Google

Du DIRO à Google

La description automatique d’images

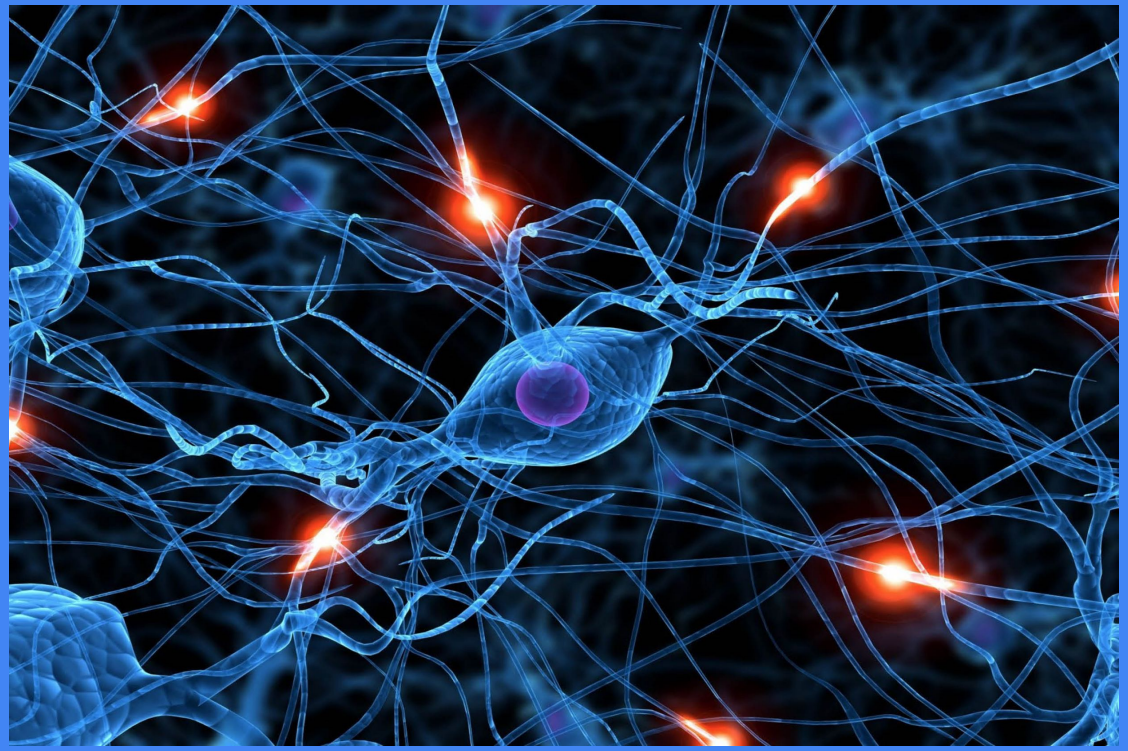
Samy Bengio

Le DIRO: une ambiance studieuse, bien sûr!

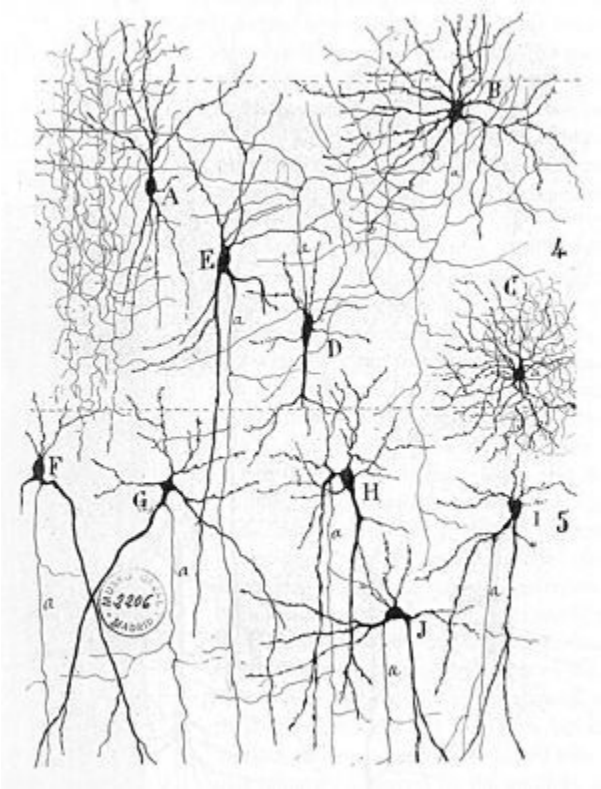
Malgré tout: un premier diplôme en poche!

C'est donc vrai!

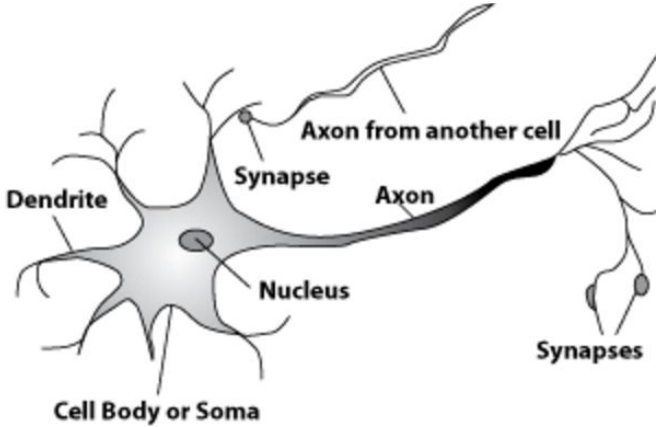
Les réseaux de neurones...



Inspiration biologique

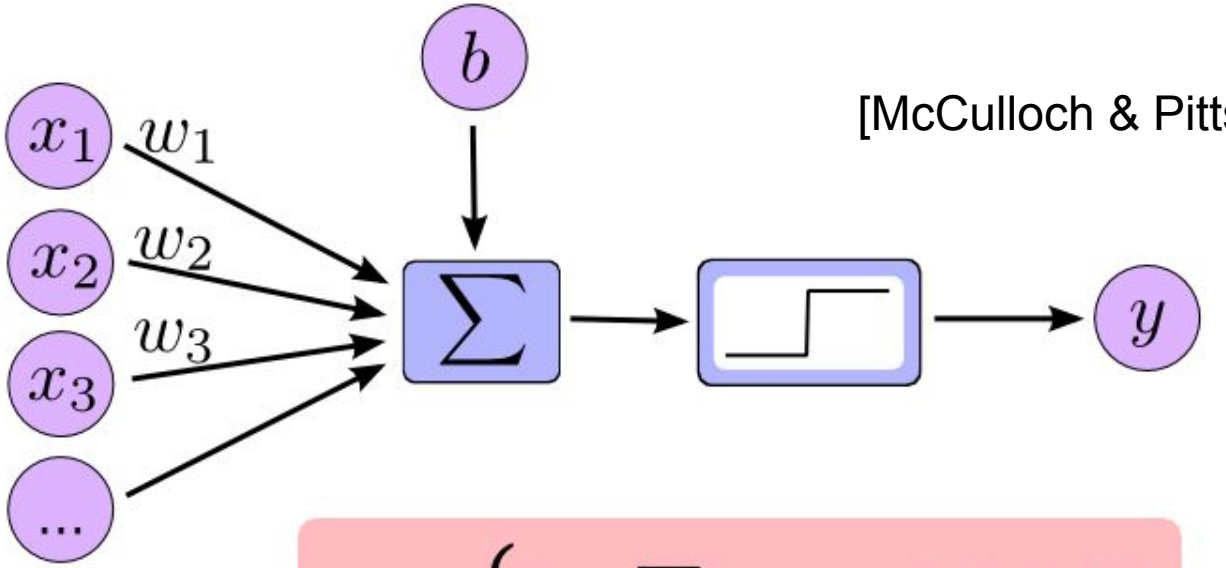


Crédit: Ramón y Cajal, 1905



… et pourquoi pas un modèle de calcul inspiré du fonctionnement du cerveau?

Commençons par un seul neurone...

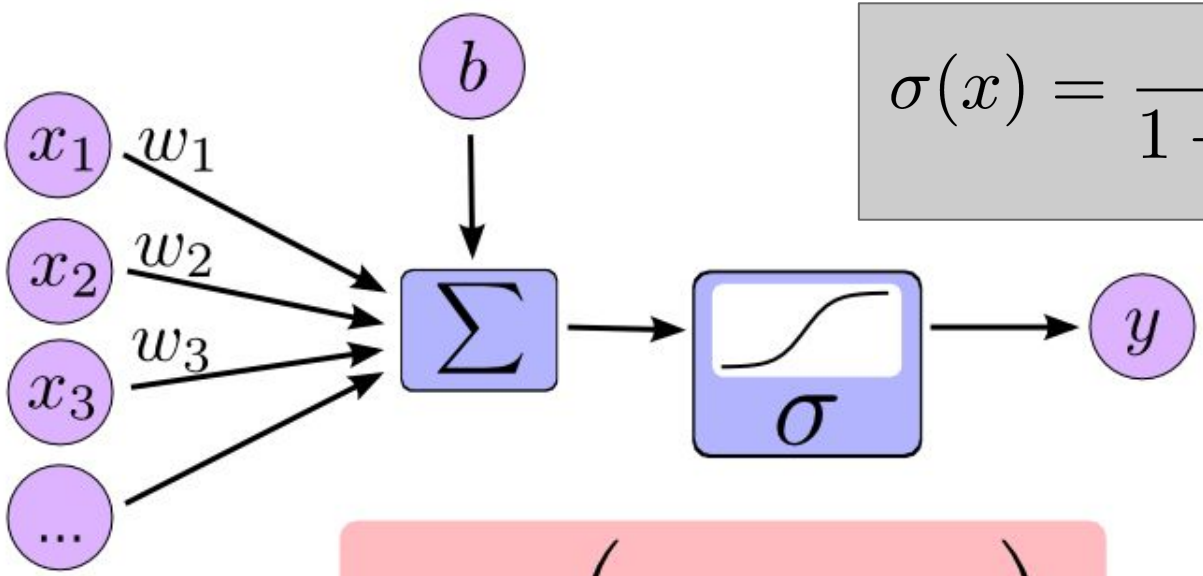


[McCulloch & Pitts, 1943]

$$y = \begin{cases} 1 & \sum_i w_i x_i + b > 0 \\ 0 & \text{else} \end{cases}$$

Crédit: Chris Olah

… on peut aussi changer la fonction...



$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

$$y = \sigma\left(\sum_i w_i x_i + b\right)$$

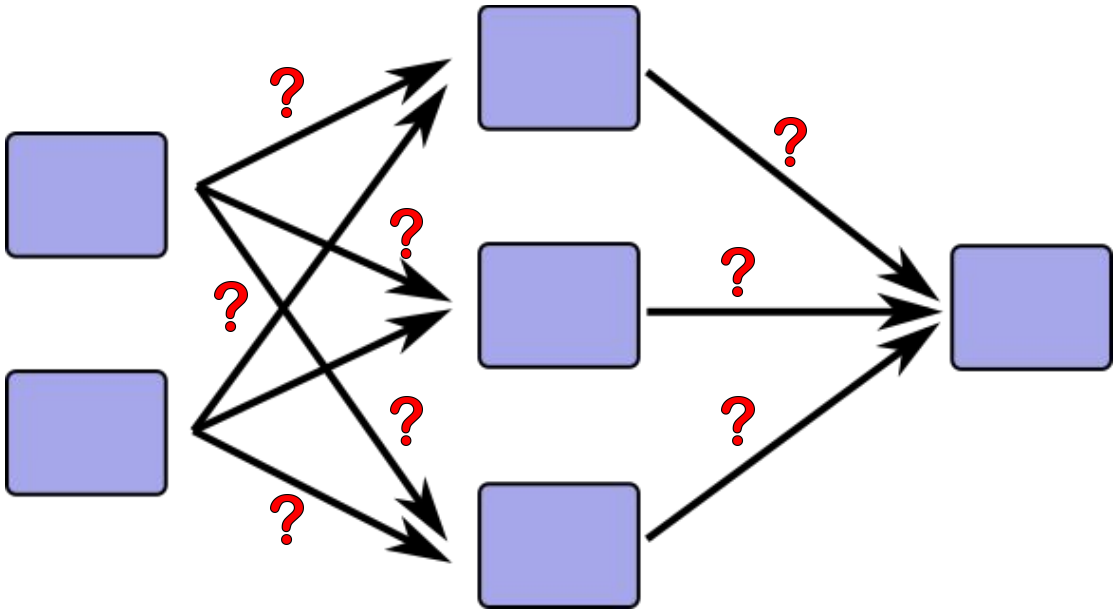
Google

Crédit: Chris Olah

… et connecter plusieurs neurones entre eux

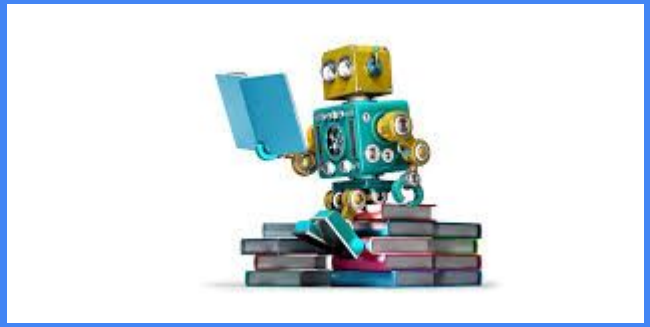
Les réseaux de neurones sont des approximateurs universels

(à condition de trouver la bonne architecture et les bons paramètres!)



Crédit: Chris Olah

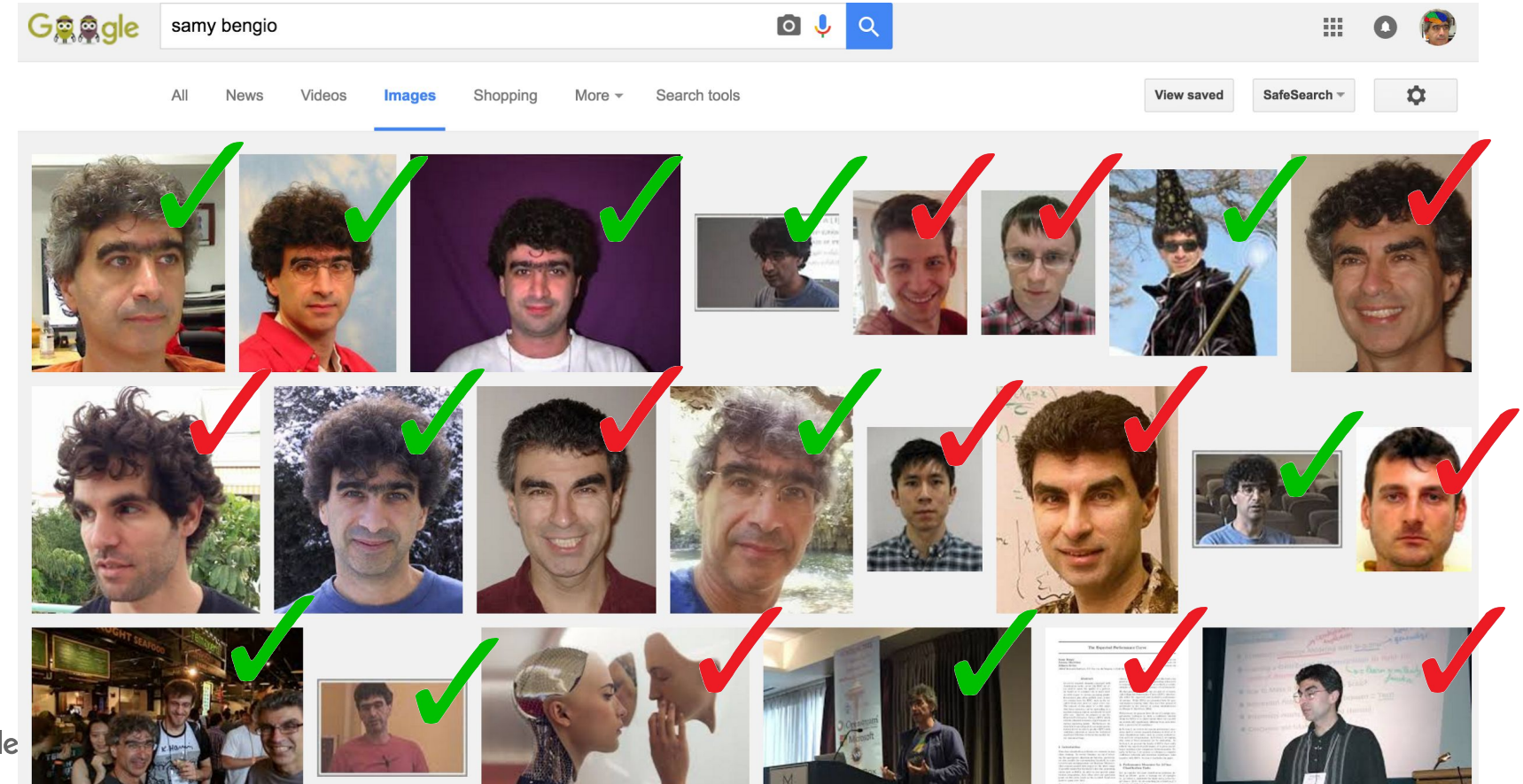
L’apprentissage machine



Apprendre à partir d'exemples

Illustration: cherchons des exemples de photos de Samy Bengio

Apprendre à classifier des images



Trouver un bon séparateur de samy/non-samy

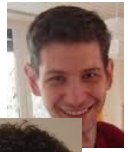




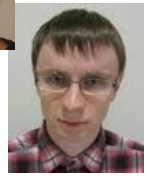






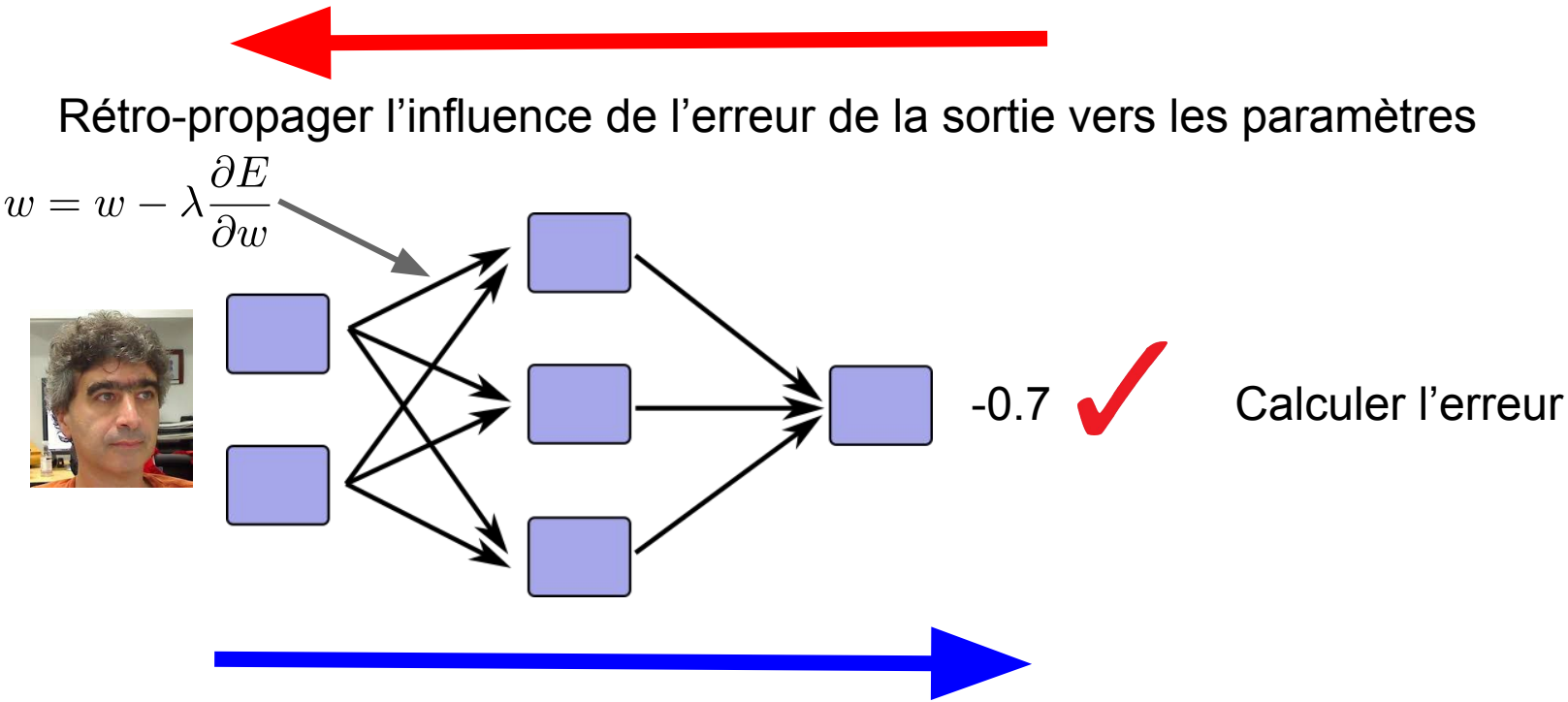





Google

Comment entraîner un réseau de neurones



Interlude sur l’évolution du domaine entre 1990 et 2016...

Les réseaux de neurones en quelques nombres

Autour de 1990:

- Nombre d’étudiants s’y intéressant au DIRO: 1
- Nombre de personnes allant à NIPS: environ 350
- Nombre de neurones dans mes modèles de l’époque: moins de 100

Les réseaux de neurones en quelques nombres

Autour de 1990:

- Nombre d'étudiants s'y intéressant au DIRO: 1
- Nombre de personnes allant à NIPS: environ 350
- Nombre de neurones dans mes modèles de l'époque: moins de 100

… et en 2016:

- Nombre de personnes au MILA: 100
- Nombre de personnes allant à NIPS: 5000
- Nombre de neurones dans les modèles récents: parfois 100 millions

Apprendre à… apprendre?

- Il existe plusieurs algorithmes d'apprentissage
- Lequel utiliser?
- En existe t-il de meilleurs?
- Ma thèse de doctorat (1993):
  - ➢ apprendre à apprendre!
- Sujet de nouveau à la mode ces jours-ci!

$$w = w - \lambda \frac{\partial E}{\partial w}$$

UNIVERSITÉ DE MONTRÉAL

FACULTÉ DES ÉTUDES SUPÉRIEURES

Attendu que le Conseil de la Faculté atteste que

SAMY BENGIO

a terminé les études du programme de doctorat

EN INFORMATIQUE

Nous RECTEUR

par décision du Conseil de l'Université
et en vertu de Notre autorité, lui conférons le grade de

PHILOSOPHIAE DOCTOR (Ph.D.)

à compter du 4 NOVEMBRE 1993 *avec tous les droits, honneurs et privilèges qui s'y rattachent.*

En foi de quoi Nous signons ce document muni du grand sceau de l'Université ainsi que de la signature du secrétaire général et de celles du doyen et du secrétaire de la Faculté.

Fait à Montréal, le 7 DÉCEMBRE 1993

Le doyen — Le secrétaire — Le recteur — Le secrétaire général

Google

Intermède Suisse: 1999 - 2007

Intermède Suisse: 1999 - 2007

idiap
RESEARCH INSTITUTE

Les réseaux de neurones: beaucoup de choix!

- Comment représenter les images? le son? le texte?
  - Pixels? caractères? fréquences? etc
- Combien de couches de neurones?
- Combien de neurones par couche?
- Quelle fonction de transfert par neurone?
- Quel objectif optimiser?
- Quelle technique utiliser pour apprendre les paramètres?
- À quelle vitesse changer les paramètres?
- Quand terminer l'entraînement?
- etc...

Une boîte à outils pour réseaux de neurones

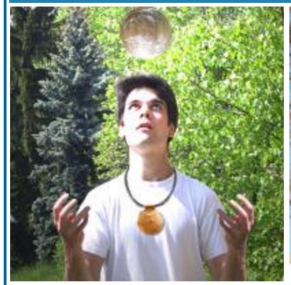
Introduction Documentation Downloads Forum Credits

Ronan Collobert (*collober [at] idiap.ch*)

What's *Torch* ?

It's a **machine-learning library**, written in *simple* **C++** and distributed **now** under a **BSD** license.
Torch is currently developed at Idiap Research Institute, in Switzerland mountains.

The Torch Team



Ronan Collobert
(The Guru)

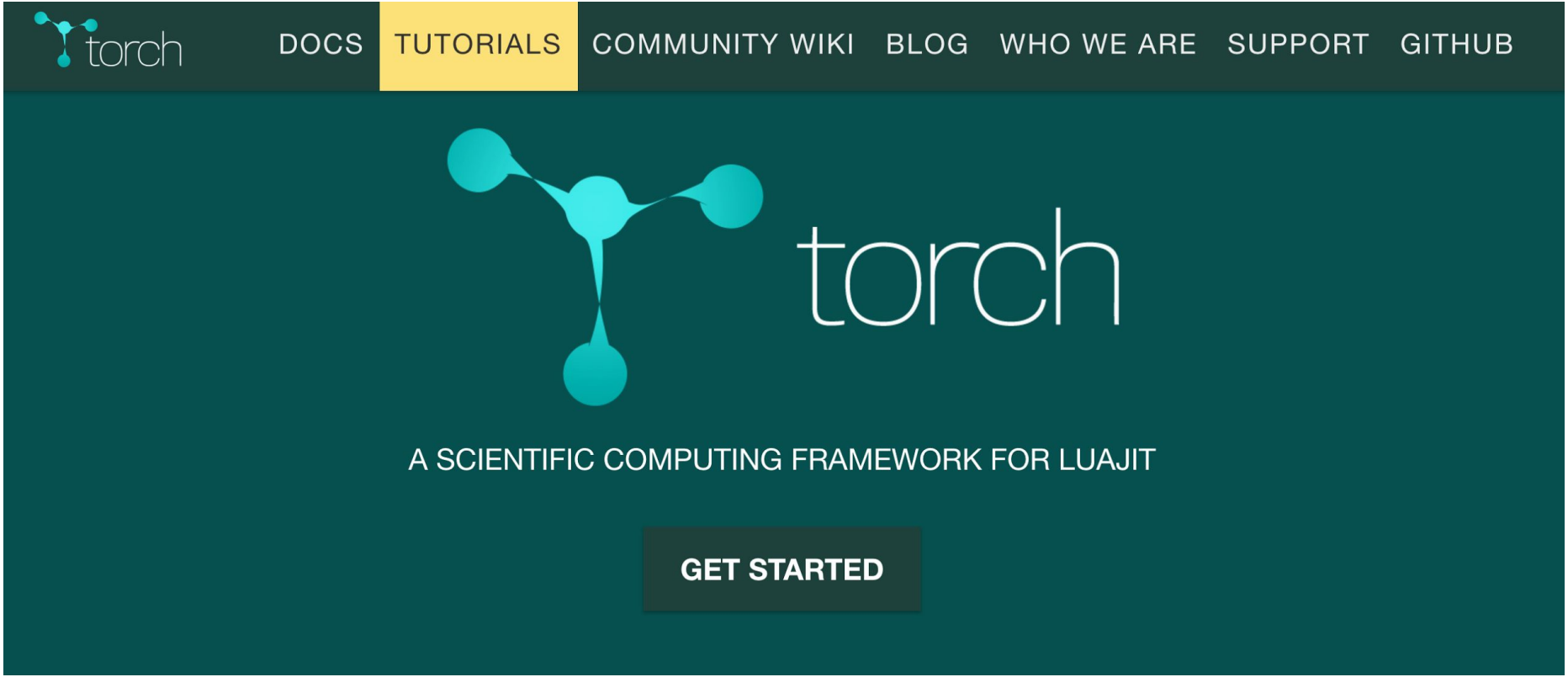
Samy Bengio
(Distributions Wizard)

Johnny Mariéthoz
(Capt'ain Debug)

2002

Google

Torch: quatorze ans plus tard...



… avec une saine compétition...

theano

Développé au DIRO par l'équipe
de mon frère Yoshua

Développé dans mon
groupe à Google

Maintenu a Facebook
(notamment par mon ancien
doctorant Ronan Collobert)

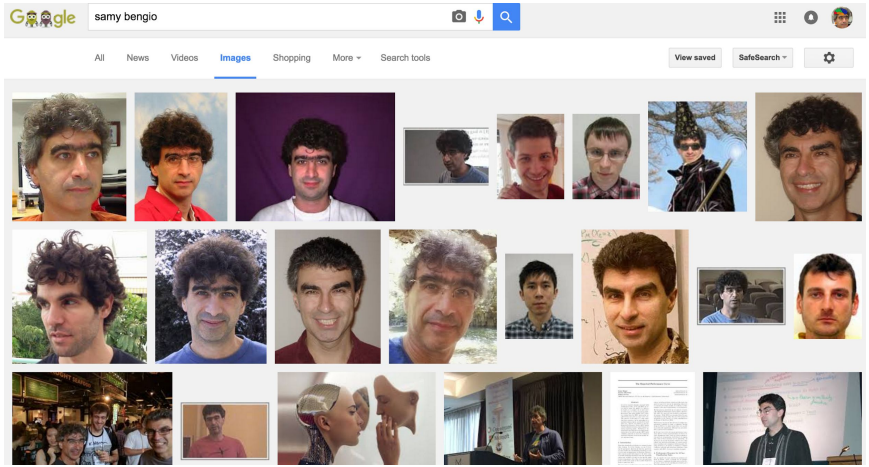
2007: à la recherche de plus de données...

Google

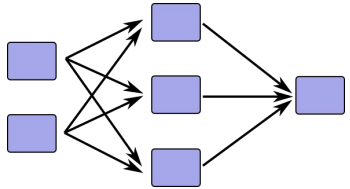
Classifier, c’est bien,
ordonner, c’est mieux

(surtout à Google)

Apprendre à ordonner des images



$$f(\ ) =$$

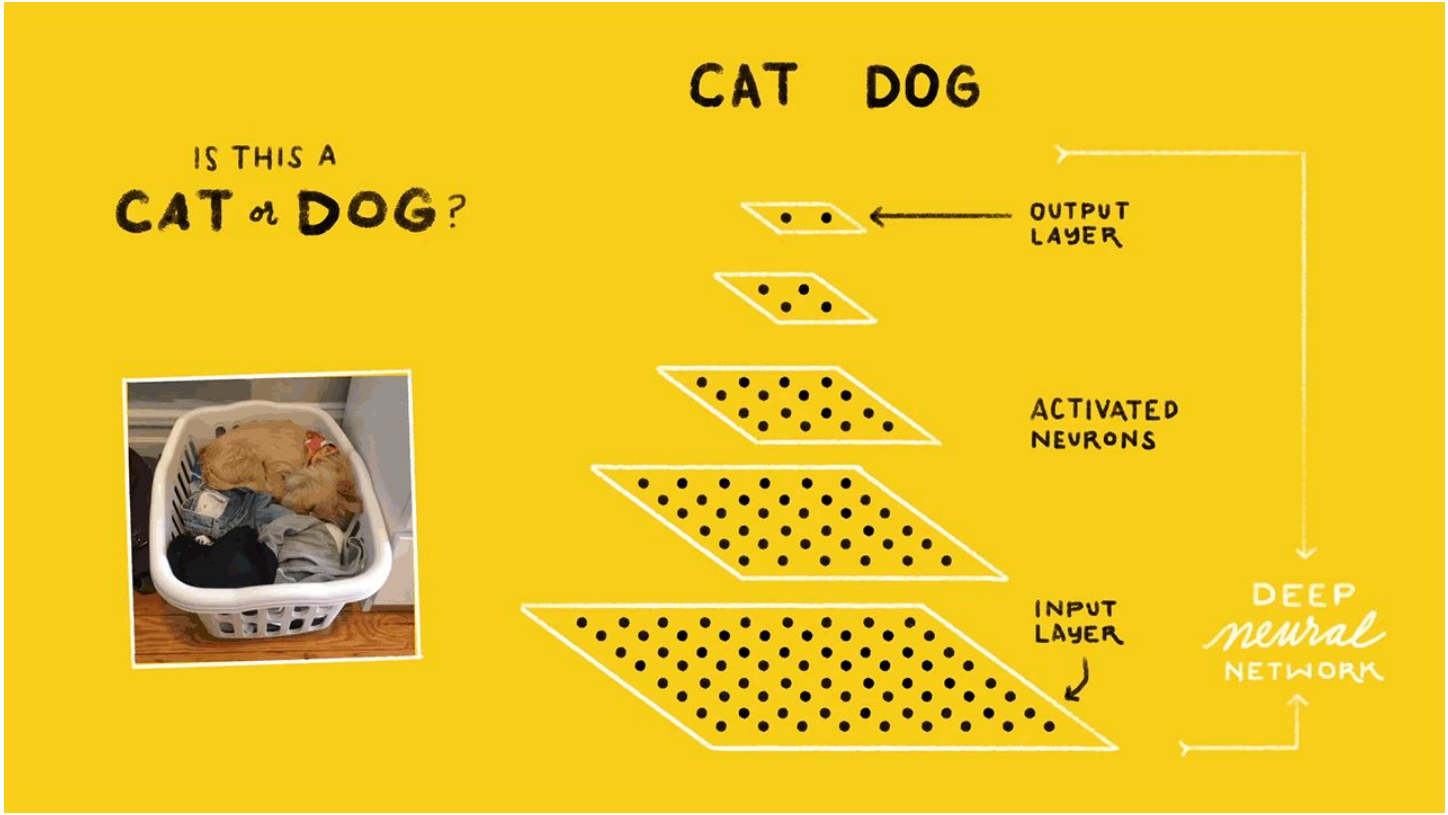


$$f\left(\ \right) > f\left(\ \right) + 1$$

Rajoutez-moi une couche s’il vous plaît

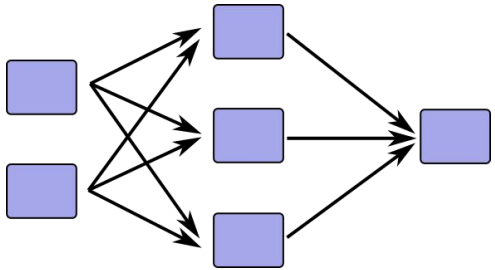
L’apprentissage profond, une révolution fomentée notamment au DIRO...

L’apprentissage profond

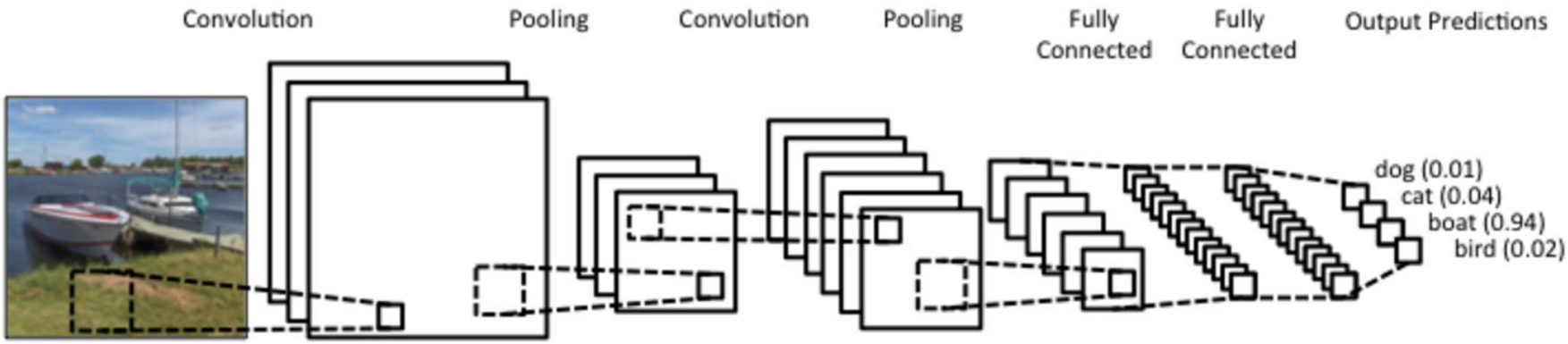


Réseau à Convolutions

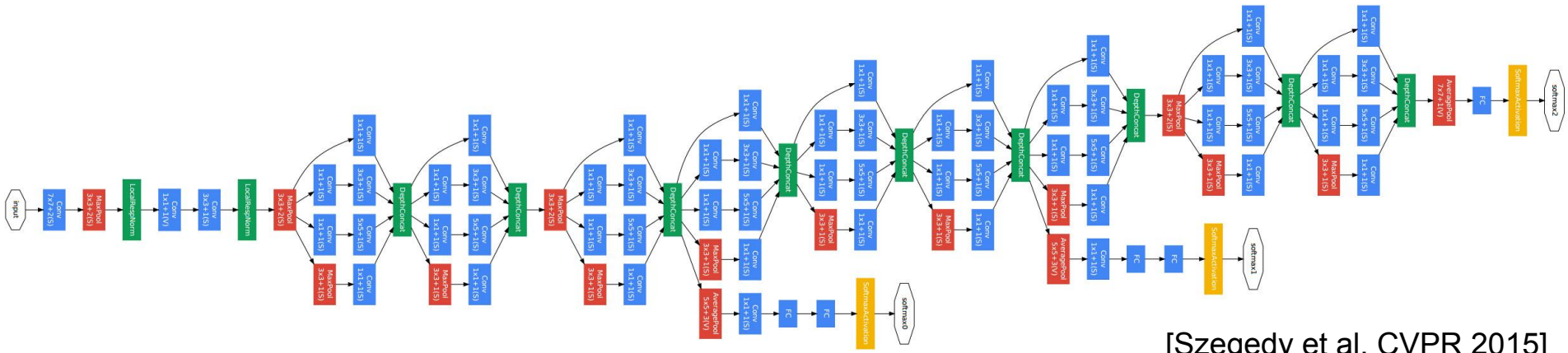
Plutot que des couches de neurones standard comme ceci:



Certaines couches agissent comme des filtres (convolutions), d'autres comme des résumés (pooling):



L’apprentissage profond - photos



[Szegedy et al, CVPR 2015]

- Compétition annuelle depuis 2010 de reconnaissance d’images: ImageNet Challenge, 1000 catégories, 1M d’images.
- En 2012, les gagnants utilisent un réseau profond pour la première fois et gagnent facilement la compétition.
- Depuis, tout le monde utilise des réseaux profonds!

Classification “fine”

“hibiscus”

“dahlia”

Généralise à différent contextes

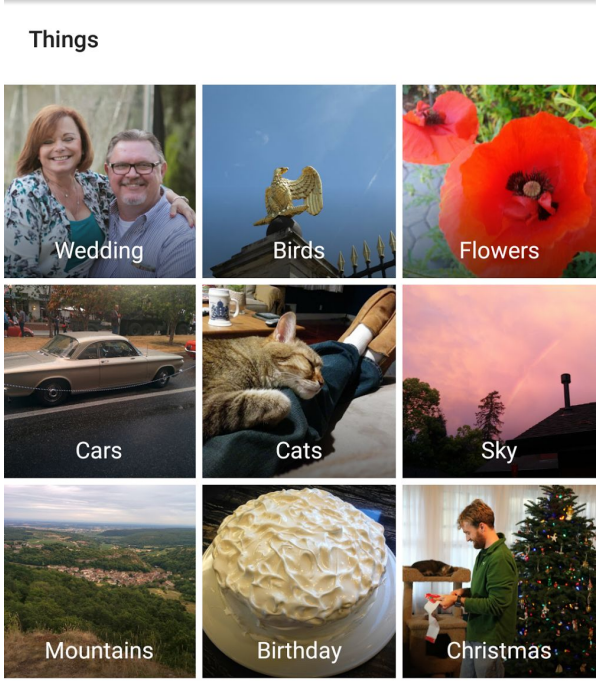
Les deux sont classés “meal”

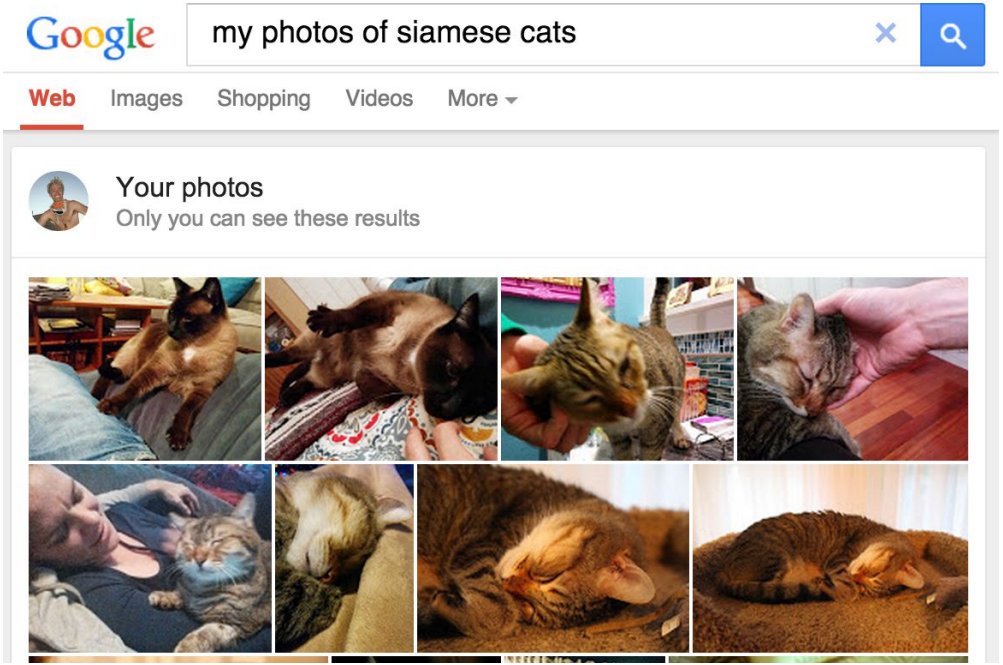
Les erreurs sont “raisonnables”

“snake”

“dog”

Google Photos Search





C’est bien beau les images, mais comment représenter des mots?

La méthode du dictionnaire

On attribue un numéro unique à chaque mot:

1. baleine
2. Berlin
3. dauphin
4. Obama
5. Paris
6. ...

On transforme les indices en codes binaires:

1. 00001
2. 00010
3. 00100
4. 01000
5. 10000
6. ...

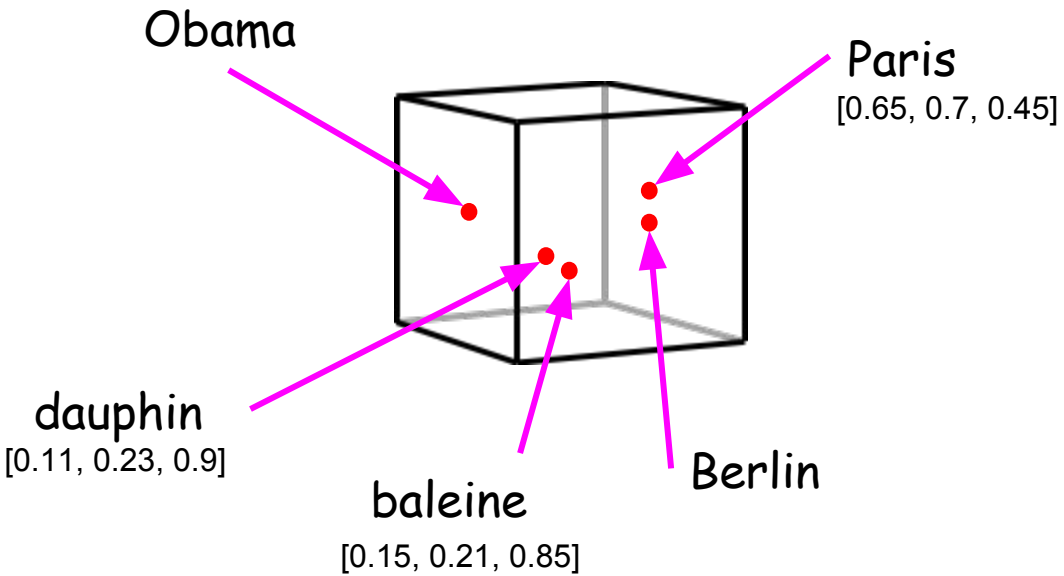
Le(s) problème(s) de cette représentation

- Le code doit être aussi long que la taille du dictionnaire.
- La représentation binaire des mots ne permet pas de les comparer:
  - Deux mots proches sémantiquement ne sont pas plus proches “en code binaire” que deux mots qui n’ont rien à voir: (par exemple en comptant le nombre de bits qui diffèrent):
    - dauphin (00100) et baleine (00001) ont 2 bits qui diffèrent
    - dauphin (00100) et Paris (10000) ont aussi 2 bits qui diffèrent

Une représentation alternative

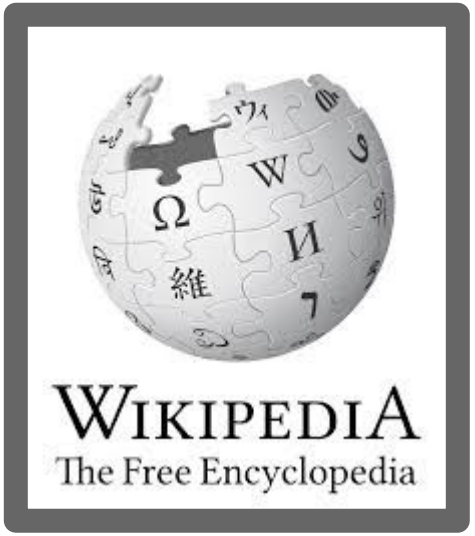
Représentons les mots par des vecteurs de nombres réels

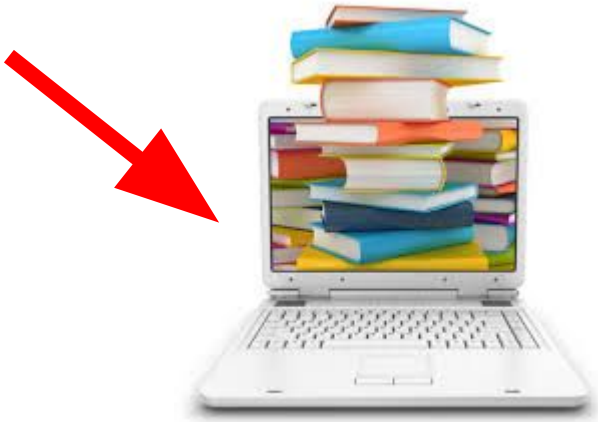
(en anglais, on parle d'*embeddings*)



Comment trouver cette représentation?

Tout commence par l'ingestion d'un grand corpus de textes, comme Wikipedia

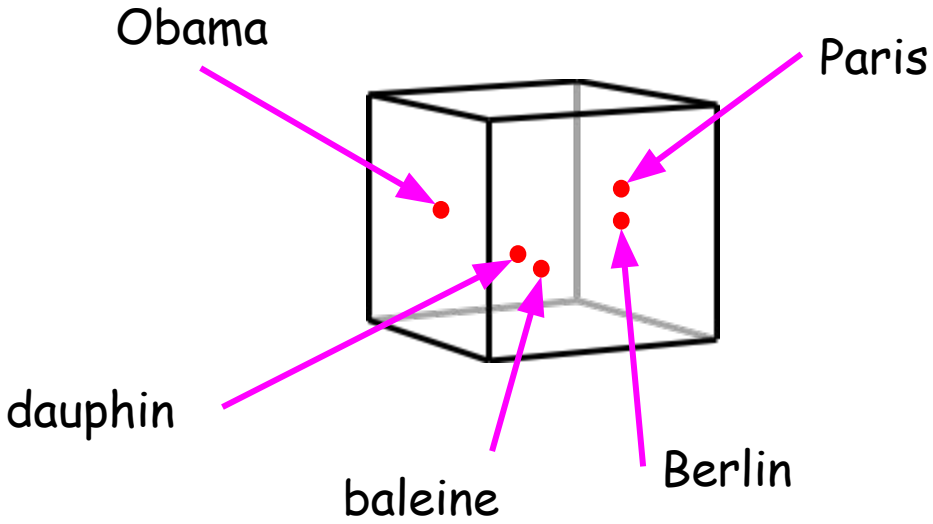




Comment trouver cette représentation?

Une approche simple (**Word2Vec**):

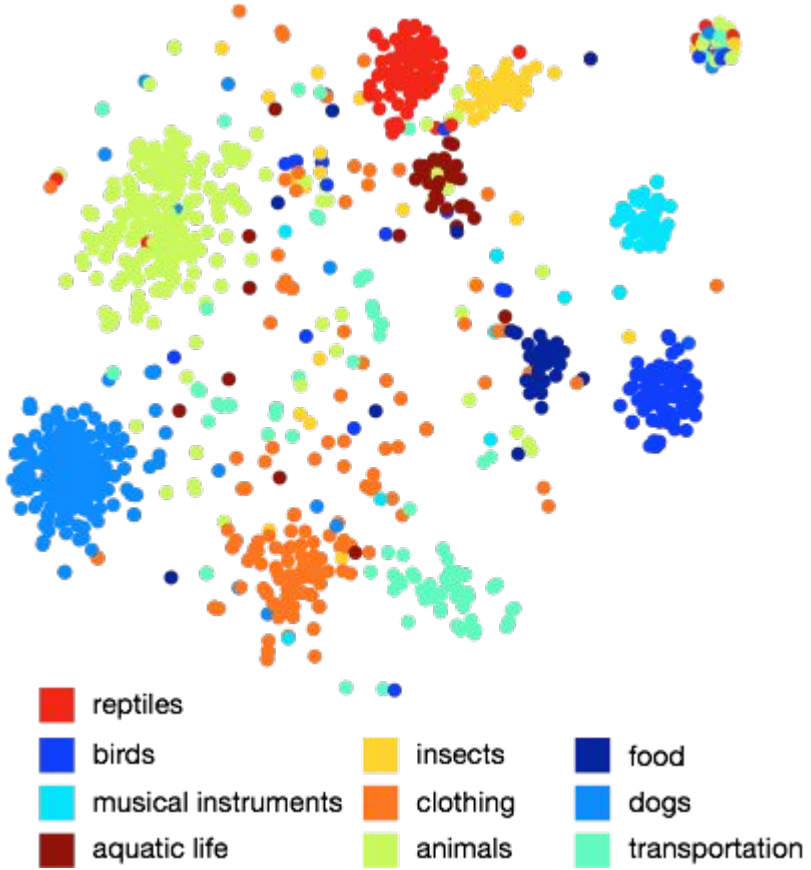
- Commencer par des représentations aléatoires de mots.
- Prendre une phrase au hasard
  - *Il habitait entre Paris et Berlin*
- Choisir deux mots au hasard:
  - *Paris* et *Berlin*
- Déplacer légèrement la représentation de chaque mot l'un vers l'autre et repousser les autres.
- Répéter autant de fois que possible.



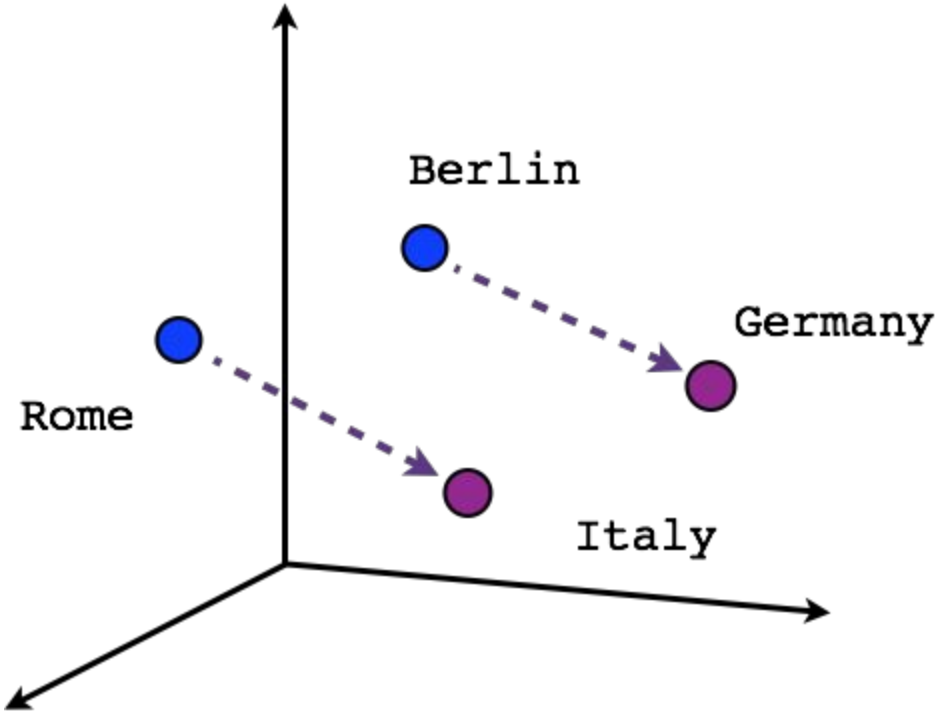
[Mikolov et al, 2013]

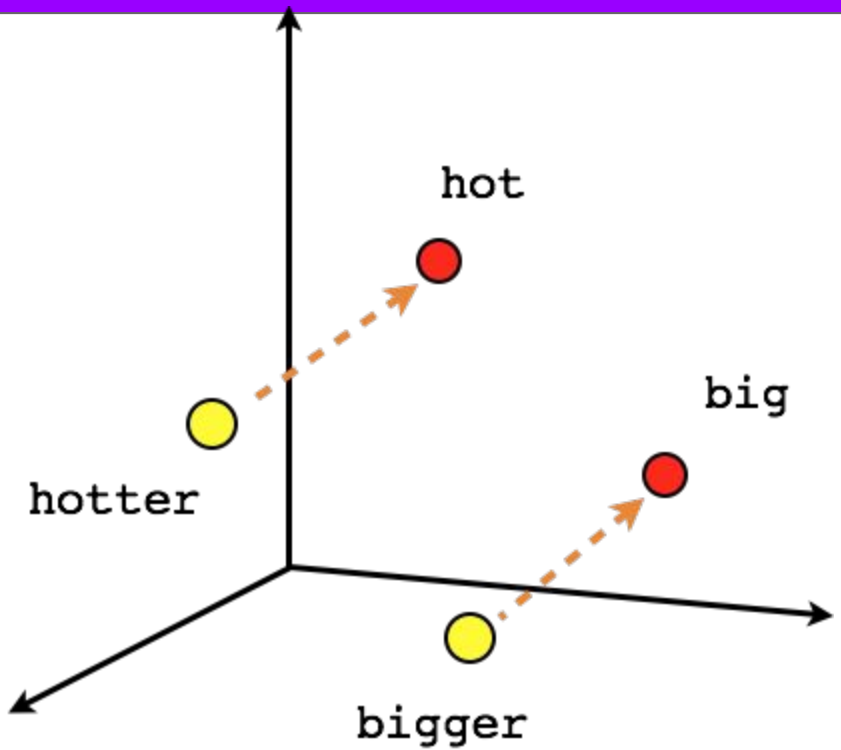
Word2Vec sur le corpus de Wikipedia

tiger shark	**car**
bull shark	cars
blacktip shark	muscle car
shark	sports car
oceanic whitetip shark	compact car
sandbar shark	autocar
dusky shark	automobile
blue shark	pickup truck
requiem shark	racing car
great white shark	passenger car
lemon shark	dealership



Étonnante propriété de cette représentation





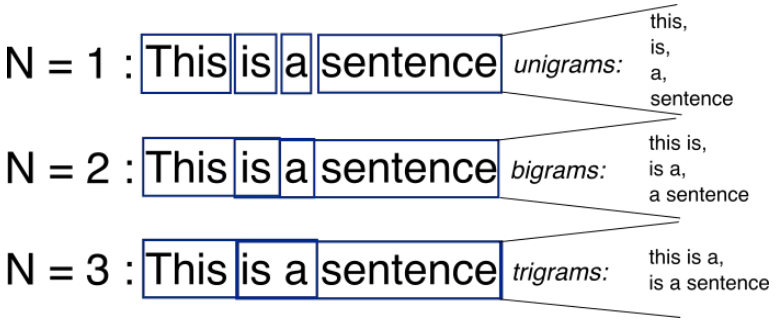
E(*Rome*) - E(*Italy*) + E(*Germany*) ≈ E(*Berlin*)

E(*hotter*) - E(*hot*) + E(*big*) ≈ E(*bigger*)

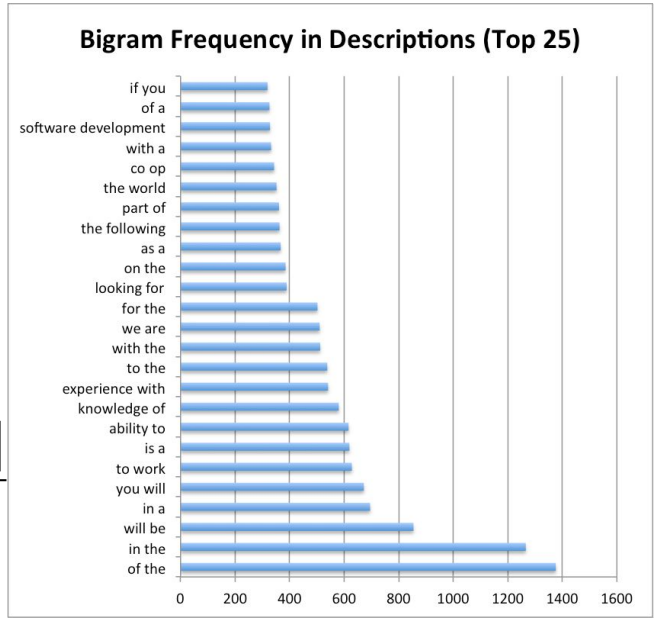
Après les mots, les phrases!

Commençons par compter

- **Quelle est la probabilité qu'une phrase soit prononcée?**
- Deux approches: demander à un linguiste ou… **observer les données**!

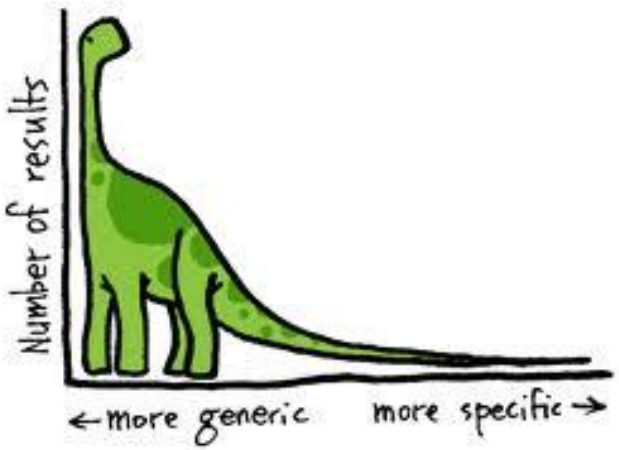


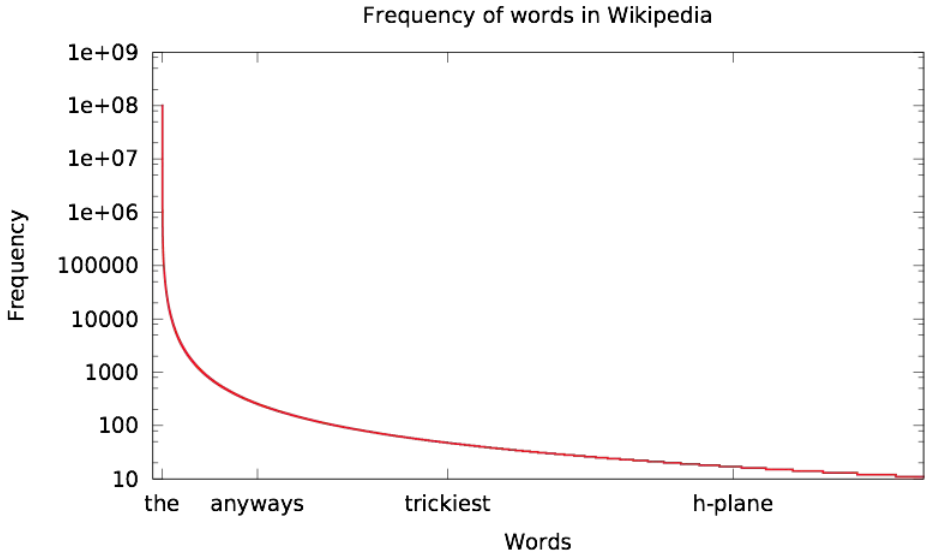
$$P([\text{Holmes}]|[\text{Sherlock}]) \approx \frac{\#[\text{Sherlock Holmes}]}{\#[\text{Sherlock}]}$$



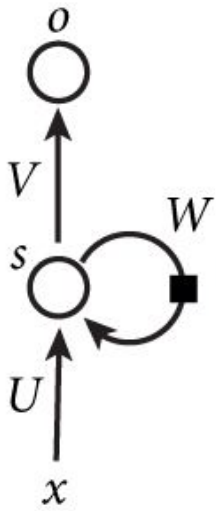
Mais que faire des cas rares?

- Quelques mots sont très populaires.
- Mais la plupart sont très rares.
- Et c'est encore pire pour les combinaisons de deux mots ou plus!



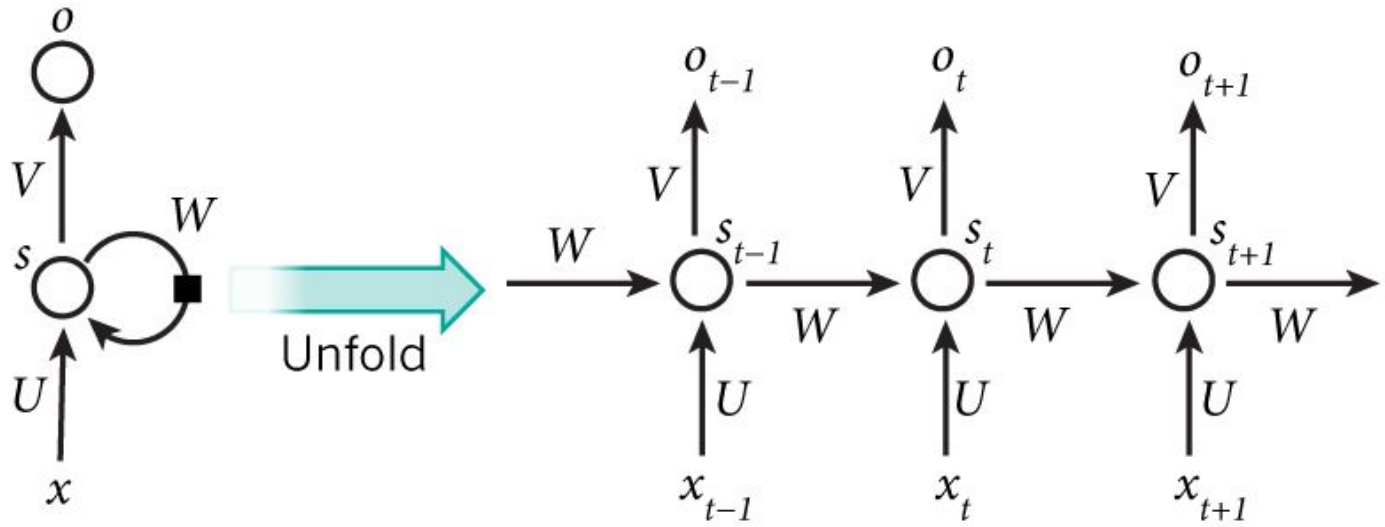


Les réseaux de neurones récurrents



Crédit: WildML

Les réseaux de neurones récurrents



$$o_t = f(s_t; V)$$

$$s_t = g(x_t, s_{t-1}; U, W)$$

Crédit: WildML

Les réseaux de neurones récurrents

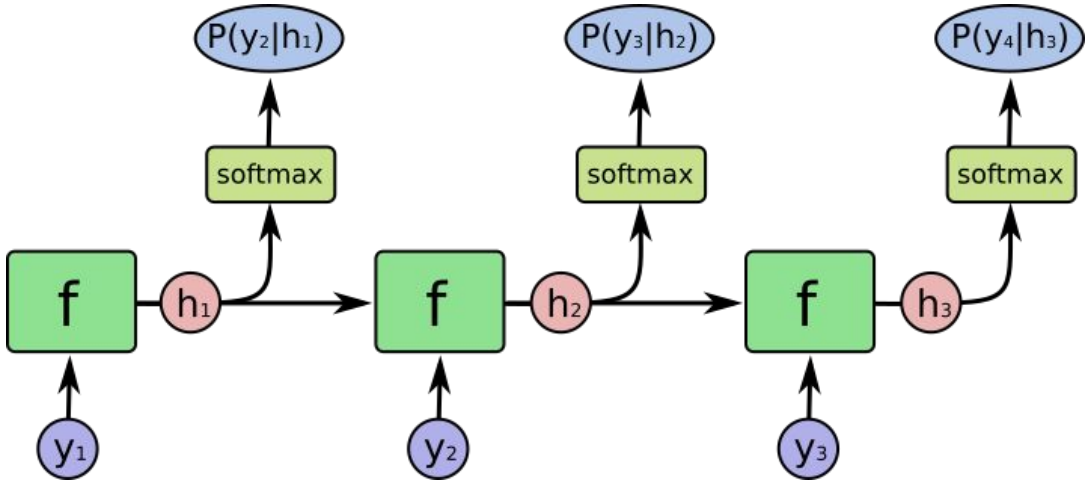
$$P(y_1, y_2, \cdots, y_T) = \prod_t P(y_t | y_1, \cdots, y_{t-1})$$

La probabilité d'une phrase peut se décomposer en produit de probabilités de chaque mot étant donné les mots précédents

Les réseaux de neurones récurrents

$$P(y_1, y_2, \cdots, y_T) = \prod_t P(y_t | y_1, \cdots, y_{t-1})$$

$$\approx \prod_t P(y_t | h_{t-1}) \text{ avec } h_t = f(y_t, h_{t-1})$$

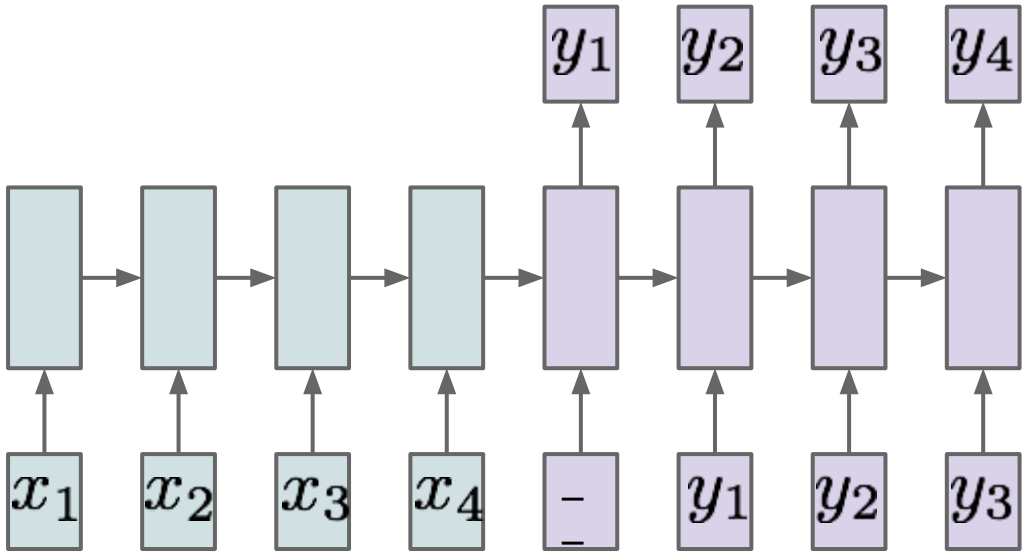


Crédit: Chris Olah

Le modèle “sequence-to-sequence”

$$p(y_1, \cdots, y_{T'} | x_1, \cdots, x_T) = \prod_{t=1}^{T'} p(y_t | y_1, \cdots, y_{t-1}, x_1, \cdots, x_T)$$

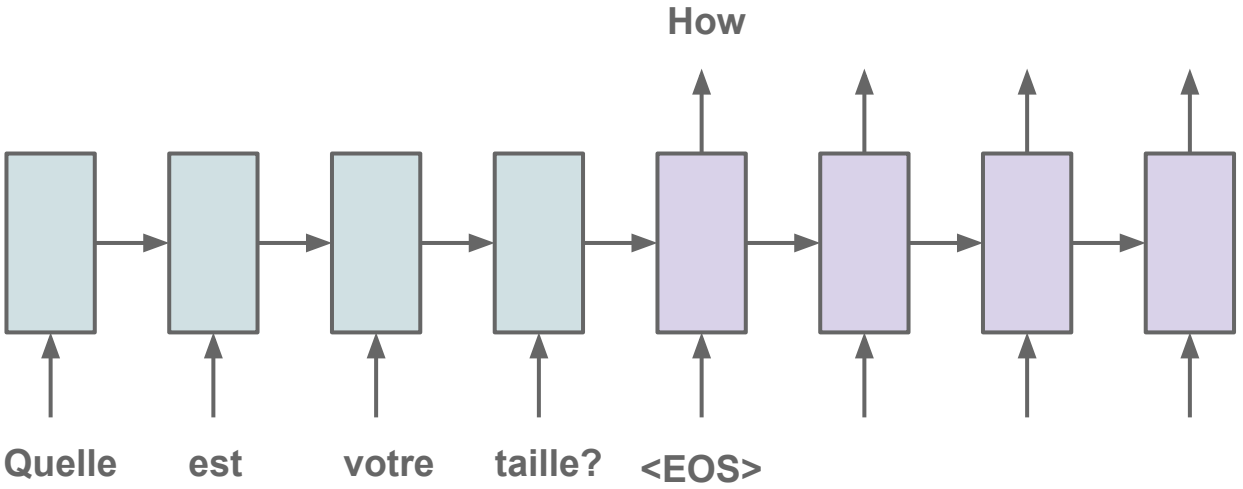
Phrase en anglais



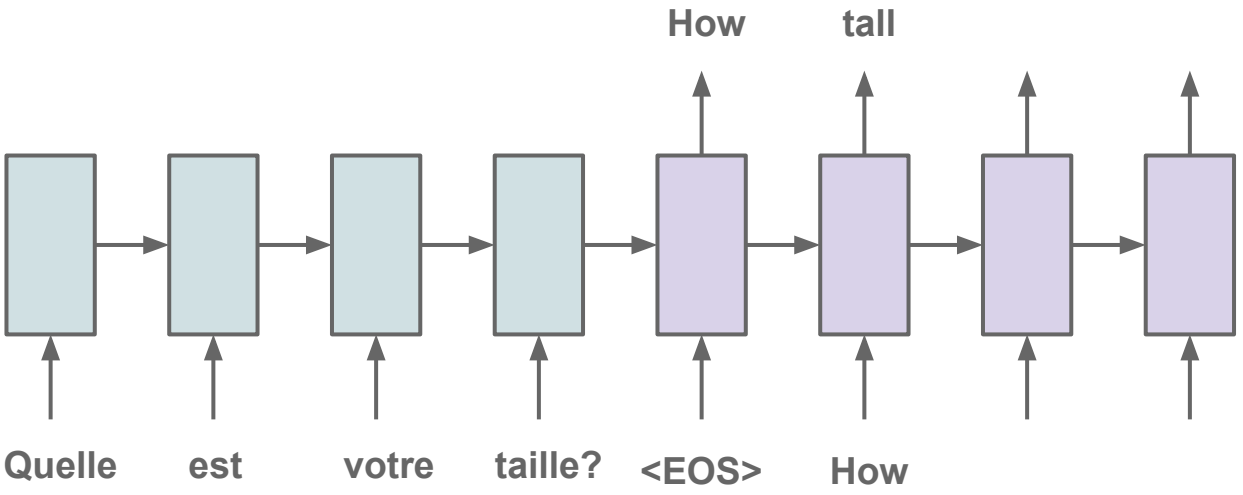
Phrase en français

[Sutskever et al, NIPS 2014]

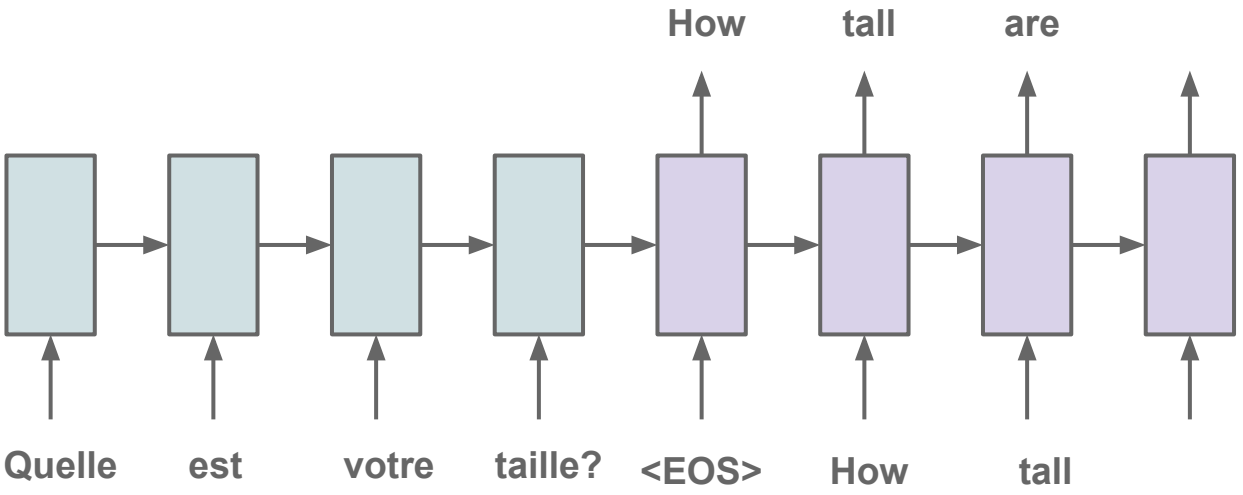
Exemple pour la traduction automatique



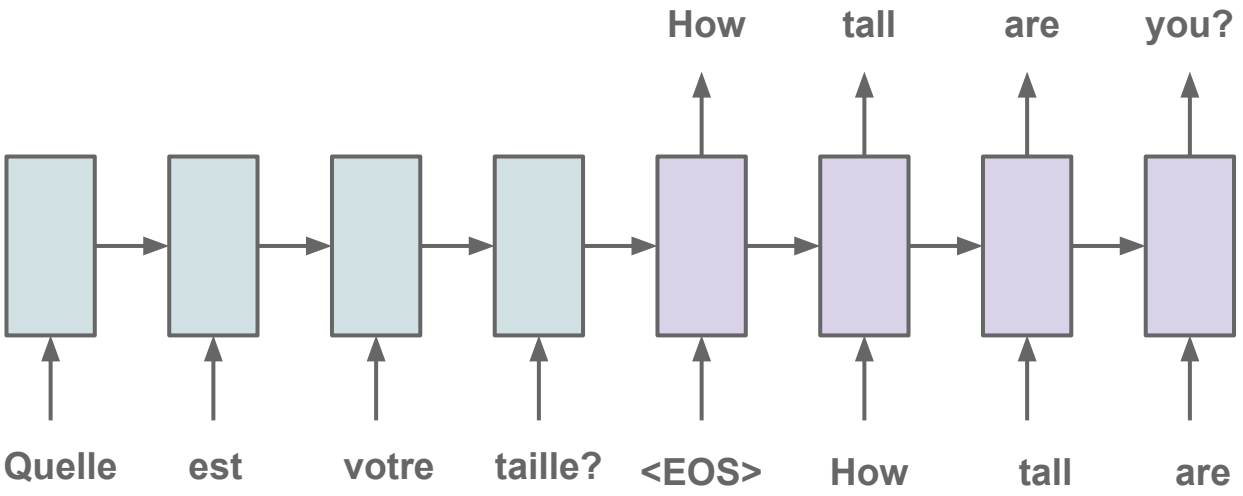
Exemple pour la traduction automatique



Exemple pour la traduction automatique



Exemple pour la traduction automatique

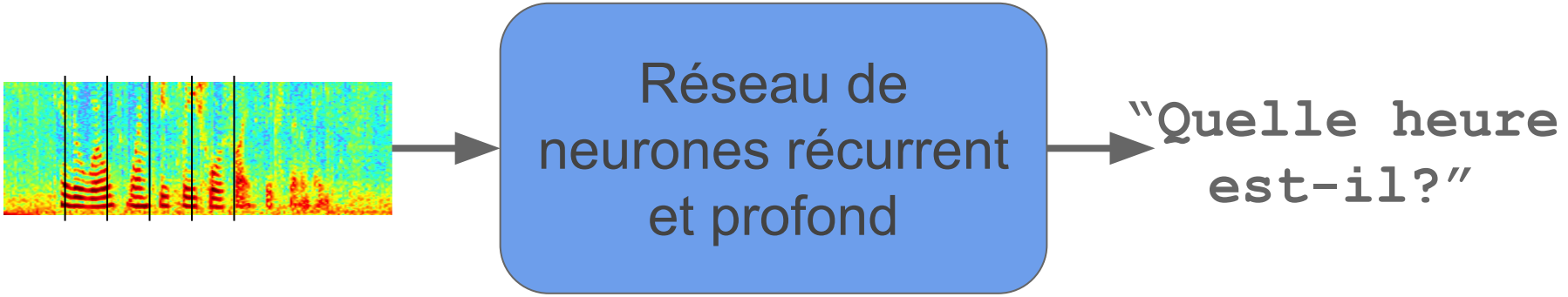


Le nouveau système de traduction automatique de Google utilise une approache basée sur cette idee...

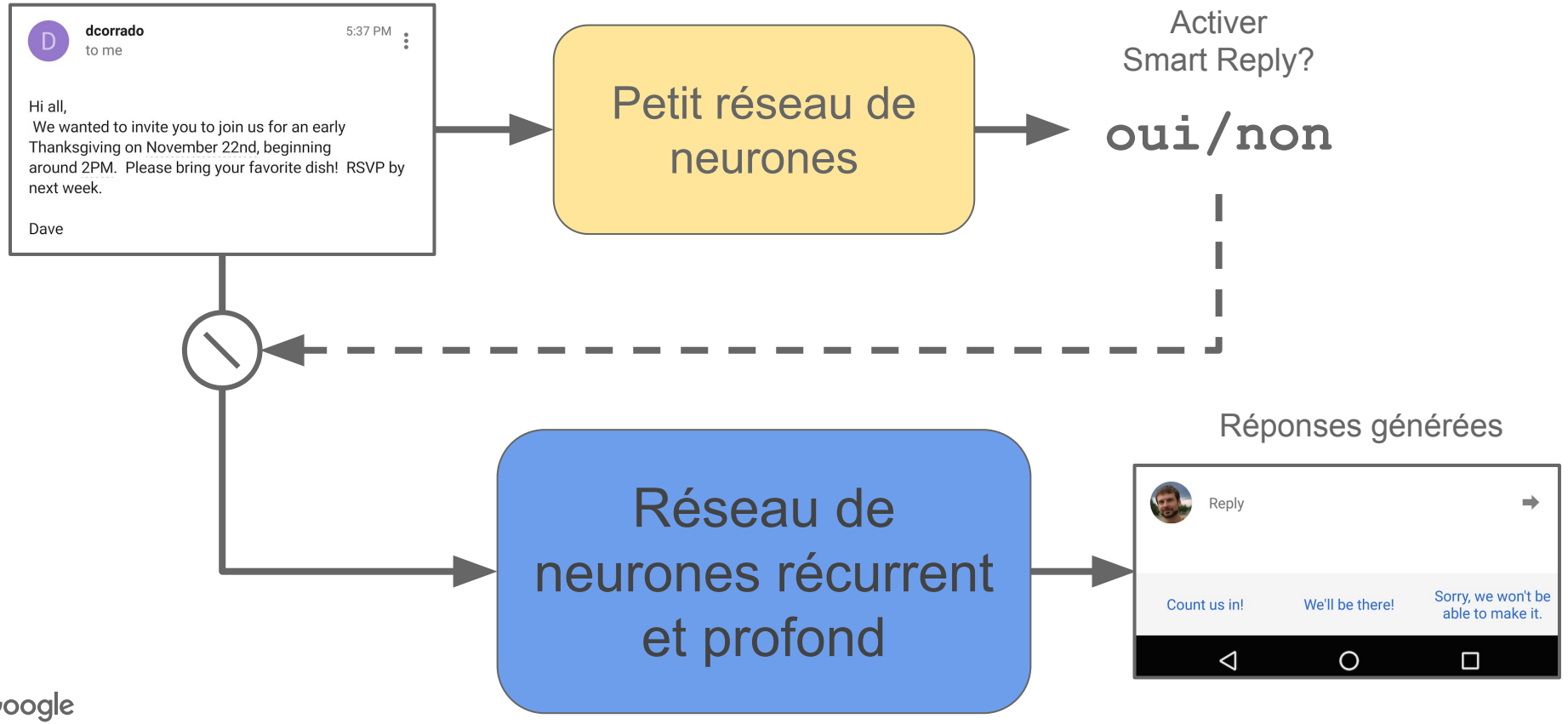
De la recherche aux produits…
… en très peu de temps!

Un environnement qui favorise l'intégration

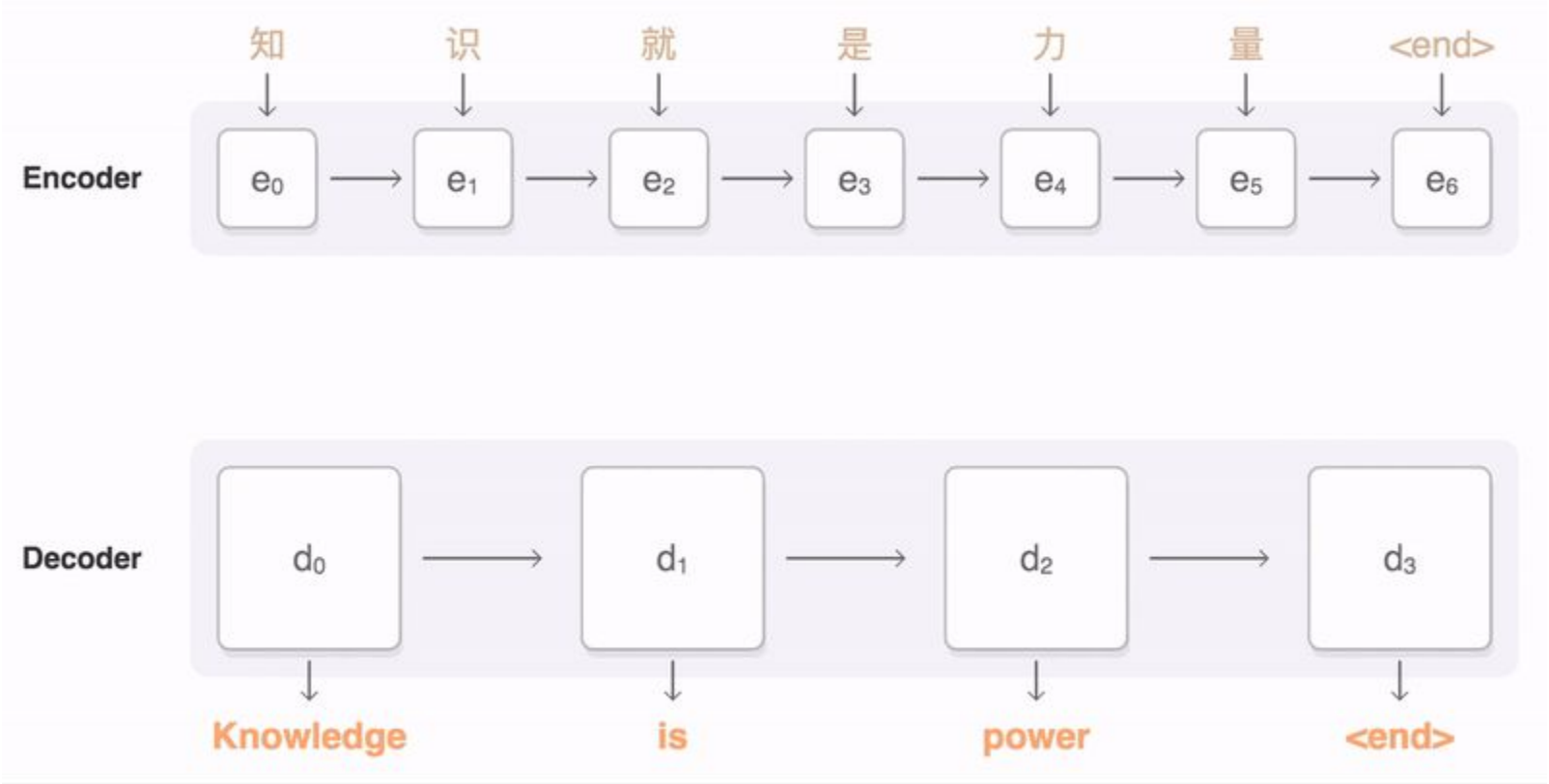
- Une seule base de code à Google!
- Lorsque du code est intégré à la base, il est utilisable par tout le monde.
- C'est pareil pour les chercheurs.
- Chaque ligne de code est révisée par d'autres informaticiens.
- En 2-3 mois, on passe de la recherche au produit.
- Exemple: la reconnaissance de la parole...



Exemple de produit: Smart Reply



Exemple de produit: traduction automatique



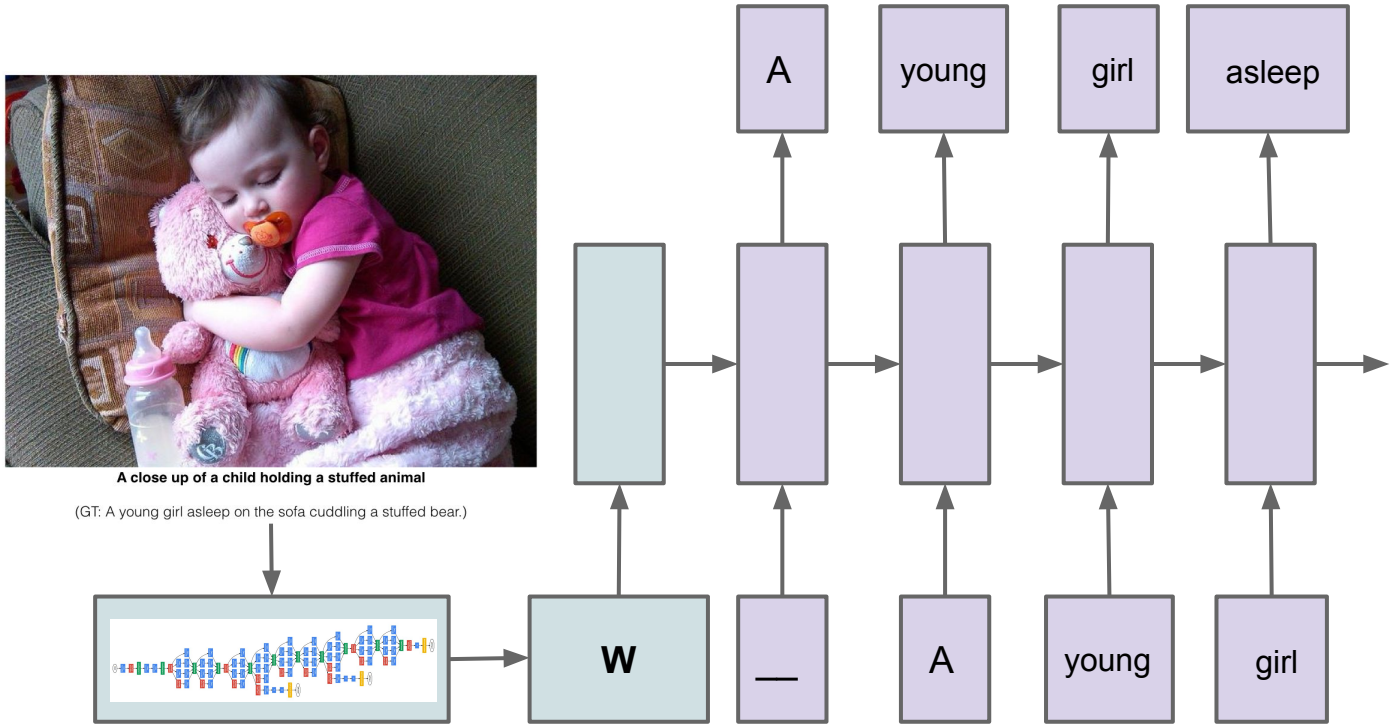
La description automatique d’images

La description automatique d’images

P(phrase en anglais | phrase en francais)

P(phrase en anglais | Image)

La description automatique d'images



[Vinyals et al, CVPR 2015]

Données et modèles

- Tout a commencé par des données: une base créée par Microsoft en 2014: MS-COCO. A l'époque:
  - 75,000 images d'entraînement
  - 5,000 images de test
  - chaque image étant pourvue de 5 descriptions différentes.
- Modèle d'images: GoogleLeNet (gagnant du concours ImageNet Challenge 2014)
- Modèle de langue: LSTM avec 512 neurones.
- Les mots sont représentés par des embeddings de taille 512.
- Petit dictionnaire de 8857 mots seulement.

Un premier essai...

NIC = Neural Image Captioning

Human: A young girl asleep on the sofa cuddling a stuffed bear.

NIC: A close up of a child holding a stuffed animal.

NIC: A baby is asleep next to a teddy bear.

Compétition MS COCO - 2015

	M1	M2	TOTAL	Ranking
Google	5	4	9	1st(tie)
MSR	4	5	9	1st(tie)
Montreal/Toronto	3	2	5	3rd(tie)
MSR Captivator	2	3	5	3rd(tie)
Berkeley LRCN	1	1	2	5th

Method	Meteor	CIDEr	LSUN
Google NIC	**0.346 (1)**	**0.946 (1)**	**0.273 (2)**
MSR Capt	0.339 (2)	0.937 (2)	0.250 (3)
UCLA/Baidu v2	0.325 (5)	0.935 (3)	0.223 (5)
MSR	0.331 (4)	0.925 (4)	**0.268 (2)**
MSR Nearest	0.318 (10)	0.916 (5)	0.216 (6)
Human	0.335 (3)	0.910 (6)	**0.638 (1)**
UCLA/Baidu v1	0.320 (8)	0.896 (7)	0.190 (9)
LRCN Berkeley	0.322 (7)	0.891 (8)	0.246 (4)
UofM/Toronto	0.323 (6)	0.878 (9)	0.262 (3)

Quelques exemples de description d'image

Human: A man and a woman in wedding attire feeding a giraffe.

NIC: A man feeding a giraffe at a zoo.

Quelques exemples de description d’image

Human: A road sign on the side of a road.

NIC: A street sign on the side of the road.

Quelques exemples de description d'image

Human: A brown dog laying in a red wicker bed.

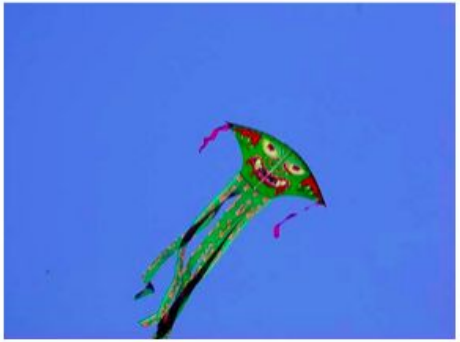
NIC: A small dog is sitting on a chair.

Encore d’autres exemples

A man holding a tennis racquet on a tennis court.

Two pizzas sitting on top of a stove top oven

A group of young people playing a game of Frisbee



A man flying through the air while riding a snowboard

Apprentissage ou généralisation?

Human captions from the training set

A man riding a wave on top of a surfboard.

A man riding a wave on top of a surfboard.

A man riding a wave on top of a surfboard.

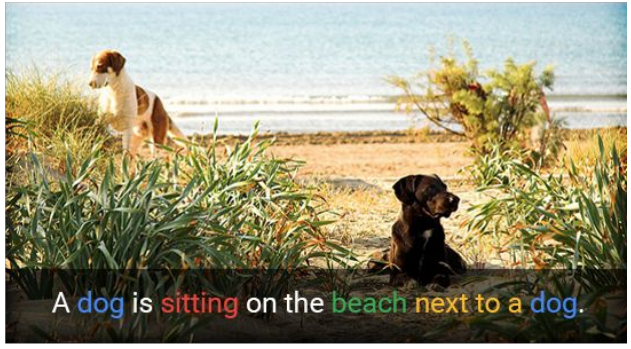
Automatically captioned

A man riding a wave on top of a surfboard.

… un peu des deux!

Human captions from the training set

Automatically captioned



Google

Mais ça ne marche pas toujours...

Human: A blue and black dress… no! A white and gold dress.

NIC: A close-up of a vase with flowers.

… il parait qu’on se ressemble: qu’en pense NIC?

Humain prenant la photo: Deux frères parlant d’intelligence artificielle.

NIC: A couple of men standing next to each other.

Conclusion

Et maintenant?

- L’apprentissage automatique et les réseaux de neurones profonds:
  - Un pas vers l’intelligence artificielle
  - De meilleures produits et services (Facebook, Google, Amazon, etc)
- Très bientôt:
  - Des progrès dans les domaines de la santé et l’éducation
- Recherche actuelle:
  - Apprendre à programmer
  - Apprendre à mieux générer du contenu (images, textes, vidéos, etc)
  - Apprendre sans supervision directe
  - Les voitures sans chauffeur et la robotique
  - Etc!
- L’université de Montréal: plaque tournante de l’intelligence artificielle!

Et à Google?

- Notre équipe, “Google Brain”: un peu comme un labo académique...
  - … mais avec aussi des produits à la clé.
- Une autre équipe cousine, à Londres: Google DeepMind.
- Beaucoup de compétition (Facebook, OpenAI, Microsoft, Amazon, etc) pour attirer les meilleurs!
- Notre CEO s’adressant aux employés de Google:
  - “Google is a machine learning company”
  - “Si vous n’utilisez pas de machine learning dans votre travail actuellement, demandez-vous pourquoi!”
  - Des milliers d’employés prennent des cours de machine learning tous les jours actuellement.

Google Brain Residency Program

- Programme d’immersion d’un an dans notre labo de recherche en apprentissage profond.
- Détails du programme disponibles ici: **g.co/brainresidency**
- Prérequis: un diplôme (BSc/MSc/PhD) en sciences/ingénierie
- Date finale pour appliquer: 13 janvier 2017
- Date du début de la résidence: juillet 2017.

Merci!

(environ1986)

Google