AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

ROADMAP ON AI TECHNOLOGIES & APPLICATIONS FOR THE MEDIA INDUSTRY

SECTION: "NLP AND CONVERSATIONAL AGENTS"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Authors	**Adrian Popescu** (French Alternative Energies and Atomic Energy Commission - CEA) **Julien Tourille** (French Alternative Energies and Atomic Energy Commission - CEA)

This report is part of the deliverable D2.3 - *"AI technologies and applications in media: State of Play, Foresight, and Research Directions"* of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

Copyright

NLP and conversational agents

Current status

Natural Language Processing (NLP) research has benefited from the introduction of deep learning models in the early 2010s. New methods and algorithms have unlocked possibilities and increased dramatically the performance on every task in the field over the last decade (see Figure 1). Specifically, the transfer learning paradigm has become the norm in the field. Large language models are trained on raw textual content scraped from the web and are fine-tuned on target tasks (e.g. named entity recognition or relation extraction). Among other model architectures that have been developed, the ***transformer architecture***[1] is the most versatile.

This new paradigm has allowed performance increases in various tasks such as named ***entity recognition and linking*** (the task of recognising entity mentions - e.g. the city of Paris - and assigning them a unique identifier - e.g. a Wikipedia URL), ***opinion mining*** (extract someone's opinion on one or several aspects of a topic) and ***argument mining*** (automatic identification and extraction of the structure of inference and reasoning of an argument).

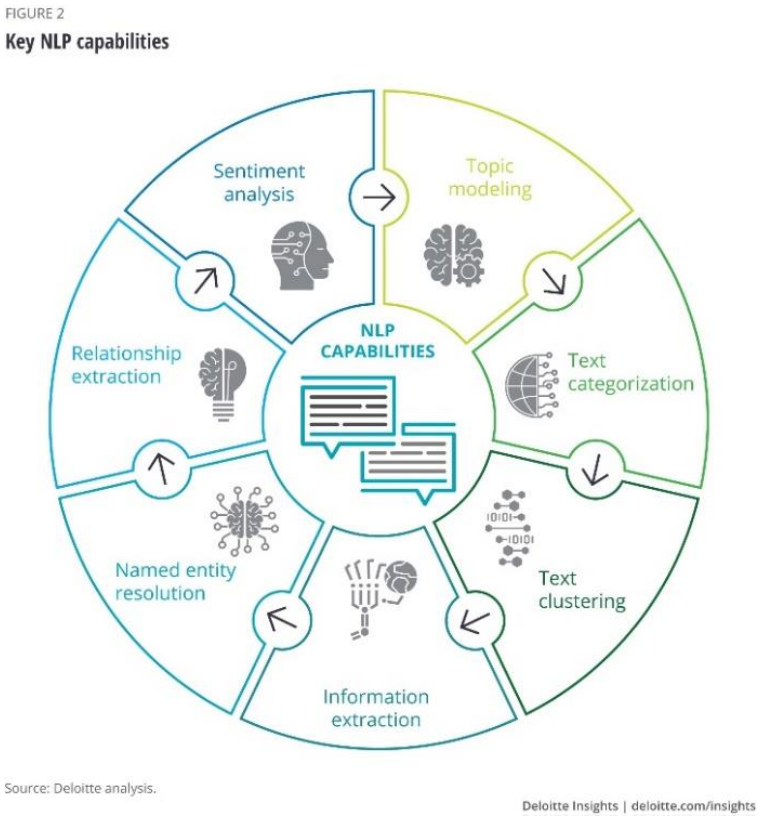

Figure 1: NLP capabilities. Image by Deloitte Insights.[2]

[1] Vaswani et al. (2017). Attention is all you need. In: NIPS

[2] Image from Deloitte Insights: W. D. Eggers, N. Malik, and M. Gracie, Using AI to unleash the power of unstructured government data: https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html (Image source:

Research on **chatbots**, also called *conversational agents*, has also benefited from this new situation. Chabots have become ubiquitous in modern societies. They are applied in various domains such as health, customer service, education and office work. They also have become part of our daily lives with the introduction of voice assistants on our connected devices (e.g. Apple's Siri, Amazon's Alexa or Microsoft's Cortana) (Figure 2).

Although academic research on chatbots has been active for the past 50 years[3], research and application on this topic have witnessed an increase in interest with the development of deep learning methods in the last decade. These new methods have improved the performance of algorithms for user intent and sentiment extraction, which are key elements in chatbot design. We have also observed the emergence of new models for natural language generation, trained on large conversational datasets. They generate accurate answers in natural language to user queries.

Figure 2: AI is listening.[4]

Research challenges

Although there was a huge leap forward over the last decade in NLP, several challenges remain. There are growing concerns around the use of large language models. They tend to ***reproduce societal biases*** that are present in the training corpus and have a ***large carbon footprint*** due to their sizes. More specifically, models such as GPT-3 are trained with huge volumes of Internet data, which are usually produced in the richest countries, in languages with higher linguistic footprint, and by communities with large representation. This results in models that fail to capture the culture of minorities and underrepresented groups and which may eventually

https://www2.deloitte.com/content/dam/insights/us/articles/4815_AI-unstructured-data/figures/4815_Figure2.png)

[3] Weizenbaum et al. (1967). *Contextual understanding by computers*. In: Communication of the ACM

[4] Illustration from XKCD at https://xkcd.com/1807/

discriminate against such groups[5]. In addition, their increased computational needs have a considerable environmental impact (e.g. it is estimated that training OpenAI's GPT-3 model produced the equivalent of 552 metric tons of CO_2, which is the equivalent of driving to the moon and back). Another interesting avenue of research is the integration of knowledge into deep learning models. There are indeed a lot of resources that are manually curated (e.g. UMLS or Wikipedia) and that contain valuable information.

Few-shot learning, i.e. the possibility to train a machine learning model with only a few examples, is also an active area of research (see section on "*Learning with scarce data*"). Being able to train new models rapidly with low annotation costs could improve the diffusion of those models outside the academic world. To understand the size and resources needed to train some of the largest language models, we note that OpenAI's GPT-3, with its 175 billion parameters, was trained with 499 billion tokens and would require $4,600,000 to train - even when utilising the lowest priced cloud GPUs on the market[6].

Concerning the research on conversational agents, Følstad et al.[7] have identified several challenges. While performance of methods and algorithms has improved over the last couple of years, extracting precisely the ***user intent and sentiment*** remain difficult and conversational agents may have breakdowns when dealing with everyday situations. Then, adaptation to the user ***conversational context*** has been identified as one of the principal challenges in the early years and has remained an active area of research. Conversational agents should adapt to the social context and literacy competences of the users. Finally, standardised evaluation of systems and algorithms need to be implemented for chabots. Open and distributed datasets for evaluation have become a key element for sound and reproducible research in NLP for many years.

Societal and media industry drivers

Vignette 1: Multi-lingual document analysis and summarisation for investigative journalism

Adèle is a journalist who writes about the economic situation in developing countries, with particular focus on energy matters. To ensure a good coverage of this highly dynamic topic, she should have swift access to large amounts of news, social media contributions and corporate documents written in different languages. She needs a set of tools which: (1) translates domain-specific documents from potentially under-resourced languages to German, English and Croatian, the reporting languages by Adèle, in a reliable manner; (2) clusters them according to both topics and opinions about these topics to give an overview of the different points of view expressed; (3) mines arguments used in the source documents and evaluates their plausibility; and (4) summarises these documents. Such NLP-based tools would greatly facilitate Adèle's documentation work and would allow her to focus on the creative side of her reporting.

[5] K. Hao, MIT Technology Review, We read the paper that forced Timnit Gebru out of Google. Here's what it says (2020): https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

[6] C. Li, OpenAI's GPT-3 Language Model: A Technical Overview (2020): https://lambdalabs.com/blog/demystifying-gpt-3/

[7] Følstad et al. (2021). *Future directions for chatbot research: an interdisciplinary research agenda*. In: Computing.

Vignette 2: Conversational agent for interactive and personalised audio content delivery in broadcast

Hector is in charge of the media delivery team of a public broadcaster. One of the innovative services of the broadcaster is to offer interactive and personalised audio content to their audience. Hector's team develops a conversational agent which adapts the content served to the user depending on their historical preferences but also on the go, based on questions asked when the content is played. The agent exploits cutting-edge NLP techniques in order to: (1) formalise historical preferences and propose material which mixes new content on topics that were already explored by the users and new topics which are likely to interest them; (2) understand the user's queries and integrate them in the dialogue model in order to be able to modify the content at any moment depending on the interaction with the user; (3) combine content from the broadcaster's archives and from open sources in a coherent material which is served to the user.

Future trends for the media sector

NLP has witnessed a true revolution during the last few years with large language models like GPT-3. NLP is expected to become increasingly mainstream in the media business though applications such as conversational agents and virtual characters, creative writing, robot journalism, interactive storytelling, voice search for image/video/audio, sentiment analysis in social media, voice dubbing, or multi-lingual translation.

On the news content production side, NLP is already in the newsroom, with large news broadcasters having their own research and development divisions. Their principal objective is to devise tools to smooth the editorial process. Among those tools, ***faceted search engines*** help journalists to look for information in databases, archives and on the web. They allow for rapid information gathering to help staff to write and edit news.

There is no extensive use of state-of-the-art NLP yet. But this may change in the future. Several academic research teams are devising ***automatic fact-checking systems*** that are able to tell whether the information is legitimate or spurious. ***Automatic argument mining*** is also actively pursued. Being able to extract the argumentation structure of a source document could help journalists to skim rapidly several sources of information.

Concerning information extraction, new methods for ***event, entity and relation extraction*** could help journalists to rapidly extract relevant documents from large databases. For instance, NLP techniques were already used to skim the large corpus available in the Pandora papers.

On the other side of the spectrum, news consumers could also benefit from the development of NLP. Automatically characterising news topics and opinions could help citizens to broaden their views and counter the effects of news recommendation systems that tend to only suggest content aligned with user's opinions.

All other media industry sectors are also expected to be impacted by the deployment of NLP algorithms. Video games, especially those involving open worlds, may benefit from conversational agents powered by AI to improve the gaming experience and offer a unique story

to each participant. This also opens interesting opportunities for the Metaverse where bots could potentially interact with users' avatars.

The film and TV industry can use new NLP techniques to automatically analyze audience reviews for fast opinion mining surveys or to predict audience engagement and movie success based on script analysis. Assisted multilingual subtitling or voice dubbing are also interesting applications for this sector while automatic or assisted translation in different languages can be a game-changer for the publishing industry.

Goals for next 10 or 20 years

NLP is a large and old academic research domain and has experienced several breakthrough changes over the last two decades. It is therefore difficult to draw objectives for the far future. However, we identified several interesting research avenues for the next few years.

As mentioned in previous sections, a lot of work has been done in representation learning, with the emergence of the transfer learning paradigm. Models have gained in complexity and ingenuity, and helped to improve the state-of-the-art for all tasks in NLP. There is an actual trend in devising more and more complex systems with billions of parameters. However, little research has been done on data centric approaches where the objective is ***to improve the overall representativity*** of the dataset that is fed to the models to increase downhill performance.

Furthermore, although new methods and datasets have been devised over the past couple of years, probing the knowledge that is embedded in these models remains difficult and needs further research. A better understanding of the content of these models could lead to improvement in other NLP areas.

Finally, the emerging topic of ***model hybridation***, i.e. combining existing curated knowledge and unsupervised language model, needs further investigation. Although recent language models manage to capture knowledge in an unsupervised fashion, they fail to learn fine-grained relations within the text.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu