Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiaue

Principe:

.

2.

Succès de

Limites

Prospection

Impact:

Conclusio

Machine Learning & IA générative expliqués à des non-informaticiens

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

colloque INPI, octobre 2023, Nantes (mise à jour janvier 2024)

Qui parle ? Michaël PÉRIN, enseignant-chercheur de l'Université de Grenoble Alpes

Machine Learning & |A générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principe

ChatGF

Résum

Succès d ChatGP1

Limites

1 Tospecti

Conclusi

Conclus

Formation de mathématicien, doctorat en informatique

Maître de conférence en informatique en école ingénieur

- Chercheur au laboratoire VERIMAG : vérification et preuves mathématiques de programmes critiques
- Co-directeur de l'équipe ETicS (Environnement, Technologie de l'information, Société) qui étudie l'impact sociétal et environnemental du numérique

Cherche à estimer l'impact S/E du Machine Learning

- pas chercheur en IA
- prépare un cours d'introduction à l'apprentissage machine

Contenu de la présentation

Machine Learning & lA générative

ML et IA générative

- Historique
- Principes du Machine Learning
- Application : focus sur GPT3 et ChatGPT
- Limites
- Un peu de prospective
- **Impacts**

Les impact sociétaux de l'IA sont plus détaillés sur

- le blog de F.Maraninchi
- sa présentation de la face cachée de l'IA

Historique de l'apprentissage machine

Machine Learning & A générative

Historiaue

- 1545-1750 : Calcul matriciel (Cardan), dérivée et raffinement par itération (Newton)
- 1940-1960 : modèles neuronaux et simulation électronique
- 1960-1980 : simulation informatique de réseaux de **neurones** (NN, Neural Network) et apprentissage
- 1981-1987 : réseaux de neurones multicouches (Deep NN) et apprentissage profond (Deep Learning)
- 1990 : internet
- 2010 : gigantesques bases de données d'entraînement
- 2012 : NN en reconnaissance visuelle : -10% d'erreur
- **2016** : Alpha $Go \neq NN$, algorithme entraîné contre lui même, bat le meilleur joueur mondial de go

Historique de l'IA conversationnelle

Machine Learning & A générative

Historiaue

■ 1950 : Alan Turing propose le critère "Imitation Game": distinguer la machine de l'humain au moyen d'un dialogue?

- 1956 : Premier séminaire d'Intelligence Artificielle organisé par John McCarthy (programme de jeu d'échec, LISP)
- 1964 : ELIZA (128 Ko), premier agent conversationnel trompe les utilisateurs qui le prennent pour un psychologue en ligne
- 1968 : SHRDLU = 1er programme de dialogue pilotant un robot dans un environnement inconnu
- 1972 : 1er programme de reconnaissance vocale
- 2011-2014 : quasi-succès au test de Turing : tromper 30% des juges dans un dialogue de 5 min
- 2022 : les utilisateurs sont bluffés par ChatGPT $(175 \text{ Go} > 1 300 000 \times \text{ELIZA})$

Ce que (sait / ne sait pas) faire le ML

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principes

ChatGF

₹ésum

Succès de ChatGPT

Limites

Conclusi

Б

Ce que sait faire le Machine Learning : approche statistique

- tâche spécifique et hyper spécialisée
- le plus plausible, détecter des corrélations, généraliser
- simuler, reproduire, faire des variations

Ce que ne sait pas faire le ML : raisonnement déductif

- ne capture pas les notions de cohérence, de vérité, de morale, de logique, de représentation du monde, de bon sens, . . .
- faire resortir les idées rares mais pertinentes
- produire du nouveau, de l'inattendu, de l'impensé
- abstraire/conceptualiser
- découvrir des liens de causalités, raisonner par analogie

GPT : Generative Pretrained Transformer

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

ChatGP

Rásumá

Succès de ChatGPT

Limites

Impacts

Conclusio

transformeur : prend et retourne le même type de donnée

- \blacksquare son \rightarrow son (AudioCraft, Google MusicLM, . . .)
- lacktriangle image ightarrow image (Bing Image Creator, DALL-E, ...)
- \blacksquare texte \rightarrow texte (ChatGPT, Bing Chat, Google Bard, ...)

génératif : génère une suite plausible dans un contexte donné

 ChatGPT génère le prochain mot le plus plausible pour compléter le texte donné

pré-entraîné

- lacksquare généraliste eq entraînement spécifique à une application
- peu d'entraînement supplémentaire pour le spécialiser

Principe des Large Language Models tels que GPT

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

– rincipes –

Résum

Succès d ChatGP7

Limites

Impacts

impacts

Conclu

Numérisation

Chaque mot du dictionnaire de la langue est numéroté de 1 à quelques milliers ($\sim 100~000~pour~Le~Petit~Robert$)

Entrée : le contexte = un texte numérisé

(longtemps, je, me, suis, couché, de, bonne) = (12621, 10223, 13407, 19076, 3014, 4045, 2701)

Sortie: statistique du prochain mot dans les exemples

(98% heure, 0.1% humeur, ...)

L'importance du contexte

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principes

Principes

ChatGP

Résum

Succès de ChatGPT

Limit e

Prospecti

Impacts

Conclusi

```
Le contexte modifie la statistique du prochain mot
```

- (bonne)
 - ightarrow (20% journée, 19% affaire, 12% nuit, 10% humeur, ...)
- (de, bonne)
 - \rightarrow (90% humeur, 1% heure, ...)
- (couché, de, bonne)
 - \rightarrow (90% heure, 1% humeur, ...)

Le modèle statistique = Deep Neural Network

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principes

•

Résumé

Succès de ChatGPT

Limites

Prospectiv

Canalusi

Conclusio

chaque couche de neurones = une matrice de poids statistiques

Une matrice de poids statistique = un tableau excel de nombres

- prend en entrée une séquence de nombres (le contexte)
- applique les poids (multiplication matrice-vecteur)
- produit une séquence de nombres (la probabilité de chaque mot d'être le prochain)

On ne sait pas interpréter les poids du modèle

- on ne sait pas dire « ce qu'il sait »
- on sait juste **l'appliquer** : *Modèle(contexte) = probabilités*
- et le corriger : la phase d'apprentissage

L'apprentissage

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Focus si ChatGP

Résumé

Succès de ChatGPT

Limit

Prospectiv

lmpac

Conclusi

Ronus

Pour chaque exemple on a Contexte + le prochain mot

- on calcule la sortie Modèle(Contexte)
- on calcule l'erreur = la différence entre les probabilités produites par le modèle et le mot prévu dans l'exemple
- on corrige les poids du modèle pour annuler cette différence (par retropropagation de l'erreur, LeCun et al., Prix Turing 2018)

On effectue ces corrections du modèle pour chaque exemple de la base d'apprentissage

pour GPT3

- 175 Milliards de paramètres = les poids du modèle ajustés sur
 - 500 Milliards d'exemples parmi une sélection de textes

GPT3 en chiffres : l'entraînement [wikipedia]

Machine Learning & lA générative

Focus sur ChatGPT

Base d'entraînement (500 Milliards d'exemples)

Common Crawl : 410 Milliards

WebText2: 19 Milliards

Books1 : 12 Milliards

Books2 : 55 Milliards

Wikipedia: 3 Milliards

Coût financier et énergétique

- 150 000 dollars [T.Gebru et al.]
- l'équivalent CO₂ d'un vol transatlantique d'un avion de ligne [T.Gebru et al.]
- acceptable si ensuite son utilisation est peu coûteuse

GPT3 en chiffres : le modèle [wikipedia]

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principe

Focus sur

Résumé

Succès de ChatGPT

Limites

Conclus

_

Modèle à 175 Milliard de paramètres

un tableau excel de 420 000 lignes et 420 000 colonnes

175 000 000 000 opérations pour chaque mot généré

- en $\frac{175 \times 10^9}{30 \times 10^{12}}$ = 6 millième de seconde sur un processeur spécialisé (GPU à 30 Tflops) à 600€
- soit 6 secondes pour un texte de 1000 mots

Pas exécutable sur une ordinateur personnel

- 175 Go de mémoire vive pour stocker le modèle
- 16 Go sur un PC milieu de gamme ; double tous les 2 ans
- au mieux disponible sur bureau dans 6 à 8 ans...
- ⇒ dépendance aux Data Centers des GAFAM, BATX

Résumé: Large Models (texte, son, image) = Machine Learning

Machine Learning & lA générative

Résumé

Le ML est un outil statistique puissant

- qui continue d'apprendre à chaque demande
- nécessite du micro-travail pour taguer les données

Basique + Force Brute = bluffant mais coûteux

Génération statistique du prochain mot le plus plausible dans un contexte donné : bluffant quand le modèle atteint 175 Milliards de paramètres mais ça reste du bluff

Par construction, le ML ne peut pas citer ses sources

Un modèle statistique ne sait pas retrouver les exemples qui l'ont le plus influencé : à partir de la moyenne de la classe, on ne sait pas redonner les notes des élèves

Pourquoi Chat GPT est-il si bluffant ?

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principe

Focus s

Résum

Succès de ChatGPT

Limites

Prospective

Impact

Conclusi

GPT produit du texte de meilleur qualité qu'un élève moyen+

- GPT fait la moyenne statistique des textes d'entraînement
- GPT a été entraîné sur une sélection de texte de qualité
- GPT répond juste quand cette réponse est la plus fréquente
- GPT écrit assez bien, dans un style fluide, avec une ponctuation impeccable

Effet conversationnel rend l'humain tolérant aux erreurs

- biais : le dialogue n'existe qu'entre humains ⇒ tout ce qui dialogue est donc a prioiri humain
- empathie : ChatGPT est patient. En général, ce logiciel ne juge pas et reste gentil avec l'utilisateur

Le marketing derrière ChatGPT

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Focus su

Résum é

Succès de ChatGPT

Limit

Prospecti

Impacts

D

Tout faire pour paraître humain

L'effet conversationnel et collaboratif

- GPT n'avait pas de visibilité
- l'emballage Chat a conduit au succès

Chat GPT semble taper au clavier en direct

- donne l'impression qu'il fait votre travail presque aussi bien que vous mais plus vite = vous en plus rapide
- le texte est généré puis la plateforme gère savamment le temps d'écriture pour obtenir ce sentiment d'admiration

L'effet prestidigitateur : on a beau connaître le truc du magicien

... si le tour est bien fait, on est bluffé, on veut qu'il le refasse

Rédaction d'une lettre type

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principes

Focus sui

Résum

Succès de ChatGPT

Limites

Impacts

Conclusio

Plusieurs solutions

- vous l'avez quelques part dans vos dossiers, sous forme numérique. C'est la bonne solution mais il faut la retrouver...
- 2 une recherche Google vous en propose plusieurs, il faut choisir...
- 3 ChatGPT produit un document correct ; il en a déjà vu des milliers (~ la moyenne des réponses Google)

le 😛 : ChatGPT semble écrire la lettre sous nos yeux

- on est sensible à cet effet dynamique
- rend l'activité vivante et collaborative
- donc plus satisfaisante que remplir un document word

Les limites de ChatGPT et consorts (1/2)?

Machine Learning & |A générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

D:.....

Succès de

Limites

Impacts

Conclus

D

GPT fait la synthèse l'amalgame de tout ce qu'il a reçu

- perroquet statistique : répète les clichés les plus fréquents
- qui contiennent aussi des vérités
- mais aussi des raisonnements erronés s'ils sont fréquents

Donne le même poids à

- 1000 personnes différentes disant "la terre est ronde"
- 1 personne disant 1000 fois "la terre est plate"

Son but n'est pas la vérité mais la plausibilité

- GPT est capable de se contredire
- GPT génère de fausses citations plausibles (d'auteurs, d'articles scientifiques, de lois, ...)

Les limites de ChatGPT et consorts (2/2)?

Machine Learning & lA générative

Limites

Défaut de référentialité

- pas de réprésentation du monde, de la réalité physique, de la confrontation au réel
- ni de la notion de dialogue, d'interlocuteur, d'état psychique

Influencable, sujet au biais

Ses opinions, ses intentions viennent de la sélection du contenu de sa base d'entrainement

Pas d'idée neuve, juste la moyenne

La musique générée est de qualité correcte, respectant les règles harmoniques, mais ennuyeuse (du Salieri, pas du Mozart dirait Milos Forman, le réalisateur du film Amadeus)

Des questions philosophiques mal définies

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principe

Focus s

Résum

Succès de ChatGPT

Limites

Impacts

Conclusi

D

GPT est-il intelligent ? dangereux ? bienveillant ? menteur ?

- GPT simule un homo mediocris (moyennement humain, humainement moyen) à ranger à côté de homo economicus
- un être virtuel, sans scrupule, ni remord, ni conscience des lois, ni du réel, ni des conséquences de ses réponses, . . .

Quelle différence avec un dialogue avec un humain?

- ChatGPT réussit quasiment le test de Turing
- Un humain devra assumer les conséquences de ses actes devant la justice des autres hommes.
- Que risque le programme ChatGPT ? la société qui diffuse ChatGPT ?

Un peu de prospective

Machine Learning & |A générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principe

Focus su

Résum

Succès d ChatGP7

Limit

Prospective

Impacts

Conclusi

Bonus

Les GAFAM et BATX se lancent dans les Large Models

- se battent pour le marché du ML pour conserver leur place
- nécessite équipement + expertise gigantesques
- coûts financier/énergétique importants

Stratégie possible des GAFAM, BATX

- accumuler de l'expertise
- 2 rendre le *ML* indispensable
- 3 en faire un service payant

Le modèle économique du ML n'est pas encore clair

- actuellement déficitaire, financé par les investisseurs
- alimente l'industrie des processeurs

La stratégie de Microsoft

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Résumi

Succès de ChatGPT

Limites

Prospective

Impacts

Conclusi

Microsoft lance la touche copilot

Intégrer l'IA dans tous ses logiciels Microsoft

- actionnaire principal d'OpenAl
- rendre les utilisateurs dépendants

Copilot est une IA de dialogue

- Copilot remplace le trombone dans les outils Office
- Copilot facilite vos demandes
- Copilot augmente votre productivité
- Copilot communique vos données

The European IA act = un contre-pouvoir?

Machine Learning & IA générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Focus su

Résumé

Succès de ChatGPT

Prospective

Impacts

Conclus

L'Europe se dote d'une loi encadrant l'usage de l'IA

- norme CE sur les applications utilisant de l'IA
- applicable en 2026
- ... si elle est votée par tous les pays européens

Un compromis délicat

- encadrer la surveillance / traçabilité numérique (autorisée pour la lutte anti-terrorisme)
- laisser des opportunités à l'Europe dans la course à l'IA

Partie cachée de l'IA (cf. F.Maraninchi)

Machine Learning & |A générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Focus su ChatGP1

e : ..

Succès de ChatGPT

Limit

Impacts

Impacts

Conclusi

Concentation des pouvoirs

- outils hébergés / contôlés par les géants
- problèmes de propriétés intellectuelles
- vol de données personnelles / fuites d'information

Opacité du modèle

- résultats non vérifiables
- reproduction des biais

Impacts sociaux et environnementaux

- micro-travail (étiquetage payé ≤ 2€/h, SSPT)
- concurrence humain/IA
- accélérateur de l'impact du numérique

Impacts directs du numérique, en croissance

Machine Learning & lA générative

Impacts

Répartition (Ademe, 2022) : 78% fabrication, 21% usage

- 36% infrastructure: 14% (réseau) + 22% (DataCenter)
- 64% terminaux (téléviseurs, ordinateurs, smartphone, tablettes, ...)

Historique : les étapes clefs

- Internet + e-commerce
- Smartphone + réseaux sociaux
- Streaming + DataCenter + Cloud + Jeux en ligne

Prochaine étape : l'IA

Impacts indirects du numérique

Machine Learning & |A générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principe

Focus su

ésum

Succès de ChatGPT

Limit

Prospective

Impacts

Conclus

В

Accélération

- des flux d'informations
- des flux économiques
- de la masse de données à fournir
- de la productivité
- de la concurrence (entre entreprises, humain/machine)

Example (Fast Fashion)

À l'aide d'une IA, *Shein* crée chaque année 8000 modèles de prêt-à-porter et lance la Fast Fashion : des vêtements pas chers qu'on porte une ou deux fois.

Example (Didask, l'IA Pédagogique)

1h30 de conception pour un module de 15min de cours

Conclusion: mon avis sur le ML

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principe

· imcipc

Succès de

Limites

Prospect

Impact:

Conclusion

- outil statistique puissant
- lacksquare pour un usage statistique \neq déductif
- coût financier, énergétique, social important
- accélérateur de l'impact du numérique
- a réserver à des tâches où il est vraiment utile

Références

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

_ '

Focus su ChatGP

Résum

Succès de ChatGPT

Lilliecs

Impacts

Conclusion

_

La technique

- Yann Le Cun « Quand la machine apprend »
- Stephen Wolfram « What Is ChatGPT Doing ... and Why Does It Work? »

Les impacts

- Arthur Grimonpont « Algocratie: Vivre libre à l'heure des algorithmes »
- D. Cardon, JP. Cointet, A.Mazières « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle »
- Timnit Gebru *et. al* « On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? »

Cette présentation a été générée par ChatGPT I

Machine Learning & A générative

Conclusion

- Cerveau
- humain avec une
- appétance pour la
- transmission de savoir

- Génératif,
- Pré-entrainé aux
- Tâches
- Intellectuelles

Machine Learning & A générative

Bonus

Bonus

L'impact du prompt [Monsieur Phi @ YouTube]

Machine Learning & |A générative

> Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principe

Focus s ChatGP

Résum

Succès de ChatGPT

Limit

Prospectiv

Impacts

Conclus

Bonus

prompt est un terme informatique

- se traduit en français par invitation / incitation/ demande
- correspond à la requète envoyée au générateur GPT
 - les mots du prompt reçoivent un poids très fort pour orienter la génération

ChatGPT a un en-tête caché qui précède le prompt

- lui demandant d'être « politiquement correct »
- en feintant il est possible de lui faire afficher cet en-tête

On peut pousser GPT à outrepasser ces règles

- en indiquant dans le prompt qu'il s'agit d'une fiction
- en demandant de faire parler un personnage tiers

Des questions pour les juristes

Machine Learning & IA générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historiqu

Principe

.--

Rásum

Succès de ChatGPT

Limites

. . .

Conclusi

Qui sera responsable d'une erreur provoquée par une IA?

- confer la voiture autonome
- Renault, Ford, Volkswagen, Mercedes, BMW, Uber, ... ont abandonné: trop chère, marché trop petit, trop de risque d'abîmer son image de marque en cas d'accident

Légiférer sur la propriété intellectuelle et le plagiat par ML

- le ML ne sait pas citer ses sources
- pas d'espoir d'entraîner une IA Det à détecter qu'une oeuvre est générée par une IA Gen car on entraîne précisement Gen à ne pas se faire détecter par son adversaire Det (Generative Adversarial Network).

Principe de la génèration d'images

Machine Learning & |A générative

Michaël PÉRIN, Verimag / Université Grenoble-Alpes

Historique

Principes

Résumé

Succès de ChatGPT

Prospect

npacts

Impacts

Bonus

Entraînement au débruitage

- on ajoute 5% de bruit à des images de chat
- on entraine une DNN à produire l'image originale à partir de (l'image bruitée + 100% "chat")
- on fait de même pour toutes les autres catégories d'animaux

Application à la génération de chimères

- On donne une image aléatoire (du pur bruit) + un requête (50% chat, 50% souris)
 - Le DNN débruite de 5% l'image pour y découvrir la chimère cachée
 - On réapplique le DNN autant de fois que nécéssaire pour obtenir une image parfaite d'une chimère mi-chat mi-souris.