Natural Language Processing in Connected Vehicle Patents Analysis

Subjects: Urban Studies
Contributor: Raj Bridgelall

Patents are a valuable source of information for understanding technological advancements, identifying emerging trends, and assessing the competitive landscape of various industries. Analysts have used NLP and ML to automate tasks such as patent classification, topic modeling, technology identification, and patent recommendation systems.

Keywords: natural language processing ; NLP model evaluation ; machine learning

1. Introduction

Patents are a valuable source of information for understanding technological advancements, identifying emerging trends, and assessing the competitive landscape of various industries. However, because of the vast and growing volume of patent data and their unstructured nature, they are challenging to analyze manually [1]. Current approaches in patent analysis have primarily focused on manual interpretations, leading to challenges in scalability and consistency.

Analysts have used NLP and ML to automate tasks such as patent classification, topic modeling, technology identification, and patent recommendation systems [2]. Even so, verifying the alignment of ML methods with human understanding of patent documents remains challenging and is an unresolved issue [3]. Hence, establishing a performance benchmark of state-of-the-art, free, open-source models will set the stage to evaluate the advancement of machines that can understand and classify patents. Such advancements could include large language models (LLMs), but accessing their application programming interface (API) is not currently free or widely available to everyone, and the models are still immature.

The goal of this research is to develop a comprehensive metric for quantifying alignment between subject matter expert (SME) and ML classification of patent topics. The author selected the field of CVs for the case study based on his more than 30 years of domain knowledge and industry experience inventing relevant products. With dozens of related patent awards in the field, the author amassed decades of experience reviewing and analyzing the nuances of patent documents in the CV and intelligent transportation systems domain.

2. Natural Language Processing in Connected Vehicle Patents Analysis

2.1. Connected Vehicles Trends

A review of the latest trends in CV technology development sets the stage for understanding the topic classifications. The transportation industry expects CV technology to revolutionize driving safety, efficiency, and convenience [4]. The recent literature on CVs focused on advancing security, traffic management, and cooperative control systems. Nkenyereye et al. (2023) examined the integration of 5G networks in vehicular cloud computing, highlighting the potential of vehicle clusters to share resources and data in a mobile cloud environment [5].

The security of CVs is another area of major concern. Shichun et al. (2023) reviewed cybersecurity techniques focusing on CVs, covering threat analysis and intrusion detection [6]. Rathore et al. (2023) focused on the cybersecurity challenges of in-vehicle communication [7], while Ju et al. (2022) examined attack detection and resilience from a vehicle control perspective [8]. Hildebrand et al. (2023) explored the use of blockchain technology to enhance security in vehicle networking [9], and Khan et al. (2023) proposed a blockchain-based secure communication system for CVs [10].

Alanazi (2023) conducted a systematic literature review of how autonomous vehicles manage traffic at junctions, exploring various methodologies that include ML [11]. Shi et al. (2023) developed a real-time control algorithm for CVs to optimize fuel use in signalized corridors [12]. Gholamhosseinian and Seitz (2022) surveyed cooperative intersection management strategies for CVs [13]. Xu and Tian (2023) proposed a method to improve arterial signal coordination using CV data [14], and Zhu et al. (2022) reviewed merging control strategies at freeway on-ramps [14]. Wang et al. (2022) discussed the development of cooperative driving systems for CVs [15].

Cui et al. (2022) reviewed cooperative perception technologies that combine local and edge sensing data for improved situational awareness [16]. Khanal et al. (2023) utilized CV data to develop crash prediction models. Gao et al. (2023) provided insights into predictive cruise control under cloud control systems, emphasizing the role of predictive algorithms in traffic efficiency and safety [17]. Islam and Abdel-Aty (2023) focused on using CV data for traffic conflict prediction [18]. Schwarz et al. (2022) examined the role of digital twin simulations in various traffic management applications [19].

The above literature review revealed that trending topics in CV development include vehicular security, traffic management, cooperative control and perception systems, and overall driving safety. These trends suggest that while the industry has made considerable progress in the development of technologies for CVs, challenges in security, real-time control, and effective traffic management remain, necessitating further research and innovation in these areas.

2.2. NLP in Patent Analysis

This part reviewed synthesizes NLP-related contributions in patent analysis from the recent literature and contrasts them with the current paper's unique contributions. Krestel et al. (2021) surveyed the use of deep learning for patent analysis, emphasizing the significant role of these techniques in automating tasks previously solvable only by domain experts [2]. Casola and Lavelli (2022) surveyed NLP approaches for summarizing, simplifying, and generating patent text for non-experts to improve the accessibility and understanding of patent information for a wider audience [1]. Their work highlighted the peculiar challenges patents pose to current NLP systems. This finding suggested that the performance of NLP and ML methods is sensitive to the type of topic domain analyzed.

Trappey et al. (2020) focused on NLP-based patent information retrieval and intelligent compilation of patent summaries [20]. Joshi et al. (2022) proposed an intelligent keyword extraction technique for patent classification by training a transformer model and comparing its performance with K-means clustering and topic modeling using the latent Dirichlet allocation (LDA) method [21]. These works highlight the challenges in effectively extracting and summarizing knowledge from patents and the opportunities that NLP techniques offer in this context.

De Clercq et al. (2019) proposed a multi-label classification approach to classify electric vehicle patents by combining LDA with ML algorithms to explore the relationships between patents and cooperative patent classification classes [22]. Their method provided a user-friendly way to analyze and visualize patent data. Hyun et al. (2020) and Wu et al. (2020) explored semantic analysis of patent data, emphasizing the role of NLP in identifying technical trends [23] and screening patents in specific domains such as communication technologies in construction [24]. These studies highlight the evolution of NLP and ML tools and the need for continuous innovation in this field.

Arts et al. (2021) demonstrated the potential of topic modeling in enhancing the understanding of patent documents and providing insights into the evolution of technologies [25]. Puccetti et al. (2023) proposed a named entity recognition (NER) method to identify technology-related entities from patent texts [26]. Their approach utilized a combination of rule-based and ML techniques to identify entities such as products, processes, organizations, and locations. Rezende et al. (2022) combined NLP techniques with algorithms such as latent semantic analysis (LSA), word2vec, and word mover's distance (WMD) to analyze patent similarity and technology trends [27].

2.3. NLP Alignment Evaluation

This part focuses on works that surveyed or compared the performance of NLP and ML methods. Kherwa and Bansal (2019) presented a comprehensive survey of topic modeling, highlighting challenges in their quantitative evaluation [28]. Abdelrazek et al. (2023) provided a recent survey categorizing topic modeling techniques into algebraic, fuzzy, probabilistic, and neural categories [29]. They reviewed the diversity and evolution of topic models and found that the research trends are moving toward developing and tuning neural topic models such as LLMs. Meaney et al. (2023) focused on methods for assessing the quality of topic models. They explored metrics such as reconstruction error, topic coherence, and stability analysis, emphasizing that different metrics capture various aspects of model fit [30]. Their findings suggested that a combination of indices, coupled with human validation, is essential for assessing the performance of topic models, a perspective that highlights the complexities of evaluating model quality. Harrando et al. (2021) empirically evaluated the challenges in systematically comparing topic modeling algorithms, revealing shortcomings in common practices and highlighting the need for a standardized approach in model benchmarking [31]. Vayansky and Kumar (2020) reviewed various topic modeling methods, focusing on their ability to handle complex data relationships, such as correlations between topics and topic changes over time. Their work encouraged diversity in the choice of topic modeling methods, particularly for complex datasets [32]. Borghesani et al. (2023) compared human and artificial semantic representations in topic modeling [3]. They demonstrated that NLP embeddings still fall short of human-like semantic representations. Rüdiger et al. (2022) compared non-application-specific topic modeling algorithms and assessed their performance against known clustering [33]. They found that metrics used so far provided a mixed picture that made it difficult to verify the accuracy of topic modeling outputs, concluding that topic model evaluation remains an unresolved issue [33]. Hoyle et al. (2021) questioned the validity of automated topic model evaluation, suggesting a disparity between automated coherence metrics and human judgment [34].

The above studies demonstrate the wide range of applications for NLP and ML in patent analysis, highlighting the complexities in evaluating algorithmic interpretations in topic modeling and the importance of human validation. Relative to these studies, the present study stands out by uniquely combining topic modeling techniques with SME expertise to quantify alignment in patent topic classification, particularly in the context of CV patents. The proposed topic alignment index, combined with an optimized NLP pipeline, fills a crucial gap in the literature by offering a more rigorous and complete approach in establishing a performance benchmark to contrast with future advancements. As NLP and ML techniques continue to advance, their role in patent analysis is likely to expand, providing researchers, practitioners, and policymakers with even more powerful tools for understanding, managing, and forecasting technological innovation. Hence, this study not only contributes to the existing body of knowledge, but also opens new avenues for future research in patent analysis.

References

1. Casola, S.; Lavelli, A. Summarization, simplification, and generation: The case of patents. Expert Syst. Appl. 2022, 205, 117627.
2. Krestel, R.; Chikkamath, R.; Hewel, C.; Risch, J. A survey on deep learning for patent analysis. World Pat. Inf. 2021, 65, 102035.
3. Borghesani, V.; Armoza, J.; Hebart, M.N.; Bellec, P.; Brambati, S.M. The Three Terms Task—An open benchmark to compare human and artificial semantic representations. Sci. Data 2023, 10, 1–13.
4. USDOT. Vehicle-to-Everything (V2X) Communications Summit: Detailed Meeting Summary: Preparing for Connected, Interoperable Deployment Nationwide; United States Department of Transportation (USDOT): Washington, DC, USA, 2023.
5. Nkenyereye, L.; Nkenyereye, L.; Jang, J.-W. Convergence of Software-Defined Vehicular Cloud and 5G Enabling Technologies: A Survey. Electronics 2023, 12, 2066.
6. Shichun, Y.; Zheng, Z.; Bin, M.; Yifan, Z.; Sida, Z.; Mingyan, L.; Yu, L.; Qiangwei, L.; Xinan, Z.; Mengyue, Z.; et al. Essential Technics of Cybersecurity for Intelligent Connected Vehicles: Comprehensive Review and Perspective. IEEE Internet Things J. 2023, 10, 21787–21810.
7. Rathore, R.S.; Hewage, C.; Kaiwartya, O.; Lloret, J. In-Vehicle Communication Cyber Security: Challenges and Solutions. Sensors 2022, 22, 6679.
8. Ju, Z.; Zhang, H.; Li, X.; Chen, X.; Han, J.; Yang, M. A Survey on Attack Detection and Resilience for Connected and Automated Vehicles: From Vehicle Dynamics and Control Perspective. IEEE Trans. Intell. Veh. 2022, 7, 815–837.
9. Hildebrand, B.; Baza, M.; Salman, T.; Tabassum, S.; Konatham, B.; Amsaad, F.; Razaque, A. A comprehensive review on blockchains for Internet of Vehicles: Challenges and directions. Comput. Sci. Rev. 2023, 48, 100547.
10. Khan, R.; Mehmood, A.; Iqbal, Z.; Maple, C.; Epiphaniou, G. Security and Privacy in Connected Vehicle Cyber Physical System Using Zero Knowledge Succinct Non Interactive Argument of Knowledge over Blockchain. Appl. Sci. 2023, 13, 1959.
11. Alanazi, F. A Systematic Literature Review of Autonomous and Connected Vehicles in Traffic Management. Appl. Sci. 2023, 13, 1789.
12. Shi, Y.; Wang, Z.; LaClair, T.J.; Wang, C.; Shao, Y. Real-time control of connected vehicles in signalized corridors using pseudospectral convex optimization. Optim. Control. Appl. Methods 2023, 44, 2257–2277.
13. Gholamhosseinian, A.; Seitz, J. A Comprehensive Survey on Cooperative Intersection Management for Heterogeneous Connected Vehicles. IEEE Access 2022, 10, 7937–7972.
14. Xu, J.; Tian, Z. OD-Based Partition Technique to Improve Arterial Signal Coordination Using Connected Vehicle Data. Transp. Res. Rec. J. Transp. Res. Board 2022, 2677, 252–265.

15. Wang, B.; Han, Y.; Wang, S.; Tian, D.; Cai, M.; Liu, M.; Wang, L. A Review of Intelligent Connected Vehicle Cooperative Driving Development. Mathematics 2022, 10, 3635.
16. Cui, G.; Zhang, W.; Xiao, Y.; Yao, L.; Fang, Z. Cooperative Perception Technology of Autonomous Driving in the Internet of Vehicles Environment: A Review. Sensors 2022, 22, 5535.
17. Gao, B.; Wan, K.; Chen, Q.; Wang, Z.; Li, R.; Jiang, Y.; Mei, R.; Luo, Y.; Li, K. A Review and Outlook on Predictive Cruise Control of Vehicles and Typical Applications Under Cloud Control System. Mach. Intell. Res. 2023, 20, 614–639.
18. Islam, Z.; Abdel-Aty, M. Traffic conflict prediction using connected vehicle data. Anal. Methods Accid. Res. 2023, 39, 100275.
19. Schwarz, C.; Wang, Z. The Role of Digital Twins in Connected and Automated Vehicles. IEEE Intell. Transp. Syst. Mag. 2022, 14, 41–51.
20. Trappey, A.J.; Trappey, C.V.; Wu, J.-L.; Wang, J.W. Intelligent compilation of patent summaries using machine learning and natural language processing techniques. Adv. Eng. Inform. 2019, 43, 101027.
21. Joshi, U.; Hedaoo, M.; Fatnani, P.; Bansal, M.; More, V. Patent Classification with Intelligent Keyword Extraction. In Proceedings of the 2022 6th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 26–27 August 2022; pp. 1–7.
22. De Clercq, D.; Diop, N.-F.; Jain, D.; Tan, B.; Wen, Z. Multi-label classification and interactive NLP-based visualization of electric vehicle patent data. World Pat. Inf. 2019, 58, 101903.
23. Hyun, Y.-G.; Han, J.-H.; Chae, U.; Lee, G.-H.; Lee, J.-Y. A study on technical trend analysis related to semantic analysis of NLP through domestic/foreign patent data. J. Digit. Converg. 2020, 18, 137–146.
24. Wu, H.; Shen, G.; Lin, X.; Li, M.; Zhang, B.; Li, C.Z. Screening patents of ICT in construction using deep learning and NLP techniques. Eng. Constr. Arch. Manag. 2020, 27, 1891–1912.
25. Arts, S.; Hou, J.; Gomez, J.C. Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. Res. Policy 2021, 50, 104144.
26. Puccetti, G.; Giordano, V.; Spada, I.; Chiarello, F.; Fantoni, G. Technology identification from patent texts: A novel named entity recognition method. Technol. Forecast. Soc. Chang. 2023, 186, 122160.
27. de Rezende, J.M.; Rodrigues, I.M.d.C.; Resendo, L.C.; Komati, K.S. Combining natural language processing techniques and algorithms LSA, word2vec and WMD for technological forecasting and similarity analysis in patent documents. Technol. Anal. Strat. Manag. 2022, 1–22.
28. Kherwa, P.; Bansal, P. Topic modeling: A comprehensive review. EAI Endorsed Trans. Scalable Inf. Syst. 2019, 7.
29. Abdelrazek, A.; Eid, Y.; Gawish, E.; Medhat, W.; Hassan, A. Topic modeling algorithms and applications: A survey. Inf. Syst. 2023, 112, 102131.
30. Meaney, C.; Stukel, T.A.; Austin, P.C.; Moineddin, R.; Greiver, M.; Escobar, M. Quality indices for topic model selection and evaluation: A literature review and case study. BMC Med. Inform. Decis. Mak. 2023, 23, 1–18.
31. Harrando, I.; Lisena, P.; Troncy, R. Apples to apples: A systematic evaluation of topic models. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Online, 1–3 September 2021.
32. Vayansky, I.; Kumar, S.A.P. A review of topic modeling methods. Inf. Syst. 2020, 94, 101582.
33. Rüdiger, M.; Antons, D.; Joshi, A.M.; Salge, T.-O. Topic modeling revisited: New evidence on algorithm performance and quality metrics. PLoS ONE 2022, 17, e0266325.
34. Hoyle, A.; Goel, P.; Hian-Cheong, A.; Peskov, D.; Boyd-Graber, J.; Resnik, P. Is automated topic model evaluation broken? The Incoherence of Coherence. Adv. Neural Inf. Process. Syst. 2021, 34, 2018–2033.