Post-doctoral position : Deep Learning for opinion mining in human testimonials related to industrial accident

Location : LITIS lab., University of Rouen Normandy, Rouen, France
Duration : 18-months, starting as soon as possible
Salary : ~2100€ / month (before income tax), including social security coverage in line with French regulations
Applications : open from 01/09/2021 to 31/12/2021

The Machine Learning team at the LITIS laboratory, the computer science laboratory of the University of Rouen Normandy, is looking for a post-doctoral researcher on a 18-months contract, starting as soon as possible. The position will be financed by the ANR research project CATCH (french acronym for "Automatic Understanding of Human Sensors Testimonials"), which involves the R&D center of the company Saagie[1], specialized in B2B DataOps solutions, Atmo Normandie[2], one of the approved French air quality monitoring associations and LITIS.

Keywords

Deep learning – Natural Langage Processing – Sentiment Analysis / Opinion Mining

Scientific context

The ambition of the CATCH project is to propose artificial intelligence and deep learning tools to take into account and automatically exploit the multitude of human testimonies related to an industrial accident and its consequences on the environment and health. By involving the population in the collection and analysis of data, particularly through social networks, and by providing effective means for interpreting this data, the proposed solution should contribute to providing answers to the worrying problem of industrial accidents and their consequences.

The overall objective is first to draw up a precise cartography of the nuisances in order to follow the propagation and the evolution of the phenomena in time, and then to analyze and characterize the sentiment of the population and its evolution throughout the crisis. To do so, we intend to exploit testimonials collected on the ODO platform[3] of Atmo Normandie, which combines these testimonies with geographical information, in conjonction with data extracted from the micro-blogging platform Twitter. Since these data are primarily texts, state-of-the-art approaches from the Natural Language Processing (NLP) field are favored, in particular, self-supervised deep learning methods such as Transformers[4] that are known to be the most performant today for a wide range of NLP tasks[5].

Research goals

---

1   https://www.saagie.com/
2   http://www.atmonormandie.fr/
3   https://www.atmo-odo.fr/
4   Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
5   https://super.gluebenchmark.com/leaderboard

The objective of this research work is twofold:
1. The automatic generation of a map of nuisances around the site of an industrial incident to monitor the propagation and the evolution of the phenomena over time.
2. The automatic recognition of the population's perception and its evolution throughout the crisis.

Related to these tasks, the post-doctoral researcher will be in charge of proposing solution for:
- extracting and linking twitter data with testimonials from the ODO dataset, which is fully labelled and associates textual testimonies with geographical data. The interest in establishing this link is to be able to enrich the data from the ODO platform to refine the mapping of nuisances in real time. This could be achieved for example, by using pseudo-labelling techniques[6] or Constrative Representation Learning methods which have recently been applied to text data[7].
- detecting in all the testimonials collected from Twitter or from the ODO platform, the presence (or absence) of several pre-identified emotions (e.g. surprise, fear, anger, sadness, disgust, etc.), several of which can be expressed at the same time.

This research work will therefore involve being familiar with the state-of-the-art NLP deep learning methods and in particular with their applications to sentiment analysis and opinion mining tasks. It will also require experience with the use and exploitation of data from Twitter in a data science context.

Application

The successful applicant will:
- possess or be in the process of obtaining a Ph.D. in computer science or applied mathematics, with a focus on machine learning or data mining.
- have strong programming skills (Java, Python, etc.) and in-depth understanding of statistics and machine learning.
- have already worked with deep learning architecture dedicated to texts (RNNs, Transformers, etc.) and/or images (CNNs, FCNs, GANs).
- have a productive publication record.

Your application should include:
- curriculum vitae
- statement of past research accomplishments, career goal and how this position will help you achieve your goals
- two representative publications
- contact information for at least two references

Contact

Application must be sent to :
- Simon BERNARD, University of Rouen Normandy, simon.bernard@univ-rouen.fr
- Clément CHATELAIN, INSA Rouen Normandy, clement.chatelain@insa-rouen.fr
- Alexandre PAUCHET, INSA Rouen Normandy, alexandre.pauchet@insa-rouen.fr

---

6   D.H Lee (2013). Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. ICML 2013 Workshop: Challenges in Representation Learning (WREPL).
7   H. Fang and S. Wang and M. Zhou and J. Ding and P. Xie (2020). CERT: Contrastive Self-supervised Learning for Language Understanding. CoRR. https://arxiv.org/abs/2005.12766