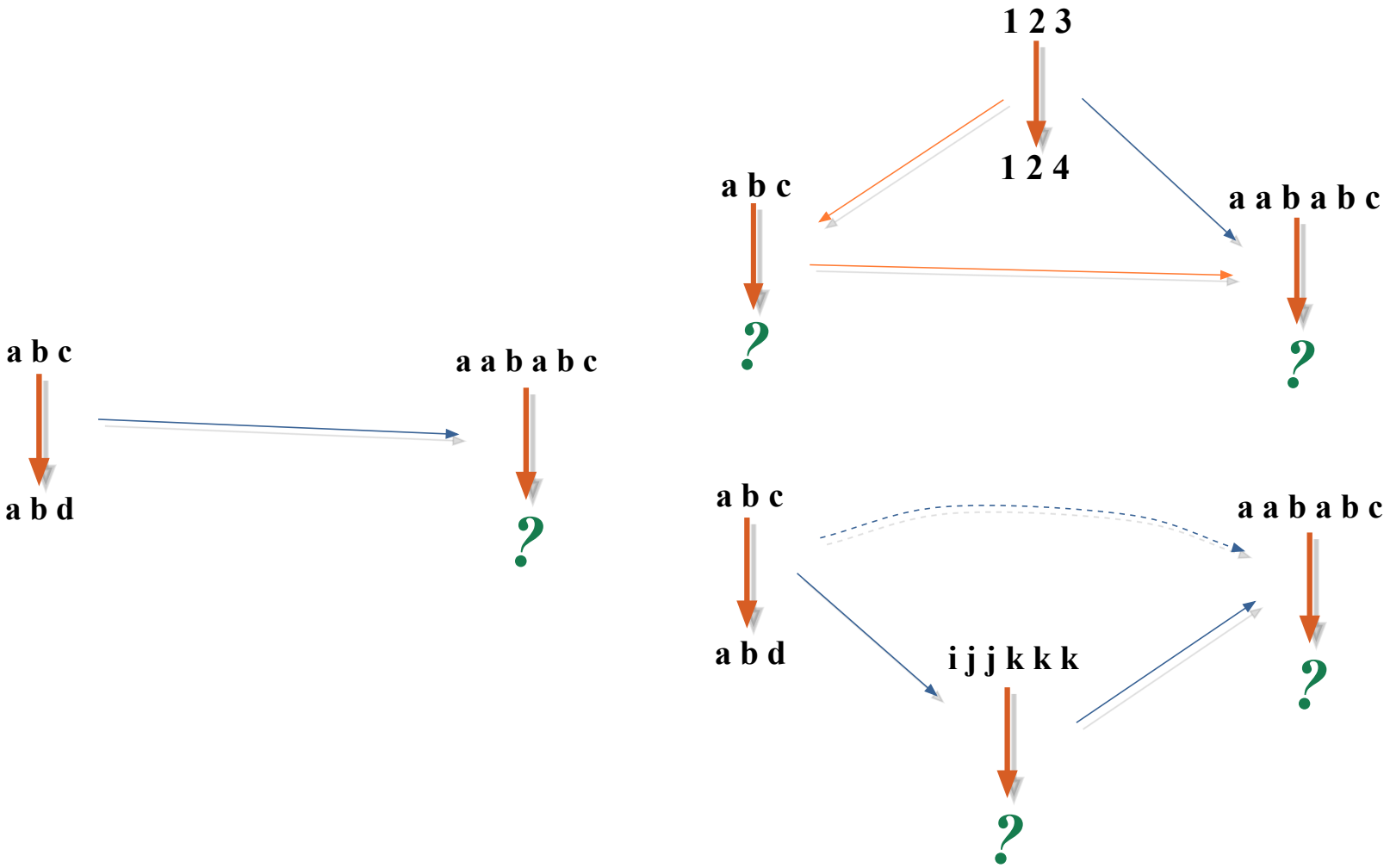

- a b d
- i i j j k d
- i i j j k l
- i i j j k k
- ?

# Transfer and analogy




Why should ‘a a b a b c d’ be any better than ‘a b d’?

# Transfer and sequence effects




- t

# 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

# 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?

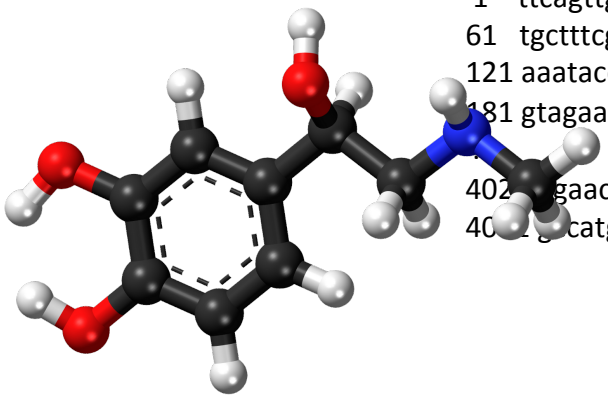
Que savons-nous de l'apprentissage

à partir de données complexes ?

# Intégration de multiple sources de données

- Annotation de protéines

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

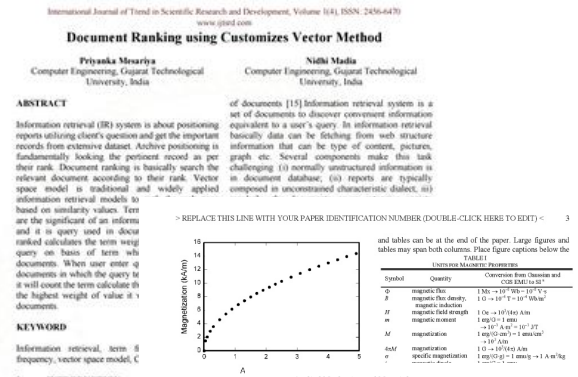


1 ttcagttgtg aatgaatgga cgtgccaaat agacgtgccg ccgccgctcg attcgcactt
61 tgctttcggt tttgccgtcg tttcacgcgt ttagttccgt tcggttcatt cccagttctt
121 aaataccgga cgtaaaaata cactctaacg gtcccgcgaa gaaaaagata aagacatctc
181 gtagaaatat taaaataaat tcctaaagtc gttggtttct cgttcacttt cgctgcctgc

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

# Intégration de multiple sources de données

- GIEC
  - Documents scientifiques multiples
  - Tableaux
  - mesures



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



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**

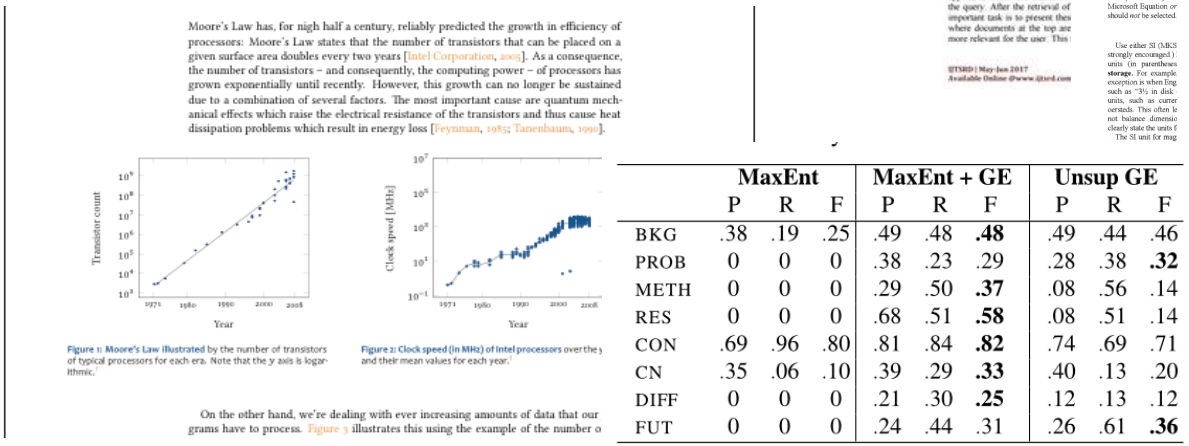
China is one of the countries most affected by the natural disasters in the world. The frequently occurred disasters affect economic development and social stability of the country, causing a great economic losses and casualties. Table 1 is part of economic losses and casualties caused by disasters choose 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



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.

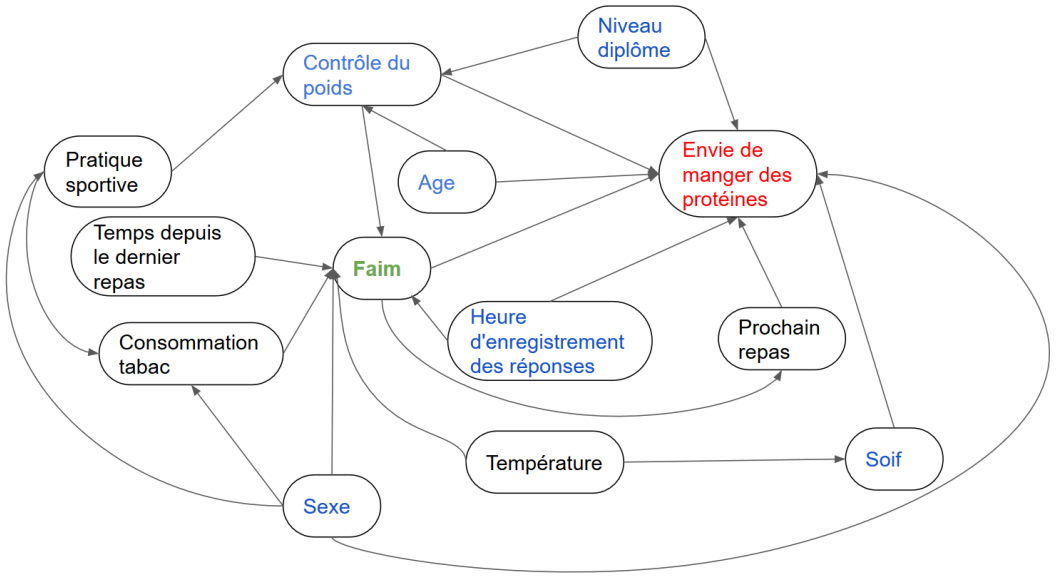
Figure 2: Clock speed (in MHz) of Intel processors over the y and their mean values for each year.

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

| | 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** |

# 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 focaliée 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 ≠ 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**

# L'avenir

# Les réseaux de neurones : jusqu'où ?

- **Reconnaissance** à partir de signal
  - Images
  - Sons
- Découverte de **descripteurs « sémantiques »**
  - E.g. Word2vec
- Problème de **l'opacité**
  - Comment coder des **connaissances a priori**
  - Comment **comprendre le raisonnement** qui conduit à une prédiction
- **Intégration** dans de plus grands systèmes
  - Avec des **systèmes qui raisonnent** (systèmes experts)
  - **Plusieurs systèmes d'apprentissage**

# Mes paris sur les directions à venir

1. Le « **small** data » : nous apprenons très souvent avec très peu
2. La prise en compte de **multiples sources de données** hétérogènes
3. La recherche de **causalités**
4. Apprendre pour **construire des théories ?**
5. L'**intégration** de multiples systèmes apprenants
6. L'apprentissage par **transfert** et **au long cours**
7. 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

# We start to pay attention to new demands

1. The need for explanations
   - Structures
   - **Causal** reasoning
   - No more only error rate
2. The need for transfer learning
   - **What** should be transferred?
   - **Conditions** for positive / negative transfer?

# We start to pay attention to new demands

1. The need for explanations
   - Structures
   - **Causal** reasoning
   - No more only error rate
2. The need for transfer learning
   - **What** should be transferred?
   - **Conditions** for positive / negative transfer?
3. Scenarios away from the i.i.d. assumption
   - Online learning / **changing environments**
   - **Curriculum** learning
   - Long-life learning

# 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?

# 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 ?**

- Pour aller plus loin


http://www2.agroparistech.fr/ufr-info/membres/cornuejols/Research/Tr-Sup-Agro-Montpellier-03-12-2018-v3x4.pdf