Chair Data Science and Artificial Intelligence for Digitalized Industry and Services

Internship project

Subject

Exploring large pre-trained representation models for textual Anomaly Detection

Possibility to continue as a PhD candidate

YES (Funding confirmed)

About the chair

The Chair Data Science and Artificial Intelligence for Digitalized Industry and Services (DSAIDIS), lead by Florence d'Alché-Buc, a Professor in the department Image, Data, Signal of Telecom Paris, unites five industrial partners: Airbus Defence & Space, Engie, Idemia, Safran et Valeo. It's general objective is to develop, in collaboration with the partners, teaching and research of the international level.

Its four principal research directions are:

1. Building predictive analytics on time series and data streams.
2. Exploiting large scale, heterogeneous, partially labeled data.
3. Machine Learning for trusted and robust decision.
4. Learning through interactions with environment.

Description of the internship

Supervision

Matthieu Labeau (https://www.telecom-paris.fr/matthieu-labeau)

Location and dates of the internship

Address : Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau

Date of the beginning of the internship : Early 2024

Team where the thesis will be written

Department IDS, Team Signal, Statistique et Apprentissage (S2A)

Keywords

Anomaly detection, Semi- and unsupervised learning, Transfer learning, Text representation learning.

Detailed subject

Anomaly Detection (AD) consists in detecting observations that deviate considerably from some concept of normality, which is mostly done by computing an anomaly score indicating its degree of anomalousness [1]. AD typically lacks labeled anomalous data. When available, those examples are not enough to characterize this notion of anomalousness. Then, the idea is to learn a model of normality from that data, identifying deviations as anomalies. In Natural Language Processing (NLP), weak signals are very difficult to capture, and even the modern, best performing Large Language Models struggle on rare words and knowledge [2]. Directly applying existing AD models to textual data is intricate. Among these difficulties are: the reliance of deep textual representation models on

pre-training on large quantities of data, containing prior knowledge which impact is difficult to gauge; the lack of interpretability of these models and of the features of their representations; and the lack of information about what an anomaly is, tied with a lack of real-world labeled data, as the vast majority of work on AD in NLP re-purposes classification datasets. Depending on the source of the data, the chosen method of representation and the AD model employed, many phenomenon could be detected (rarity, being out-of-domain, noise, semantic variation) which may or may not be related to the intended application, and the interpretability of any textual AD model and its decisions is severely limited.

The main purpose of this internship is to explore the very large array of possible combinations in the AD field (comprising model, loss, and when explicit, regularization method) with text representational models, besides few recent works [3,4]; a key factor to consider being the transfer of linguistic knowledge from large pre-trained models. We will need to clearly control assumptions made by the model on anomalies, which is seldom done in the current literature [5]. We can also explore the use of existing, but less popular textual representation methods, such as spherical representations [6], and on the best way to include knowledge transfer in a textual AD framework for possible domain adaptation [7]. Secondly, to make up for the lack of general purpose textual AD datasets, and in order to extend evaluation, we plan to re-purpose existing classification datasets [8] to be much closer to the assumptions of the task. This first investigation, aiming at providing a more holistic perspective of textual AD, has the purpose be continued into a PhD thesis, funded by the ANR project CFTextAD.

Candidate profile

Student having master 2 research

- Practical and theoretical understanding of Machine Learning
- A previous experience or interest in Natural Language Processing
- Good level of programming (Python)
- Good command of English

Application

To send on matthieu.labeau@telecom-paris.fr:

- Curriculum Vitae
- Personalized motivation letter that explains interest of the candidate in the subject (can be directly in the body of the email)
- Grade reports for recent years
- Contact of a person willing to give recommendation

References

[1] Lukas Ruff et al. “A Unifying Review of Deep and Shallow Anomaly Detection”. In: Proceedings of the IEEE (2021). url: https://doi.org/10.11092Fjproc.2021.3052449.

[2] Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. Large Language Models Struggle to Learn Long-Tail Knowledge. 2022. arXiv: 2211.08411 [cs.CL].

[3] Lukas Ruff, Yury Zemlyanskiy, Robert Vandermeulen, Thomas Schnake, and Marius Kloft. “Self-Attentive, Multi-Context One-Class Classification for Unsupervised Anomaly Detection on Text”. In: ACL. 2019. url: https://aclanthology.org/P19-1398.

[4] Andrei Manolache, Florin Brad, and Elena Burceanu. “DATE: Detecting Anomalies in Text via Self-Supervision of Transformers”. In: NAACL-HLT. Online: Association for Computational Linguistics, 2021. url: https://aclanthology.org/2021.naacl-main.25.

[5] Udit Arora, William Huang, and He He. “Types of Out-of-Distribution Texts and How to Detect Them”. In: EMNLP. 2021. url: https://aclanthology.org/2021.emnlp-main.835.

[6] Yu Meng, Jiaxin Huang, Guangyuan Wang, Chao Zhang, Honglei Zhuang, Lance Kaplan, and Jiawei Han. “Spherical Text Embedding”. In: Advances in Neural Information Processing Systems. 2019. url: https://proceedings.neurips.cc/paper/2019/file/043ab21fc5a1607b381ac3896176dac6-Paper.pdf.

[7] Keyang Xu, Tongzheng Ren, Shikun Zhang, Yihao Feng, and Caiming Xiong. “Unsupervised Out-of-Domain Detection via Pre-trained Transformers”. In: ACL. 2021. url: https://aclanthology.org/2021.acl-long.85.

[8] N. Gupta, S. Bohra, Y. Prabhu, S. Purohit, and M. Varma. “Generalized Zero-Shot Extreme Multi-label Learning”. In: ACM SIGKDD. 2021.