DOI: orcid.org/10.20535/kpisn.2024.1-4.315079
UDC 004.89:303.732.4

Volodymyr Savastiyanov[1*], Mykhailo Stoliar[1]
[1]Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
[*]Corresponding author: vvs.in.ua@gmail.com

SYSTEM APPROACH TO THE COMBINED USE OF LARGE LANGUAGE MODELS AND CLASSICAL MODELS IN FORESIGHT TASKS

Background. Large Language Models (LLMs) and their associated agents have spread wide technology and represent a significant advancement in recent times. These state-of-the-art models expose valuable potential, but they are not devoid of restrictions, inefficiencies, and limits. This article investigates the exploration of these constraints within specific domain areas and prediction problems as examples.
Objective. The article highlights features offered by GPT-based models and compares the conclusions with classical methods of textual data analysis in classification tasks using the prediction methodology as an example. The purpose of the study is to develop a system approach to the combined use of traditional machine learning approaches as a practical alternative to LLMs in foresight tasks using the example of STEEP analysis, which provides an opportunity to obtain valuable information from textual data.
Methods. The study is structured into four segments, each addressing distinct parts: Data Mining, text pre-processing using LLMs, text pre-processing utilizing Natural Language Processing (NLP) methods, and comparative analysis of results. Data Mining includes data collection and data pre-processing stages for train and test observations. For the utilization of LLMs, chains of thought techniques and prompt engineering were used.
Results. Throughout this study, it was acknowledged that the LLMs can be used in combination with classical machine learning methodologies for domain-specific areas in STEEP analysis under Foresight tasks. The outcome revealed a model that was developed significantly faster and with less complexity compared to LLMs such as GPT and Mistral. Increasing the number of models employed leads to more stable results.
Conclusions. The main result of the proceeding is that the patterns that reveal LLMs under certain settings can also be identified by classical models. Moreover, augmenting the deployment of LLMs during the data preparation stages contributes to heightened stability in outcomes. Using classical models combined with LLMs speeds up response times during inference and reduces operating costs for running models.
Keywords: system analysis; foresight; textual analytics; classification; LLM; NLP.

Introduction

The new area of digitalization has opened the potential insights from textual data, including news articles, social media posts, emails and user comments. A significant challenge for researchers and business lies in structuring this information for future demands, such as content generation, possible targeted recommendations, and value enhancement [1]. These tasks are equally relevant in the government sector, particularly in the realm of strategic management and planning, where ongoing analysis of information from external and internal sources not only helps to address current needs but also enables long-term planning for decades ahead. Specifically tailored for long-term planning within government structures, the foresight methodology [2] has been developed, and actively implemented across all levels of government management bodies [3]. In numerous foresight processes, it is crucial to engage experts, stakeholders, and citizens in innovative workshops and strategic dialogues aimed at making sense of complex issues (or futures). According to the methodology, these tasks involve creative and mathematical techniques for constructing the future through various workflows, which process formatted and formalized knowledge as semi-structured and textual data through the cognitive efforts of experts,

Пропозиція для цитування цієї статті: В.В. Савастьянов, М.В. Столяр, "Системний підхід до комбінованого використання великих мовних та класичних моделей в задачах передбачення", *Наукові вісті КПІ*, № 1–4, с. 40–48, 2024. doi: orcid.org/10.20535/kpisn.2024.1-4.315079

Offer a citation for this article: Volodymyr Savastiyanov, Mykhailo Stoliar, "System Approach to the Combined Use of Large Language Models and Classical Models in Foresight Tasks", *KPI Science News*, no. 1–4, pp. 40–48, 2024. doi: orcid.org/10.20535/kpisn.2024.1-4.315079

specialists, and mathematical models. A substantial hurdle presents itself in automating these workflows, particularly from the initial stage of processing incoming information. Some of these workflows and creative methodologies can leverage modern NLP approaches, yet they encounter typical NLP task problems such as data annotation. This process remains labor-intensive and time-consuming for numerous NLP tasks. While there are several methods available to generate pseudo-data labels, these methods are typically specific to particular tasks and initially require a substantial volume of labelled data. The introduction of the extensive language model GPT-3, which boasts 175 billion parameters, has significantly enhanced performance across various few-shot learning tasks. In [4] was investigated the potential of utilizing GPT-3 in combination with human labelling as an economical data labelling tool to train other models. Shuohang Wang and colleagues provide numerous tables and indicators that identify effective combinations of a person and GPT for specific datasets. However, they have not shown which traditional machine learning models can achieve similar performance to any Human&LLM combination within a smaller budget. In this article, we are going to investigate such powerful tools like Large Language Models (LLMs) in specific domains, focusing on the machine learning paradigm of Zero-Shot Classification, which operates without pre-labelled data, new progressive developed chain of thought and Prompt Engineering methodology [5]. Using it in LLM is important for several reasons: clarity (clear and understandable information), coherence (the logical connection between ideas), persuasiveness (cases more compelling), credibility (well-reasoned insights), problem-solving (systematically analysing the problem) etc. Overall, a special chain of thought and Prompt Engineering in communication with LLMs enhances their ability to engage in dynamic, contextually aware, and insightful interactions, leading to more meaningful and productive communication experiences. The second step is to estimate the influence of text data on the domain. Our dataset, parsed from the beginning of 2011 to the end of 2023, includes over 120,000 records of news headlines extracted from trusted and popular sources like The Guardian, Time, 9News, FreightWaves, Journal of Commerce, and The Times of Earth. The domain-specific area is the logistics sector within the USA market, aiming to provide a comprehensive overview over the specified period. The data includes information not only from the target domain. It was chosen for a more general result. To be more precise, this study is deep into whether traditional machine learning methods, for example, TF-IDF and Logistic Regression, can be used as valuable alternatives to LLMs for specific analytical tasks (classification). A successful substitution would suggest that the data contains identifiable patterns that LLMs and traditional methods can accurately recognize. Moreover, it would underscore the importance of these patterns in informing the effectiveness and reliability of various computational techniques employed in linguistic analysis. Additionally, the paper would elucidate the implications of such findings for advancing the field of NLP and optimizing the performance of LLMs in real-world applications. Our research is not to solely investigate the classification prowess of LLMs across diverse domains. Rather, we aim to explore the feasibility of substituting LLMs with conventional approaches grounded in established methodologies. As a result, we can decrease the amount of energy and expense used for inference LLMs. As we know, to be comfortable working with LLMs you need GPU accelerators. It is expensive and not environmentally friendly (CO_2 emissions during inference). This exploration could potentially challenge the dependence on LLMs and represent the objectivity of the knowledge base and their hallucination effect.

Problem Statement

In recent years, there has been significant research into NLP technologies, including both LLMs and traditional machine learning techniques. The application and estimation of these methods have become critical in understanding their relative strengths and weaknesses. For example, in [6], the potential of NLP, specifically topic modelling, is discussed for identifying differences between citizen-derived foresight and institutional foresight, enhancing strategic foresight processes. In [7], the use of machine learning (ML) and NLP for analysing large sets of media texts are investigated, aiming to underline the foresight process by identifying future-oriented statements and trends. This paper seeks to bridge the gap by evaluating the performance of classical machine learning techniques against LLMs in natural language processing tasks, addressing the extreme popularity in the media of new progressive models, and estimating their practical applications. The authors discovered the utilization of textual analytics within the Foresight process [8], yet encountered a formidable challenge in surpassing the capabilities of cutting-edge textual analytics technologies such as LLM. In the related works [9–12], GPT models were explored on tasks using machine learning paradigms zero-shot, one-shot, and few-shot classification. They found that GPT zero-shot performance is significantly weaker than few-shot performance in tasks like reading comprehension, question answering and classification. You can see it by differences in the f1-score (and other provided metrics) for different classification datasets. The main hypothesis is that without few-shot observations, it becomes more difficult for models to perform effectively. Overall, these findings shed light on the nuanced dynamics underlying the performance of LLMs in different learning problems, highlighting the pivotal influence of data availability, size and quality, and model training configurations like optimization algorithm, batch size, learning ratio, and so on for task effectiveness. To address the combination challenge of union Foresight studies [13] and LLMs, an approach is introduced to integrate future modelling into the existing learning frameworks [14]. By employing the subject trajectory, which serves as a structured representation of consecutive frame sequences, as a learning objective, the goal is to establish dependencies between past and future contexts. They have proposed two innovative methodologies, namely Foresight Pre-Training (FPT) and Foresight Instruction-Tuning (FIT), to endow MLLMs with predictive capabilities, drawing inspiration from the contemporary learning paradigms observed in Language Models (LMs) [15]. In our work, we are going to investigate the classification task on the example of STEEP analysis, as a part of foresight study, using special techniques like prompt engineering and chain of thought in a domain-specific area - the logistic market in the USA. The main goal of the work is to figure out that LLMs can see patterns in text data (news headlines) and strictly recognize them and then create ML techniques that solve this task with similar quality to save computational time and fee for LLM clouds. Such an approach brings benefits not only within scientific circles but also across various domains. Industries can extract the predictive capabilities of LLMs and NLP methodology empowered with foresight studies to enhance decision-making processes, optimize resource allocation, and anticipate market trends.

Methods

The process flow of the system approach, as illustrated in Fig. 1, involves using panels, experts, analysts, and tools to generate or collect text documents in a foresight study (data discovery step). These documents are tagged using LLMs in STEP 1, despite the potential cost inefficiency. In STEP 2, a classical machine learning model is trained with the tagged data. STEP 3 involves deploying the trained ML model for predictions or analyses. This systematic approach focuses on the comparative evaluation and cost assessment of LLMs and classical machine learning techniques, to achieve an optimal balance between performance and cost-efficiency in NLP tasks within the context of foresight studies. It emphasizes the combined use of large language models and classical models to enhance the effectiveness and efficiency of foresight-related analyses.

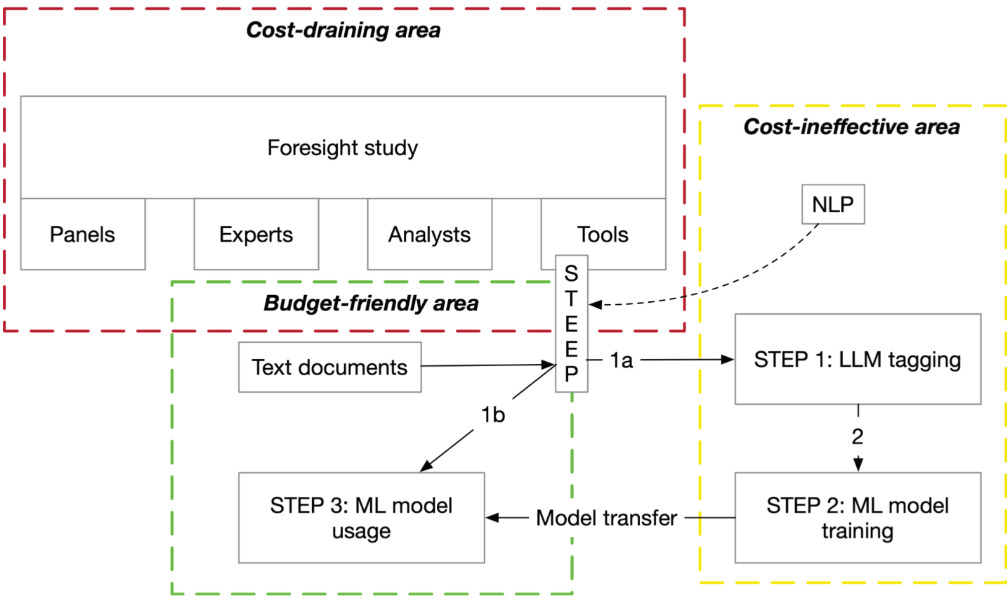

Fig. 1. Integrating Cost-Effective Strategies in Foresight Studies: A Dual-Step Approach with LLM Tagging and ML Model Optimization

The cost implications encompass three distinct zones: the cost-draining area, the cost-ineffective area, and the budget-friendly area, distinguished by red, yellow, and green dashed borders respectively. According to our methodology, the Foresight process falls within the cost-draining area due to its reliance on labor-intensive organizational and creative activities. However, in our study, we employ NLP tools for Foresight activities (STEEP analysis as an example, there can be other approaches depending on classification), potentially automating certain tasks and optimizing budget allocation. System approach to the combined use of large language models and classical models in a budget-friendly area aims to streamline processes and enhance cost-effectiveness in Foresight.

Data Collection

Data collection is a key component in the major of research works [16], especially for those, who are centered around language models (LMs). Within this domain, an array of diverse textual datasets exists, each offering unique insights and challenges. It can be books, papers, subtitles, messages, news, to name a few. In this study, a dataset has been compiled, encompassing news headlines starting from 2011 to the end of 2023. We have to admit that emphasizing the inclusion of data from the most recent timeline (2023) is important for analysis, as certain LLMs may have been trained on older data, potentially influencing their performance. While our primary focus is not centered on the utilization of historical data in LLM training processes, we prioritize minimizing data leaks [17] and ensuring the stability of our results. The main sources of data are The Guardian, Time, 9News, FreightWaves, Journal of Commerce, and The Times of Earth. The data includes more than 120,000 records for the targeted period.

Large Language Models as experts

In recent years, the intersection of Foresight methodologies and Language Models (LLMs) has become more promising across academic, commercial, and industrial domains. Foresight, the ability to anticipate future developments, trends or needs, has long been a wished tool in various fields. Meanwhile, LLMs have revolutionized natural language processing, demonstrating remarkable proficiency in understanding and generating human-like text, also describing the logical chain of future reasoning. There are a lot of LLMs that are accessible to users and researchers like GPT, LLAMA, Mistral, BART, GEMINI, etc. The primary difference between them lies in architecture variance, number of parameters, and text corpora for training. In these experiments, we stop on GPT-3.5-turbo and Mixtral-8x7B-Instruct-v0.1 (Mistral) as base LLMs. The anchor parameter was chosen $top\_k = 50$ for text generation for both models [18]. The selection of benchmark LLMs such as GPT and Mistral for our study is underpinned by several factors, each contributing to the comprehensive evaluation and comparison: established performance (robust performance across various NLP tasks), architectural significance (built upon transformer architectures), large-scale pre-training (pre-training on vast text corpora, diverse linguistic contexts and domains), availability and accessibility (publicly available). In this paper, we consider GPT and Mistral models as leading experts in the field of logistics. These models show robust results in news classification tasks [19]. The hypothesis is that they are two independent experts, whose knowledge base was fulfilled during the training procedure. To extract that knowledge we use prompt engineering and chain of thoughts. For the zero-shot classification, we use STEEP categories [20] (as an example) according to Foresight methodology [2]. The process of the STEEP analysis involves a deep exploration of external environmental factors. This analytical framework divides the scope into Social, Technological, Economic, Environmental, and Political categories, which lead to a comprehensive view of external influences on trends. It is used in decision-making systems, strategy planning, and other Foresight tasks. As we mentioned before, our LLMs hold expertise in the area of logistics. The utilization of appropriate instructions or prompts is imperative to effectively extract knowledge from LLMs, and then to transfer them to domain experts. This process involves carefully selecting prompts that guide the attention layers of models toward the desired information, thereby enhancing its ability to generate relevant and accurate responses with the minimization of hallucinations. The team has compiled a comprehensive list of prompts, guided by our perspective. They are well-suited to the task of determining the impact of current headlines on the logistics industry, assigning a rating ranging from 1 to 10. The combined prompt includes instructions for classification and estimation tasks for multiple observations, see Algorithm 1.

Algorithm 1

Base prompt

[INST]
You are a TOP LEVEL LOGISTIC ANALYST with twenty years of experience.
The task is to classify news headlines in [INPUT]. Estimate the impact of each headline on the USA LOGISTIC MARKET.
[Rules]
For classification use STEEP analysis.
There are a total of five classification categories: Social, Technological, Economical, Environmental, and Political.
Each text has one, two or three categories. Be aware. Estimation measures from 1 to 10. The higher rank the stronger impact of headline on the USA LOGISTIC INDEX.
Use the Example to format the output data: headlines | classes | impact
[End of Rules]
[Examples]
Pics of the week: Fields of red honour the fallen on Remembrance Day | Social, Environmental | 2
Rare echidna found for the first time in 62 years | Environmental | 3
Afghan girls win EU prize for farm robot | Technological | 1
Ice slip victims filmed instead of warned | Political | 1
Fuel shortage leads to increase in shipping costs for logistics companies | Economic | 8
[END Examples]
[/INST]
[INPUT]
"{text}"
[OUTPUT]

The prompt consists of the instruction, which transfers LLM to be an expert in the domain area (USA logistic market), the rules of output generation, which determines a consistent chain of generation in strict order, examples of desired output, and field of "text" inputs. In other words, the prompt determines not only the role of the assistant (identity and instructions) but also includes represented examples of the desired output. This is the way we utilize prompt engineering and chain of thought methodology to achieve efficiency in generation with a minimum of hallucinations.

ML model training and inference

Natural Language Processing (NLP) stands at the frontline position of modern computational linguistics methodology, offering a multifaceted approach to understanding and processing human language. NLP methods serve as a bridge to the gap between human communication and machine understanding [21]. NLP includes not only approaches to work with text data but also has metrics to evaluate achieved results. These metrics provide quantitative measures of model accuracy, coherence, fluency, and other relevant attributes. In our study, the text corpus is pre-processed by lemmatization and the removal of stopwords, ensuring that only meaningful content remains for subsequent analysis. Following pre-processing, the transformed text data is then encoded into machine-readable features using the Term Frequency-Inverse Document Frequency (TF-IDF) [22–23] approach with bigrams. This transformation allows us to represent the textual data in a numerical format. Logistic Regression is a fundamental statistical and machine-learning technique used for classification tasks. In this study, it was used as a baseline machine learning algorithm. The core idea behind the method is to find a relationship between features and the probability of a particular outcome. As the TF-IDF approach was used in the text pre-processing step, Logistic Regression is a proper choice due to the number of input features [24]. For result estimation, standard metrics such as precision, recall, and F1-score are utilized. Given the prevalence of imbalanced classes within the real data, similar to our dataset, we opt for the use of macro average and weighted average aggregation methods to ensure a more precise evaluation of the final metrics.

Results

In the previous steps of the work: the dataset was collected, consisted of news headlines, configured experts (prompt with generation parameters for GPT and Mistral models), and generated classes with estimated impact value to domain area for each observation. The hypothesis of this research aims to explore potential dependencies within the dataset and target STEEP classes, and estimated impact on domain area. Additionally, it seeks to level out the hallucination effect of generated information. The Proof of concept was achieved with the help of classic machine learning methods in the Natural Language Processing (NLP) field. In the first step, we split data to train and test in proportion 4:1 with stratification by target class. A text corpus is pre-processed with the TF-IDF approach with bigrams. It transforms text into features. In the second step, we take Logistic Regression as a benchmark machine learning model. The classification report for the STEEP classes is represented in Table 1 by the first expert assumption.

***Table* 1.** STEEP classification report (GPT)

Class	Precision	Recall	F1 score	Sample
Economic	0.84	0.89	0.86	7006
Environment	0.89	0.7	0.78	1383
Politic	0.83	0.85	0.84	5266
Social	0.79	0.76	0.78	4206
Technology	0.78	0.75	0.76	3548
weighted avg	0.82	0.82	0.82	21468

The classification report for the STEEP classes is represented in Table 2 by second expert assumption.

***Table* 2.** STEEP classification report (Mistral)

Class	Precision	Recall	F1 score	Sample
Economic	0.76	0.85	0.8	8089
Environment	0.76	0.56	0.65	1317
Politic	0.72	0.78	0.75	5705
Social	0.66	0.56	0.6	3290
Technology	0.71	0.63	0.66	4499
weighted avg	0.73	0.73	0.72	22900

The classification report for the STEEP classes is represented in Table 3, where both experts have a consensus in class prediction.

***Table* 3.** STEEP classification report (Consensus)

Class	Precision	Recall	F1 score	Sample
Economic	0.88	0.94	0.92	5799
Environment	0.94	0.69	0.8	866
Politic	0.88	0.9	0.89	3959
Social	0.84	0.76	0.8	2136
Technology	0.85	0.83	0.83	2578
weighted avg	0.87	0.87	0.87	15338

In the third step, an attempt is made to replicate the experiment to estimate the impact in the domain area. The results are represented in Table 4 for the combination of models. It is evident from the figure that even the predictions made by a multitude of models are not recognizable by conventional methods.

***Table* 4.** Classification report of impact (Consensus)

Class	Precision	Recall	F1 score	Sample
1	0.73	0.57	0.64	371
2	0.37	0.15	0.21	279
3	0.32	0.28	0.3	382
4	0.26	0.46	0.33	521
5	0.20	0.11	0.13	393
6	0.25	0.26	0.26	461
7	0.25	0.28	0.30	510
8	0.32	0.25	0.28	370
9	0.45	0.33	0.38	289
weighted avg	0.34	0.32	0.31	3576

In exploring the cost-efficiency of Large Language Models, it is important to analyse how our approach can reduce operational expenses. OpenAI's GPT-3.5 Turbo, for example, operates at a cost of $\$6 \cdot 10^{-4}$ per 1000 tokens. By optimizing calculations to run on CPUs, our model significantly reduces this cost to $\$3.5 \cdot 10^{-12}$ per 1000 tokens.

Conclusions

This study has demonstrated the feasibility and efficacy of employing traditional machine learning methods as viable substitutes for Large Language Models (LLMs) in conducting classification tasks like STEEP analysis, for example, as a part of foresight study, within specific domain areas. This study is applicable to any classification tasks inside the foresight methodology. By systematically comparing the performance of classic classification models, like logistic regression to a valuable substitution of state-of-the-art LLMs, which are overestimated in social prospect. Logistic regression has a better performance based on labelling from GPT than the Mistal model by weighted average F1 score. It can be explained as the varying number of parameters in LLMs and text of different size (including text streams) in the training corpora. Also, taking a swarm of models and their combination for prediction (consensus), we can achieve significantly improved and more robust results. Logistic regression has a weighted average F1 score of 0.87, see Table 3. This experiment proves that the data has some special features (keywords and dependencies), which are recognizable by LLMs and classic machine learning approaches. The second experiment is less successful, primarily due to two distinct reasons. Firstly, the estimation of dependencies in the domain of logistics is much more complicated, and classic methods are not a proper instrument for it. Secondly, there are possible hallucinations of LLMs in their output. This phenomenon can be presented as insufficient data during the training process of LLMs and difficulty in generating numerous characteristics precisely.

Moreover, the study revealed that traditional models provide a considerable advantage in terms of cost optimization. The proposed approach is cheaper in $1.7 \cdot 10^8$ times per 1000 tokens processing. Implementing these models requires significantly fewer computational resources compared to LLMs, which are often resource-intensive due to their complex architectures and large-scale data requirements. This aspect of cost-effectiveness is particularly crucial for organizations aiming to enhance their analytical processes without incurring prohibitive expenses. In conclusion, our research not only validates the practicality of combining LLMs with more cost-efficient and equally effective traditional machine learning models but also highlights the broader applicability and benefits of the proposed system approach. These findings encourage further exploration and adoption of classic machine learning techniques, promoting a more accessible, economical, and flexible analytical framework in various industries.

Furthermore, this study briefly touched upon the impact of various pre-processing techniques on the performance of LLMs compared to traditional natural language processing methods. While logistic regression offers simplicity and computational efficiency, it may not adequately capture complex relationships that more sophisticated models like random forests or neural networks could elucidate. The proposed system approach of a combination of LLMs and traditional ML models presents robust performance and results. Future research could also explore the integration of additional LLM platforms beyond the GPT-3.5-turbo and Mixtral-8x7B-Instruct-v0.1 used in this study, to assess their effectiveness across a broader range of applications and datasets. This would provide a more comprehensive understanding of how different LLMs perform in diverse analytical contexts.

Author Contributions: Volodymyr Savastiyanov: Conceptualization, Methodology, Resources, Writing – Review & Editing. Mykhailo Stoliar: Methodology, Software, Investigation, Data Curation, Writing – Original Draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The information supporting this research is readily accessible. https://www.kaggle.com/datasets/tomatobuster/headline-news-data

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

References

[1] Li et al., “Knowledge structure of technology licensing based on co-keywords network: A review and future directions” *International Review of Economics & Finance*, vol. 66, pp. 154–165, 2020. doi: 10.1016/j.iref.2019.11.007.

[2] UNIDO, UNIDO Technology Foresight Manual, vol. 1. United Nations Industrial Development Organization, Vienna (2005).

[3] European Commission. 2020. “Strategic Foresight Report – Charting the course towards a more resilient Europe”. Available online at: https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/strategic-foresight/2020-strategic-foresight-report_en.

[4] Shuohang Wang, Yang Liu, Yichong Xu, Chenguang Zhu and Michael Zeng, “Want To Reduce Labeling Cost? GPT-3 Can Help”, arXiv preprint arXiv.2108.13487, 2023, doi: 10.48550/arXiv.2108.13487.

[5] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing”, ACM Computing Surveys, vol. 55, no. 9, pp. 1–35, 2023.

[6] A. B. Rosa, N. Gudowsky, & P. Repo, “Repo Sensemaking and lens-shaping: Identifying citizen contributions to foresight through comparative topic modelling”, Futures 129, pp. 1–15, 2021, doi: 10.1016/j.futures.2021.102733.

[7] M. Vignoli, J. Rurden, D. Wasserbacher and S Kimpeler, “An Exploration of the Potential of Machine Learning Tools for Media Analysis to Support Sense-Making Processes in Foresight”, Frontiers in Communication 7:750614, 2022. doi :10.3389/fcomm.2022.750614.

[8] N. Pankratova, V. Savastiyanov, “Assessment of situations in the field of social disasters based on the methodology of foresight and textual analytics”. Proceedings of the 2019 IEEE. Second International Conference IEEE UKRCON-2019. pp. 1207–1210, ISBN 9781728138831.

[9] Callie Y. Kim, Christine P. Lee and Bilge Mutlu, “Understanding Large-Language Model (LLM)-powered Human-Robot Interaction”, In Proceedings of the 2024 ACM/IEEE, International Conference on Human-Robot Interaction (HRI ’24). Association for Computing Machinery, New York, NY, USA, pp. 371–380, 2024, doi: 10.1145/3610977.3634966.

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., “Language models are few-shot learners,” Advances in neural information processing systems, vol. 33, pp. 1877–1901, 2020.

[11] B. Zhang and H. Soh, "Large Language Models as Zero-Shot Human Models for Human-Robot Interaction", 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, pp. 7961–7968, 2023, doi: 10.1109/IROS55552.2023.10341488.

[12] F. Pourpanah et al., "A Review of Generalized Zero-Shot Learning Methods", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4051–4070, 1 April 2023, doi: 10.1109/TPAMI.2022.3191696.

[13] C. Mühlroth and M. Grottke, "Artificial Intelligence in Innovation: How to Spot Emerging Trends and Technologies", in *IEEE Transactions on Engineering Management*, vol. 69, no. 2, pp. 493–510, April 2022, doi: 10.1109/TEM.2020.2989214.

[14] M. Stoliar and V. Savastiyanov, "Using LLM classification in foresight studies": Scientific Collection "InterConf", vol. 157, pp. 367–375, 2023.

[15] En Yu, Liang Zhao, Yana Wei, Jinrong Yang, Dongming Wu, Lingyu Kong, Haoran Wei, Tiancai Wang, Zheng Ge, Xiangyu Zhang and Wenbing Tao, "Merlin: Empowering Multimodal LLMs with Foresight Minds", arXiv preprint arXiv:2312.00589, 2023, doi: 10.48550/arXiv.2312.00589.

[16] Y. Luo, S. Xu and C. Xie, "E-commerce Big Data Classification and Mining Algorithm based on Artificial Intelligence", 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, pp. 1153–1155, 2022, doi: 10.1109/ICETCI55101.2022.9832370.

[17] Chan A., "GPT-3 and InstructGPT: technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry", AI Ethics 3, pp. 53–64, 2023, doi: 10.1007/s43681-022-00148-6.

[18] Dave Wild, "Futurework – A Guidebook for The Future of Work", Aotearoa New Zealand: Smith & Wild. 218 p., 2023, ISBN 978-0-473-66594-4.

[19] Mowafak Allaham and Nicholas Diakopoulos, "Supporting Anticipatory Governance using LLMs: Evaluating and Aligning Large Language Models with the News Media to Anticipate the Negative Impacts of AI", arXiv preprint arXiv:2401.18028, 2024, doi: 10.48550/arXiv.2401.18028.

[20] Cambria Erik, Bebo White., "Jumping NLP curves: A review of natural language processing research", IEEE Computational intelligence magazine 9.2, pp. 48–57, 2014, doi: 10.1109/MCI.2014.2307227.

[21] A. Jalilifard, V.F. Caridб, A.F. Mansano, R.S. Cristo, F.P.C. da Fonseca, "Semantic Sensitive TF-IDF to Determine Word Relevance in Documents", Thampi, S.M., Gelenbe, E., Atiquzzaman, M., Chaudhary, V., Li, KC. (eds) Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 736. Springer, Singapore, 2021, doi.org: 10.1007/978-981-33-6987-0_27.

[22] B. Ahmed, G. Ali, A. Hussain, A. Baseer and J. Ahmed, "Analysis of Text Feature Extractors using Deep Learning on Fake News", *Eng. Technol. Appl. Sci. Res.*, vol. 11, no. 2, pp. 7001–7005, Apr. 2021.

[23] Lucky Agarwal, Kartik Thakral, Gaurav Bhatt, and Ankush Mittal, "Authorship Clustering using TF-IDF weighted Word-Embeddings", In Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE '19). Association for Computing Machinery, New York, NY, USA, pp. 24–29, 2019, doi: 10.1145/3368567.3368572.

[24] Fanghui Liu, Xiaolin Huang, and Jie Yang, "Indefinite Kernel Logistic Regression", In Proceedings of the 25th ACM International Conference on Multimedia (MM '17). Association for Computing Machinery, New York, NY, USA, pp. 846–853, 2017, doi: 10.1145/3123266.3123295.

Володимир Савастьянов, Михайло Столяр

СИСТЕМНИЙ ПІДХІД ДО КОМБІНОВАНОГО ВИКОРИСТАННЯ ВЕЛИКИХ МОВНИХ ТА КЛАСИЧНИХ МОДЕЛЕЙ В ЗАДАЧАХ ПЕРЕДБАЧЕННЯ

Проблематика. Великі мовні моделі (LLMs) та пов'язані з ними агенти широко застосовують у різних сферах життя та є одними з технологічних проривів за останні роки. Ці найсучасніші моделі демонструють вражаючий потенціал, але в деяких ситуаціях вони демонструють неефективність. У цій статті досліджено виявлені обмеження у конкретних галузевих сферах і на прикладах задач передбачення.

Мета дослідження. У статті висвітлено можливості, які пропонують моделі на основі GPT, та зіставлено висновки із класичними методами аналізу текстових даних у задачах класифікації на прикладі методології передбачення. Метою дослідження є розроблення системного підходу до комбінованого використання традиційних підходів машинного навчання як практичної альтернативи LLMs у задачах передбачення (форсайту) на прикладі STEEP-аналізу, який дає можливість отримувати цінну інформацію із текстових даних.

Методи. Це дослідження структуроване на чотири сегменти: Data Mining, передобробка тексту з використанням LLMs, передобробка тексту за допомогою класичних методів обробки природної мови (NLP) та порівняльний аналіз результатів. Data Mining включає етапи збирання даних і попередної обробки даних для навчальних і тестових спостережень. Для використання LLMs було застосовано підходи "chains of thought" та "prompt engineering".

Результати. За результатами дослідження було встановлено, що LLMs можуть бути застосовані у комбінації із класичними методами машинного навчання для доменних специфічних сфер у STEEP-аналізі завдачах прогнозування. Результати показали, що запропонованих підхід є значно швидшим і має меншу складність порівняно з LLMs, такими як GPT і Mistral. Збільшення кількості використаних моделей приводить до більш стабільних результатів.

Висновки. Основний результат роботи полягає в тому, що патерни, які виявляють LLMs за певних налаштувань, також можуть бути виявлені класичними моделями. Понад те збільшення кількості використаних LLMs на етапах обробки даних сприяє підвищенню стабільності результатів. Використання класичних моделей у комбінації з LLMs прискорить час відповіді та зменшить експлуатаційні витрати на запуск моделей.

Ключові слова: системний аналіз; методологія передбачення; текстова аналітика; класифікація; LLM; NLP.

Рекомендована Радою
НН інституту прикладного системного аналізу
КПІ ім. Ігоря Сікорського

Надійшла до редакції
9 жовтня 2024 року

Прийнята до публікації
19 листопада 2024 року