Evaluating Public Consultation in Urban Planning via Neural Language Models and Topic Modelling

Christian Caton[1], Gabriele Pergola[2], Tessio Novack[3], and Yulan He[2]

[1]*CDT Urban Science, Warwick Institute for the Science of Cities, University of Warwick, UK*
[2]*Department of Computer Science, University of Warwick, UK*
[3]*Centre for Interdisciplinary Methodologies, University of Warwick, UK*

Urban planning has the fundamental role of managing the cooperative development of ever bigger cities and the community's cultures inhabiting those places. To make the best decisions, urban planners need to analyse relevant data and community responses, spread through abundant reports. This process can be labour intensive and might lead to overlooking less prominent but still relevant concerns of minorities. In this study, we present a Natural Language Processing framework to assist the analysis of consultation reports. The framework leverages state-of-the-art techniques that enable the urban planners to easily describe the issues of interest in free text and, as a result, to accurately identify the emerging concerns from different stakeholders. A first assessment of the London Plan's Green Belt policy has shown the capability of detecting specific community interests about urban planner problems, allowing quick identification of minorities' issues, otherwise overlooked due to the vast amount of data.

Keywords: urban planning; the London Plan; Green Belt; diversity and minorities; natural language processing; neural language models

History: received on 23 October 2021; accepted on 21 November 2021; published on 8 December 2021

1 Introduction

Urban planning is an fundamental process in which cities and urban areas are constructed based on several aims defined by the government in conjunction with local councils. This process affects the whole infrastructure of a city, such as: transportation, layout, housing densities, and culture. Without such planning, cities are bound to become less sustainable, efficient, and create more urban environments riddled with social inequalities. Urban planning aims to assure that all people have access to good quality living environments and sufficient housing in attempt to tackle the challenges of social inequity in cities (Smit et al., 2011).

The role of the urban planner is to identify requirements of specific communities and devise a plan in addressing these requirements, through development, expansion or refurbishment in specific areas (Park and Lee, 2009). To identify such requirements, planners need to analyse relevant data and community responses which they might gather through an abundance of reports. This process can create difficulties for planners as it requires them to deal with many long, detailed reports that can be considered very labour intensive, and thus, result in a potential lapse in identification of less prominent minorities. This creates the issue of planning discourse not considering the different perspectives emerging from multiple stakeholders or communities around a particular issue; enhancing the concept of barriers and segregation within communities and minorities where people might feel ignored (Higgins et al., 2005).

C Caton, G Pergola, T Novack, and Y He (2022): *Evaluating Public Consultation in Urban Planning via Neural Language Models and Topic Modelling*. In: FB Mocnik and R Westerholt (eds.), *Proceedings of the 3rd International Symposium on Platial Information Science (PLATIAL'21)*, pp. 49–54

https://doi.org/10.5281/zenodo.5767188

PLATIAL'21 Third International Symposium on Platial Information Science (PLATIAL'21)
Enschede, the Netherlands; 15–17 December 2021

To assist this process, natural language processing (NLP) is a promising field of artificial intelligence (AI), developing computational techniques for learning, understanding, and producing human language content (Hirschberg and Manning, 2015). Cai (2021) conducted a systematic literature search highlighting the recent and growing interest in NLP among urban researchers, arguing that the application of NLP in urban research has its advantages of raising new research questions, improving the use of big urban data, and reducing research costs. In this way, the collection of big data regarding public responses can have an impact on environmental characteristics through the policy-maker decisions and each response can be different amongst people; culture, background, and context can all play a part through the community responses, altering the space people reside in (Klettner et al., 2013).

In this study, we propose a new NLP framework, tailoring the recent advances in language and topic modelling to assist the urban planner in analysing consultation reports and ease access to concerns and perspectives about urban issues of interest. As a case study, we focus our analysis on extracting interests of different stakeholders towards the Green Belt policy (Greater London Authority, 2021, Chapter 8), a set of guidelines part of the broader London Plan aiming at preventing the urban sprawl of Greater London. The proposed framework first extracts the numerous topics discussed within the consultation reports via topic modelling (Blei et al., 2003). Then, the relevant topics are matched via clustering (Pergola et al., 2021) with the urban issue of interest (e.g., Table 2), whose description is provided as input to the framework and encoded using a contextualized neural language model (Reimers and Gurevych, 2019). The outcome of this process is a concise description of the main concerns and perspectives of different stakeholders (e.g., business, borough, local communities, etc.) about problems of interest for the urban planners.

To the best of our knowledge, this is the first attempt to combine contextualized language models and topic modelling in the context of urban planning, easing accurate access to concerns of urban communities that could be easily neglected in the vast amount of text. This approach could be one of the key tools in aiding urban planners in making better decisions and easing their involvement for future planning discourse. In the following, we first introduce the Green Belt consultation reports in more detail (Section 2). Then, we provide an overview of the framework and a more technical description of how we adapted the NLP methodologies to the problem (Section 3). We conclude with an assessment of the preliminary results showing the proposed framework's potential and accuracy, and a discussion about future works and new promising applications (Section 4).

2 Case Study and Datasets

The London Plan is a statutory spatial development strategy for the greater London area in the UK, prepared and published by the mayor of London (Greater London Authority, 2021, Chapter 8). The objectives are to ensure London is a city that can meet the demands of population and economic growth, exist as a successful competitive city internationally, improve the environment, provide easily accessible, secure, diverse and well-built neighbourhoods, and offer safe and easy opportunities for people to access public facilities and jobs. Since 2011, the overall strategic plan for London sets out an integrated economic, environmental, transport and social framework for the development of London over the next 20–25 years (Greater London Authority, 2021).

This study explores the draft of the new statutory spatial development strategy for Greater London, particularly focused on the Green-Belt area. This project took place between December 2017 and March 2018 and received many consultation responses from various sources, categorized into eight main consultation groups: London Borough, Authorities outside London, Professional Bodies, Campaign Groups, Government and Agencies, Community Groups, Business, and Individuals. The consultations are publicly accessible online (London Assembly, 2017), and each group offers a wide variety of social-demographic response data, with responses for each group ranging from 40 to 1000. Table 1 summarizes the statistics on the consultation data collected.

3 Framework to Assist Urban Planners

The overall framework consists of the following steps: (i) framing of the urban issues of interest, (ii) collecting and preprocessing the textual data, (iii) extracting topics from the consultation re-

Table 1: Data statistics for the documents of each consultation group

Consultation Group	Number of Documents	Avg. Number of Word Tokens	Number of Tokens	Vocabulary Size
Authorities outside London	54	1667	90,022	478
Borough	40	14,954	598,145	1468
Businesses	363	5047	1,832,125	2557
Campaign Groups	134	3755	503,132	1691
Community Groups	135	2791	376,746	1431
Government and Agencies	77	4754	366,066	1103
Individuals	958	262	251,112	981
Professional Bodies	72	6321	455,091	1465

Table 2: Identified topics regarding the urban planners' issue 'levels and scale'. Communities can encounter Green Belt policy when planning applications are submitted for site, which presents a lack of understanding of how places and people are directly affected at a local level or village/town. For a given issue described by the urban planner (e.g., *levels and scale*), the table reports the topics extracted from two consultation groups, encoded by neural language model, and matched to the issue.

Consulted Groups		Topics of Interest
Businesses	B.1	centre town area office use caz support growth development business
	B.2	policy, development, use, plan, support, land, space, site, area, need
	B.3	cost, value, development, profit, finance, disposal, plan, fee, rent, build
	B.4	housing, plan, home, need, rent, market, year, student, target, delivery
Campaign	C.1	space, venue, child, play, biodiversity, area, include, site, sport, community
	C.2	woman, food, sector, issue, community, year, centre, work, organisation, area
	C.3	com, city, science, health, www, poverty, inequality, job, document, report
	C.4	people, gender, woman, community, feel, year, lgbtqi, experience, queer, identity

Table 3: Example of topics extracted from the *campaign* group

#	Topics
1	space, venue, child, play, biodiversity, area, include, site, sport, community
2	tree, cent, exhibition, product, policy, court, page, borough, forest, woodland
3	woman, food, sector, issue, community, year, centre, work, organisation, area
4	com, city, science, health, www, poverty, inequality, job, document, report
5	people, gender, woman, community, feel, year, lgbtqi, experience, queer, identity
6	land, bed, flat, site, use, value, child, nhs, development, number
7	policy, plan, support, development, need, space, community, use, infrastructure, transport
8	housing, home, rent, need, build, plan, policy, target, year, income
9	energy, building, carbon, use, heat, emission, design, development, network, policy
10	policy, planning, plan, development, land, area, building, site, exist, protect

sponses, (iv) matching the issue of interest with the most related topics. The process is graphically synthesized in Figure 1.

Framing the Planner Issues. First, we provide a concise description of the urban issues as input to the framework, analogously to what an urban planner would do when interested on specific problems related to the Green Belt. As a case study, we consider the problem of *level and scale*, arising when stakeholders submitting applications for site, without taking into account the Green Belt policy on small and large communities and places. The description used is reported in Table 2.

Topic Extraction. Simultaneously, we collect and provide the model with the textual data to be analysed. In particular, we employ the aforementioned consultation responses on the Green Belt, organized by sources. For each consultation group, we extract the main topics discussed by means of the Latent Dirichlet Allocation (LDA; Blei et al. 2003). This process allows the exploration of different concepts being discussed within each of the responses of the consultation groups. For example, Table 3 reports ten topics generated for all the responses in the campaign group consultation, highlighting concisely the main concerns of this group (e.g., child playing, biodiversity, poverty inequality, community space, etc.). To extract the topics, the LDA's hyperparameters (i.e., iterations i, number of topics k, learning decay λ) vary depending on the size of the response group and the theme redundancy across the documents. Thus, they were determined by grid search, comparing the perplexity (inverse of log-likelihood) on small held-out sets of responses. Standard preprocessing steps have been performed to remove the most frequent and rare words, along with English stop words. The top-10 words of each LDA's word-topic distribution are considered representative of a topic.

Encoding and Clustering. In the final step of the framework, we aim at matching (i) the issues of interest described ideally by the urban planner with (ii) the topics pointing out the different concerns raised from various stakeholders (e.g., businesses, local communities, and so on). To perform the matching, we encode into the same mathematical space the issues of interest and the topics via a methodology commonly denominated as *topics* or *sentence embedding*: both the topics from the responses and the sentences of the issues of interest are encoded into vectors of a high-dimensional geometric space using a language model, namely Sentence-BERT (Reimers and Gurevych, 2019). Finally, in this geometrical space of sentence and topic vectors, we run an unsupervised clustering algorithm, namely k-means (Garbade, 2018), to match the issues of interest with the most related topics. The result is a set of matches between the issues of interest and the topics from different consultation groups, similar to the one reported as an example in Table 2, concisely highlighting the most relevant concerns of consultation groups regarding the particular issue of interest.

Qualitative Analysis. Table 2 showcases the LDA topics generated from the *campaign* and *businesses* groups responses, and clustered around the *levels and scale* urban planner issue. For the *campaign* group, the main concerns regard community spaces/venues for children to play, biodiversity, health, inequality regarding jobs, experiences for different gender, sexual identity, and sexuality. This suggests

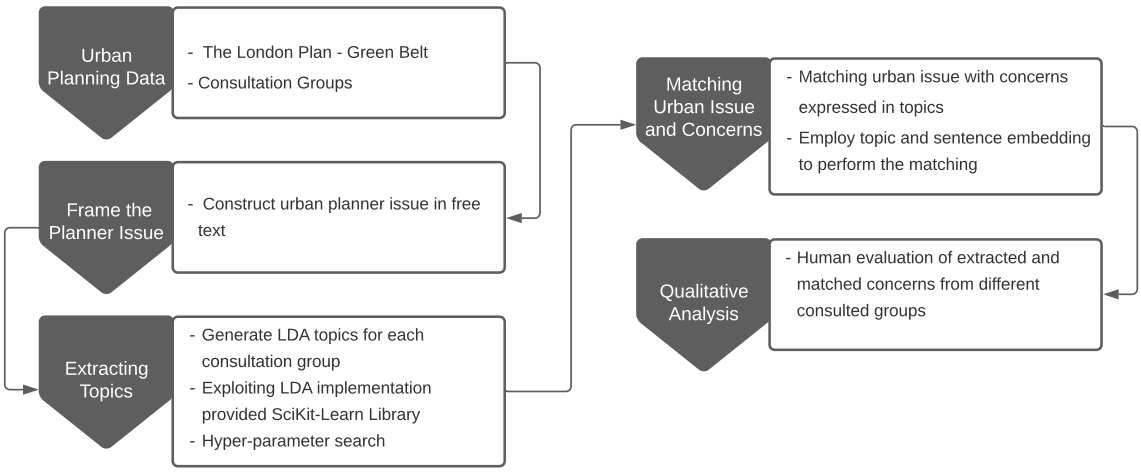

Figure 1: Diagram synthesizing the main steps of the framework

that campaigns' responses are considering those from a 'smaller scale': it appears that the discussions around this issue are themed around diversity problems about their neighbourhood, jobs and community, such as poverty, inequality, and community charities, such as LGBT. On the other side, from the *businesses* group, the main themes clustered around the urban planner issues regarding growth and development for businesses, offices and usage of town centres, cost and profit, need to provide care, student accommodation and making housing targets. These themes suggest that the responses from the businesses consultation group are focused less at a local level, and rather focusing on how Green Belt policies affect people on a larger scale.

4 Conclusion and Future Work

In this preliminary study, we presented an NLP framework in the context of urban planning to assist the exploration and analysis of consultation reports. The framework employed state-of-the-art techniques in NLP, enhancing the accurate identification of stakeholders' concerns about urban issues of interest for the urban planners. We focused our pilot analysis on the Green Belt policy, one of the most relevant ongoing guidelines for the cooperative development of urban spaces and communities in Greater London, UK. A first qualitative analysis has shown the possibility of detecting specific community concerns about urban planner issues described in free text, endowing the planner to quickly detect problems related to minorities, otherwise overlooked due to the vast amount of data.

We conclude by highlighting some of the potential extensions and research lines developing from the presented study:

Perspective Analysis. A first extension to the framework would consist of exploring marginalized groups in more detail by drawing attention to conflicting perspectives that even the same community might have on the same urban planning issues. In particular, we plan to reveal the different perspectives using a Latent Argument Model (LAM) model (Vilares and He, 2017) and perform hierarchical clustering.

Quantitative Evaluation. The consultation responses for the Green Belt policy were manually examined by experts of the council (Greater London Authority, 2017), producing structured reviews of the main themes in them. For example, their reports made available a column denominated 'Topics' that could be directly used to compare the automatically extracted topics of interest. Overall, this structured data could serve as a gold standard to evaluate the results generated automatically by our framework using the human-produced analysis, providing more fine-grained feedback on the effectiveness of the proposed process in easing the urban planner examination.

Author Contributions

C Caton contributed the core paper concept. G Pergola, T Novack, Y He, and C Caton contributed to the paper methodology.

Funding

This research is supported in part by the EPSRC Centre for Doctoral Training in Urban Science and Progress (EP/L016400/1).

ORCID

Tessio Novack https://orcid.org/0000-0003-2817-5972
Yulan He https://orcid.org/0000-0003-3948-5845

References

Blei, David M; Ng, Andrew Y; and Jordan, Michael I: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 3, 2003, 993–1022

Cai, Meng: *Natural language processing for urban research: a systematic review*. Heliyon, 7(3), 2021, e06322. doi: 10.1016/j.heliyon.2021.e06322

Garbade, Michael J: *Understanding K-means clustering in machine learning*. https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1, 2018. Retrieved 1 November 2021

Greater London Authority: *Integrated impact assessment scoping report – summary of consultation responses*. Tech. rep., London, UK, 2017. Retrieved 1 November 2021

—— *The London Plan*. Tech. rep., London, UK, 2021. Retrieved 1 November 2021

Higgins, Marilyn; Hague, Cliff; Prior, Alan; et al.: *Diversity and equality in planning – a good practice guide*. Tech. rep., Office of the Deputy Prime Minister, London, UK, 2005. Retrieved 1 November 2021

Hirschberg, Julia and Manning, Christopher D: *Advances in natural language processing*. Science, 349(6245), 2015, 261–266. doi: 10.1126/science.aaa8685

Klettner, Silvia; Huang, Haosheng; Schmidt, Manuela; and Gartner, Georg: *Crowdsourcing affective responses to space*. KN – Journal of Cartography and Geographic Information, 63(2), 2013, 66–73

London Assembly: *Consultation responses on the draft new London Plan*. https://www.london.gov.uk/what-we-do/planning/london-plan/new-london-plan/consultation-responses-draft-new-london-plan, 2017. Retrieved 1 November 2021

Park, Jae-Gil and Lee, Young A: *The role of urban planning in the process of making livable cities in Korea*. In: Horita, Masahide and Koizumi, Shinichi (eds.), *Innovations in collaborative urban regeneration*, Tokyo: Springer, 2009. 133–146. doi: 10.1007/978-4-431-99264-6_9

Pergola, Gabriele; Gui, Lin; and He, Yulan: *A disentangled adversarial neural topic model for separating opinions from plots in user reviews*. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, 2870–2883. doi: 10.18653/v1/2021.naacl-main.228

Reimers, Nils and Gurevych, Iryna: *Sentence-BERT: sentence embeddings using Siamese BERT-networks*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, 3982–3992

Smit, Warren; Hancock, Trevor; Kumaresen, Jacob; et al.: *Toward a research and action agenda on urban planning/design and health equity in cities in low and middle-income countries*. Journal of Urban Health, 88(5), 2011, 875–885. doi: 10.1007/s11524-011-9605-2

Vilares, David and He, Yulan: *Detecting perspectives in political debates*. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, 1573–1582. doi: 10.18653/v1/D17-1165