

INFO-F308 - Printemps des Sciences '18

Fiction ?

Introduction

Réalité ou fiction ? Volontairement ou inconsciemment, l'Homme a toujours observé, étudié, rêvé, imaginé, fabriqué, fantasmé, inventé. À notre époque et à chaque seconde, des milliards d'informations transitent de par le monde à la vitesse de la lumière. L'information se présente à nous, mise en forme, modelée, ciselée suivant nos goûts ou opinions : actualité, histoire, faits divers, société, politique, économie, santé, sciences nous sont servis sur un plateau d'argent. On like, on s'émeut, on partage. Et pourtant, jamais le besoin de vérifier l'exactitude de ces informations ne s'est révélé aussi crucial : comment séparer le vrai du faux ? Est-il vrai que le 17 août, Mars apparaîtra, dans le ciel, aussi grosse que la Lune ? Que l'industrie pharmaceutique nous cache le remède anti-cancer ? Que la vaccination est à l'origine de tous nos maux ?

Mais manipuler la réalité, la transformer peut aussi la magnifier : qui n'a jamais rêvé devant ces somptueuses reconstitutions historiques, ces mondes mythiques dans lesquels de flamboyants dragons côtoient de valeureux gladiateurs, ces récits de voyages interstellaires à couper le souffle ? Où donc se situe la frontière entre la science et la fiction, entre le plausible et le farfelu ? Que diraient Léonard de Vinci ou Jules Vernes de notre monde ? Notre réalité a-t-elle dépassé leur fiction ? Cette réalité, elle-même, existe-t-elle puisqu'elle n'est que la perception, la vision que nous avons de ce qui nous entoure ? La réalité ne serait-elle qu'une fiction ?

1. Fake news - Identification vérité alternative / faits alternatifs



Présentation du sujet

Les réseaux sociaux informatisés (e.g. Facebook, Twitter, etc...) permettent très facilement à ses utilisateurs de publier des informations et de les partager avec leurs contacts. Grâce aux développements technologiques, ce partage peut se faire à partir de tout dispositif connecté à ce réseau social (e.g. smartphone, tablette, ordinateur). De plus, les auteurs de ces informations ne sont ni obligés de vérifier l'authenticité de ces informations, ni de fournir des sources fiables. La combinaison de ces deux aspects permet de manière très simple de diffuser des fausses informations, également appelées faits alternatifs/vérité alternative.

D'un autre côté, le format et la disponibilité croissante de ce type d'information pourraient potentiellement permettre leur traitement automatique ou un apprentissage automatique des éléments caractéristiques des vraies/fausses nouvelles.

Travail à réaliser

Nous disposons de plusieurs jeux de données (en Anglais) constitués de messages publiés sur ces réseaux sociaux, ainsi que des métadonnées sur le message et la classification du message (i.e. vrai/faux)

Avec ces données, nous voudrions réaliser un système (mobile ou non) qui, après une phase d'apprentissage sur les données disponibles, réaliserait une classification des nouveaux messages postés, afin d'aider son utilisateur à mieux interpréter des nouvelles informations. Afin de rendre le système plus accessible pour le visiteur du Printemps de Sciences, nous vous suggérons une implémentation ludique (cf. [Fake news game](#)).

Jeux de données

- <https://github.com/BuzzFeedNews/2016-10-facebook-fact-check>
- <https://www.kaggle.com/mrisdal/fake-news>
- <http://www.fakenewschallenge.org/>
- <http://compsocial.github.io/CREDBANK-data/>
- <http://sbp-brims.org/2017/challenge/> (Data Set section)

Littérature scientifique

- [1] Wang, W. Y. (2017). " Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. arXiv preprint arXiv:1705.00648.
- [2] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4.
- [3] Rubin, V. L., Conroy, N. J., Chen, Y., & Cornwell, S. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In *Proceedings of NAACL-HLT* (pp. 7-17).
- [4] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). Fake review detection: Classification and analysis of real and pseudo reviews. Technical Report UIC-CS-2013-03, University of Illinois at Chicago, Tech. Rep.
- [5] Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2921-2927).

Contacts

- Jacopo De Stefani (Jacopo.De.Stefani@ulb.ac.be)

2. Miroir, mon beau miroir... - Détection de la manipulation d'une image



Présentation du sujet

Durant les 30 dernières années, grâce à une évolution technologique continue, le domaine de la photographie a pu connaître une véritable révolution. Les limites de la pellicule, l'ancien format de stockage analogique ont progressivement été remplacés par un format numérique. De plus, la miniaturisation des capteurs a permis l'intégration des caméras dans les dispositifs les plus divers, des téléphones portables, jusqu'aux drones. Ceci a mené à une augmentation sans précédent du volume des images produites par ces dispositifs ce qui rend de plus en plus difficile la vérification l'authenticité de ces images.

L'amélioration continue des logiciels de traitement d'images rend la manipulation des images très facile et à la portée de tous. La combinaison de ces facteurs rend toujours plus difficile distinguer entre une image réelle et une image modifiée ad hoc.

Travail à réaliser

Nous disposons de plusieurs jeux de données constitués d'images manipulées artificiellement (avec l'introduction/suppression d'éléments), ainsi que les images réelles correspondantes.

Avec ces données, nous voudrions réaliser un système (mobile ou non) qui, après une phase d'apprentissage sur les données disponibles, permettrait, de manière automatique, l'analyse d'une image donnée, et la détection d'une manipulation sur l'image même.

Jeux de données

- <https://www5.cs.fau.de/research/data/image-manipulation/>
- <https://revealproject.eu/the-deutsche-welle-image-forensics-dataset/>
- <http://forensics.idealtest.org/>

Littérature scientifique

- [1] Dadkhah, S., Manaf, A. A., Hori, Y., Hassanien, A. E., & Sadeghi, S. (2014). An effective SVD-based image tampering detection and self-recovery using active watermarking. *Signal Processing: Image Communication*, 29(10), 1197-1210.
- [2] Christlein, V., Riess, C., Jordan, J., Riess, C., & Angelopoulou, E. (2012). An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on information forensics and security*, 7(6), 1841-1854.
- [3] Dirik, A. E., & Memon, N. (2009, November). Image tamper detection based on demosaicing artifacts. In *Image Processing (ICIP), 2009 16th IEEE International Conference on* (pp. 1497-1500). IEEE.
- [4] Mishra, M., & Adhikary, F. (2013). Digital image tamper detection techniques-a comprehensive study. *arXiv preprint arXiv:1306.6737*.
- [5] Verdoliva, L., Cozzolino, D., & Poggi, G. (2014, December). A feature-based approach for image tampering detection and localization. In *Information Forensics and Security (WIFS), 2014 IEEE International Workshop on* (pp. 149-154). IEEE.

Contacts

- Jacopo De Stefani (Jacopo.De.Stefani@ulb.ac.be)
- Matthieu Defrance (Matthieu.DC.Defrance@ulb.ac.be)

3. Simulation

Présentation

Avec l'émergence de l'informatique et l'explosion de la puissance de calcul associée, la simulation informatique est devenue un outil incontournable dans des domaines tels que la physique, la chimie, la météorologie ou encore les sciences humaines. Elle permet par exemple de donner des perspectives sur le développement de systèmes complexes.

Travail à réaliser

Nous nous intéresserons plus particulièrement au paradigme multi-agents et à son utilité dans la simulation de modèles de sociétés. Un système multi-agents (SMA) est un système composé d'un ensemble d'agents, situés dans un certain environnement et interagissant selon certaines relations. Un agent est une entité caractérisée par le fait qu'elle est, au moins partiellement, autonome [1].

Pour ce faire, nous nous concentrerons sur la modélisation d'une fourmilière. L'objectif étant de reproduire l'expérience décrite dans [2] à l'aide du langage Prolog.

Références

[1] https://fr.wikipedia.org/wiki/Système_multi-agents

[2] Drogoul, A., & Ferber, J. (1992, July). Multi-agent simulation as a tool for modeling societies: Application to social differentiation in ant colonies. In *European Workshop on Modelling Autonomous Agents in a Multi-Agent World* (pp. 2-23). Springer, Berlin, Heidelberg.

Jeux de données

- <http://www.lirmm.fr/~ferber/ProgAgent/index.html>

Littérature scientifique

- [1] Drogoul, A., & Ferber, J. (1992, July). Multi-agent simulation as a tool for modeling societies: Application to social differentiation in ant colonies. In *European Workshop on Modelling Autonomous Agents in a Multi-Agent World* (pp. 2-23). Springer, Berlin, Heidelberg.
- [2] Michel, F. (2004). Formalisme, outils et éléments méthodologiques pour la modélisation et la simulation multi-agents. *These de doctorat, Université Montpellier II*, 129.
- [3] Ferber, J. (1995). *Les systèmes multi-agents: vers une intelligence collective*. InterEditions.

Contacts

- Guillaume Duvillié: gduvilli@ulb.ac.be

4. Déterminisme - Nombres pseudo-aléatoires, entropie, compression, etc.



Présentation du sujet

Générer un nombre aléatoire en python? Rien de plus facile!

```
import random  
print(random.random())
```

Hors, les ordinateurs sont des machines déterministes. Contradiction!

Que veut dire aléatoire? Qu'est-ce qu'un nombre aléatoire? Comment peut-on mesurer la stochasticité ("aléatoirité") d'un nombre? Peut-on la mesurer? Quelle quantité d'information est contenue dans un livre, un morceau de musique, une photo, un film? Nous pouvons compresser cette information (formats PDF, MP3, JPEG, H.264). Quelles sont les limites de cette compression? Peut-on créer un système complexe à partir de règles simples? Toutes ces questions sont liées. Nous tenterons d'y répondre.

Travail à réaliser

Le travail comporte une partie théorique qui vous introduira au concept de décidabilité vu au cours "Calculabilité & Complexité" du Master 1 (INFOF408) ainsi qu'une partie implémentation

qui vous permettra d'expérimenter avec les générateurs de nombres pseudo-aléatoires, les tests de stochasticité, les algorithmes de compression, les fractales, le mélange de paquets de cartes, les automates cellulaires, le jeu de la vie de Conway, etc.

Références

1. [VSauce - What is random?](#)
2. [Veritasium - What is NOT random?](#)
3. [Fisher–Yates shuffle](#)
4. [Mandelbrot set](#)

Jeux de données

- <https://random.org>
- <https://www.fourmilab.ch/hotbits>
- <https://randomizer.org>

Littérature scientifique

1. Sipser, M. (2012). *Introduction to the Theory of Computation*. Cengage Learning, Chapters 3 & 4.
2. Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1), 3-30.
3. Marsaglia, G., & Tsang, W. W. (2002). Some difficult-to-pass tests of randomness. *Journal of Statistical Software*, 7(3), 1-9.
4. Ziv, J., & Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Transactions on information theory*, 23(3), 337-343.
5. Gage, D., Laub, E., & McGarry, B. (2005, October). Cellular automata: Is rule 30 random. In *Proceedings of the Midwest NKS Conference, Indiana University*.
6. Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223(4), 120-123.

Contacts

- Aurélien Ooms (aureooms@ulb.ac.be)