JOIV : Int. J. Inform. Visualization, 8(3-2): IT for Global Goals: Building a Sustainable Tomorrow - November 2024 1662-1670

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

PRiSm: Policy Recommendation Systems in Cadastral Survey Using National Public Opinion Big Data

Kihoon Lee [a], Yohan Chang [a,*]

[a] *National Infrastructure & Geospatial Information Research Division, Korea Research Institute for Human Settlements, Sejong-si, Korea*

*Corresponding author: *ycanns@krihs.re.kr*

***Abstract*— Cadastral surveys are vital for assessing individual properties and generating national statistics. In South Korea, rapid territorial changes have exposed personal-level issues, while current systems limited public engagement capabilities have constrained policymaking. Although efforts have been made to bridge this gap, they have been hindered by the lack of an effective medium. This study introduces a novel framework, PRiSm (Policy Recommendation Systems in Cadastral Survey), which leverages National Public Opinion Big Data. We collected and analyzed two key data sources: 1) public opinion data from 2018 to 2023, which correlates strongly with cadastral resurveys across South Korea, and 2) content from 54 major news media outlets over the same period. The first data set represents bottom-up opinions at the individual level, while the second reflects top-down perspectives on national issues. The PRiSm system, developed in this study, utilizes Natural Language Processing (NLP) and advanced Machine Learning (ML) techniques, including Word2Vec and a Genetic Algorithm for hyperparameter optimization, to process over a thousand inquiries and news articles. Our results highlight how different groups engage in discussions shaped by their interests and concerns, revealing key sensitivities and recommending terms invaluable for stakeholders and policymakers. We anticipate that PRiSm will offer meaningful insights for the public and decision-makers. Additionally, with more advanced ML and/or Deep Learning algorithms, there is significant potential for further advancements in NLP within the PRiSm framework.**

***Keywords*— Cadastral survey; public opinion big data; natural language processing; policy recommendation systems; prediction.**

Manuscript received 20 Apr. 2024; revised 9 Jul. 2024; accepted 12 Sep. 2024. Date of publication 30 Nov. 2024.

I. INTRODUCTION

Cadastral surveys are quite critical in South Korea, impacting both individuals and governmental operations by directly influencing property rights. Since 2012, approximately 861,076 lots, representing 15.6% of 5,535,971, have been processed by 2022 [1]. Despite significant advancements in cadastral processing, various concerns have arisen due to limited communication channels available to the government. Key among these concerns is the determination of lot boundaries, which frequently escalates into related secondary issues such as legal disputes, administrative challenges, and planning complications [2]-[4]. These issues often lead to protracted legal battles, wasting time and financial resources and inhibiting cadastral updates' progress.

The government frequently and routinely needs more resources to capture public opinion, which remains underrepresented in policy discussions [5]. This worsens a negative feedback loop, hindering positive advancements in cadastral management [6]. Despite the availability of extensive resources that could utilize public opinions—ranging from individual complaints or suggestions to nationwide media coverage—the integration of this data into cadastral survey processes has been neglected [7].

This study introduces a novel framework named Policy Recommendation System in the cadastral survey, PRiSm, which utilizes the National Public Opinion Big Data (NPOBD) and National News Big Data (NNBD) from South Korea. The NPOBD, managed by the Anti-Corruption and Civil Rights Commission (ACCRC), comprises data collected from diverse sources such as phone calls, texts, smartphone applications, webpages, and in-person documents. Concurrently, the NNBD collects and services national-level news from various media outlets, including major broadcasting networks and local newspapers nationwide. By leveraging these two datasets, PRiSm aims to investigate bottom-up public opinion and top-down media perspectives on cadastral issues.

To analyze these rich textual data, PRiSm employs two natural language processing (NLP) techniques: graph theory-based text mining and Word2Vec(W2V) technique. W2Vis artificial neural network-based supervised fashion machine learning model. The first identifies key themes and examines interactions between public and media contexts, while the latter examines the representation of these themes at national and individual levels. This dual approach can highlight the differences between public interests and current practices and predict potential issues that might arise shortly, offering strategic insights for policy adjustments. Suggested sensitivity information in both numeric and text manners and recommendations for the next plausible issue can be excellent indicators for stakeholders and policy decisions.

Cadastral surveying is a critical but under-researched area due to the difficulties of gathering public opinion and the sensitivity of related policy resolution approaches. The mainstream of existing bodies has been lying on improving the accuracy and economy of cadastral surveys [8]-[11], management of cadaster in a three-dimensional approach [12]–[19], and with particular emphasis on the use of unmanned aerial vehicles (UAVs) equipped with some advanced image processing sensors and technologies. [20]-[25]. This shows a general trend in literature that prioritizes technical improvements over participatory or data-driven approaches.

The first attempt to integrate news media data into the cadastral survey domain was conducted in 2014 [26]. Their study used news media data between 2012 and 2014 and involved extracting major keywords from news articles using Word Cloud and graph theory-based approaches. However, the scope of their research was merely to visualize the relationships among keywords without any advancing methodological frameworks or machine learning-based approaches and extracting and developing a simple policy direction as well [26]. Although these studies demonstrated the potential research expansion of news articles in the cadastral survey domain, they needed to develop a systematic framework to use this data effectively.

Another study explored public opinion data at the jurisdictional level, focusing on term frequency (TF) and inverse document frequency (IDF) techniques to capture the context of public sentiment [27]. By leveraging geospatial information techniques—specifically kernel density analysis, one of the well-known methods in geospatial clustering analysis—their research succeeded in integrating detailed insights in geographically with public opinion data, the findings not commonly explored in broader studies due to the expansive nature of the data complications; their study was able to ample geospatial information in conjunction with public opinion data since subject area in their study was relatively limited.

Recent advances in machine learning and deep learning have stimulated many applications exploring large quantities of text data from diverse sources, including social networks, news data, and survey outcomes [28]–[30]. These developments imply that increasing the value of public opinion data is a significant asset not only for academic research but also for practical policy decision-making [30]-[32]. The growing highlight of such data signifies its potential as a crucial tool across various domains, including cadastral surveying.

This paper is organized as follows: Chapter 2 describes the NPOBD and NNBD datasets in detail and explores the NLP techniques and experimental design employed in this study; Chapters 3 and 4 discuss the findings and propose directions for further research. To the authors' knowledge, this is the first data-driven approach in the cadastral domain to incorporate national public opinion data to develop policy recommendations.

II. Materials and Method

This chapter introduces two primary data sources used in this study and shows how to access, process, utilize, and research methods.

A. National Public Opinion Big Data

The primary data source for this research is the National Public Opinion Big Data, acquired from the e-People service. This service, organized by the Anti-Corruption and Civil Rights Commission, facilitates the submission of individual petitions and collects opinion data nationwide through various channels, including phone, text, smartphone applications, webpages, and in-person documents. The data was queried from another public portal, the Public Data Portal [26], using public API access. To tailor the data for this research, we extracted all contexts that may contain specific cadastral survey-related keywords such as "cadastral," "cadastral (re)survey," "liquidation in cadastral survey," "scale changes in cadastral survey," and "continuous cadastral map." Given the informal nature of the petitions, which often include slang and irregular expressions, a rigorous word correction and refining process was necessary. After removing duplicated contexts, we identified 132 unique inquiries with ample contexts from 2018 to 2023, each comprising more than 500 words. Most of them relate to local jurisdictions in South Korea, and a few of them address central government issues such as proposed laws and legislation. Most petitions originated from urban-rural complex land use areas, reflecting these regions' prevalent rebuilding and planning challenges.

B. National News Big Data

The second source of the data is News Big Data. News media aim to deliver and service unbiased and comprehensive information nationwide. This rich and well-structured text provides valuable insights into various issues and topics. In this manner, news media can be an excellent tool for monitoring and hearing in the public domain [26], [32]. This study used news media data to capture public voices and policy concerns from a broader perspective. The data was obtained from the BigKinds platform [34], which collects daily news media from 54 different media outlets, including broadcasters and newspapers across the country. Using the same keywords and period as for the NPOBD for the equivalent comparisons, we identified 4,542 news articles. Most articles highlighted national-level issues, with some focusing on the need for accurate cadastral surveys. Similar to NPOBD, most news originated from urban-rural complex areas, primarily from local news media.

C. Data Pre-processing and Additional Opinion Data

The unique nature of the petitions allows for the capture of detailed regional voices, although many texts require careful attention due to informal language, typos, and rough expressions. Such characteristics necessitate a secondary text cleaning and refining process, critical for reliable analysis outcomes. Compared to petitions, news media content generally requires little text cleaning as it is formally recorded and well-organized. However, due to the inherent nature of news media, multiple articles might cover the same topic in different regions, introducing potential biases. Additional data were sourced from current practices in eight major cities to mitigate these biases and enhance keyword extraction accuracy. These cities were selected based on their mix of urban and rural characteristics, reflecting the common land use issues observed in the NPOBD. Surveys conducted in these jurisdictions yielded the most frequently discussed keywords, which included 'land boundary,' 'amount of liquidation,' 'property conflicts,' 'conversation,' and 'zoning.'

D. N-grams-based Network Analysis

In PRiSm, we apply two distinct NLP techniques to capture the multi-dimensions of our data sources. One such technique is N-grams-based network analysis. Initially, we conduct a co-occurrence analysis to count the frequency of word pairs within the context, preserving their relational information. This analysis facilitates the construction of a network graph in subsequent steps, where correlations among words are visualized; from the outcome of co-consequences analysis, the network graph can be extracted in the next step by matching the correlations among words. The graph allows us to explore degree centrality, highlighting frequently used words that could indicate critical topics within the inquiries; the more frequently used words in the context may become a clue to show how these words were closely used in the context and presumed as top keywords for the inquiries. Moreover, the phi coefficient can be drawn to present how a pair of words is relatively correlated with other pairs of words, and this can be regarded as an interest against the set of words. For example, consider two words, X and Y; we examine four scenarios: both words present (let denote as a), only X present (let denote as b), only Y present (let denote as c), and neither present (let denote as d). The phi correlation is then calculated as follows [35]:

$$phi\ correlation = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (1)$$

This study extracted a set of sensitivities from words of interest by counting degree centrality and phi correlation from the context. It used open-source R with several core libraries, such as string, KoNLP, dplyr, tidy text, graph, etc.

E. Word2Vec

W2V, which functions based on a Generative Pre-trained Transformer (GPT) architecture, generates vector representations of words, capturing contextual information from their surrounding words [29]. The W2V model utilizes the vector to capture word information from its surrounding words in the context. In technical saying, continuous Bag-of-Words (CBOW) under feedforward neural network architecture enables decreasing computational complexity while continuous Skip-gram architecture increases the probabilities of the outcomes with considering the maximum distance of the words from the context [37], [38]. Generalized objective functions are (2) and (3) for CBOW and Skip-gram, respectively.

$$Maximize\ \frac{1}{T}\sum_{t=1}^{T}\log p(w_t|C_t) \quad (2)$$

$$Maximize\ \frac{1}{T}\sum_{t=1}^{T}\sum_{-c\le j\le c, j\ne 0}\log p(w_{t+j}|w_t) \quad (3)$$

Where T is the total number of words in the word group, commonly called a corpus, c is the size of context for training, w_t is the target word, w_{t+j} and C_t are surrounding or context word(s) around w_t. In this study, we used the W2V model with a Korean tokenizer, Okt, controlling hyper-parameters, such as vector size and window, to have better outcomes. The hyper-parameters were selected simply by applying the Genetic algorithm (GA), one of the most well-known heuristic optimization techniques. The model is implemented under open-source Python with several libraries, such as Natural Language ToolKit (NLTK), genism, etc.

TABLE I
TOP KEYWORDS IN EACH DATA SOURCE

Rank	NPOBD	NNBD	Survey
1	(Re)survey	(Re)survey	**Land boundary**
2	Start construction	Measurement	**Amount of liquidation**
3	**Planning**	District/Zoning	Property conflict(s)
4	**Land boundary**	Land	Conversation
5	**Administration**	**Land boundary**	Zoning
6	Territory/Land	Proceed	
7	Subject	Territory/Land	
8	Announcement	Information	
9	Transportation	Subject	
10	Acts	Empty lot	
11	Subway	Construction	
12	Grade	Reality	
13	(Re)considering	Reduction	
14	Eligible person	Explanatory meeting	
15	Objection	Drone/UAV	
16	Validity	Agreement	
17	Consultation	Transportation	
18	Examination	Digital	
19	Reject	**Administration**	
20	Confliction	Decision	

F. TF-based Keywords Exploring

The exploration of TF-based keywords reveals significant terms across various data sources. Table 1 lists the top 20 keywords from NPOBD, NNBD, and survey data; both keywords from NPOBD and NNBD were obtained after several NLP processing, such as tokenization, stop words, stemming, and lemmatization as a rule of thumb. Notably, the term '(re)survey' ranks highest across in-person petitions and news media, indicating a primary focus on this issue. The distribution of other keywords varies between sources. For instance, 'land boundary' appears in all three within different ranks. In contrast, terms such as 'administration,' 'territory/land,' and 'zoning' appear in only two sources. While NPOBD keywords often relate to personal interests and factors affecting housing prices (e.g., 'transportation,' 'act,' 'subway', 'eligible person'), NNBD keywords are more aligned with urban planning and administrative issues (e.g., 'district/zoning,' 'information,' 'construction', 'agreement'). Emerging cadastral survey tools like 'drone/'Unmanned Aerial Vehicle (UAV)' and 'digital' are also featured prominently in the rankings. Although this outcome was based on the TF order for each data source, the order of keywords for each data source may represent distinct interests and understandings.

G. Research Design

This section delineates the methodologies used in this study, PRiSm, a novel framework for policy recommendation in cadastral surveys. PRiSm aims to capture public opinion from detailed personal experiences to broader societal perspectives, offering diverse insights that stakeholders can utilize. Following the general flowchart in Fig. 1, this study incorporates data cleaning, filtering, and refining processes for each data source.

Basic processes such as data cleaning, filtering, and refining were performed upon holding each data source. The next step is exploring TF-based keywords. In this step, the petition-based top keyword set, namely KW_A, and the news media-based keyword set KW_B, were extracted from their respective contexts. Additionally, the keyword set, namely KW_C, was drawn from the current practice with the surveys mentioned above. This process consists of two distinct approaches: N-grams-based network analysis and W2V analysis. N-grams-based network analysis will create two different graph networks by calculating their centrality values from the NPOBD and NNBD contexts, respectively, and then check their sensitivity against the combination of keyword sets: KW_A with KW_C and KW_B with KW_C. By calculating sensitivity values, the results can show how the in-person-based and national issue-based opinion groups react differently or similarly, ultimately revealing the gaps between the two data sources. Similar to the first analysis, this analysis will also examine not only the keyword sets from its own context but also cross-keyword sets from different contexts: i.e., NPOBD with KW_A/C, NPOBD with KW_B/C, NNBD with KW_A/C, and NNBD with KW_B/C. This cross-keyword checking process will demonstrate how trained opinion contexts from different sources differ or are similar among keywords. Ultimately, the results from these two analyses will show how current practices may closely proceed or deviate from personal and community perspectives.

III. Results and Discussions

This study offers a novel and unique framework with diverse public opinion sources to enhance policy recommendations in cadastral surveys. The findings are detailed in the subsequent discussion and visualized in the provided Tables and Figures.

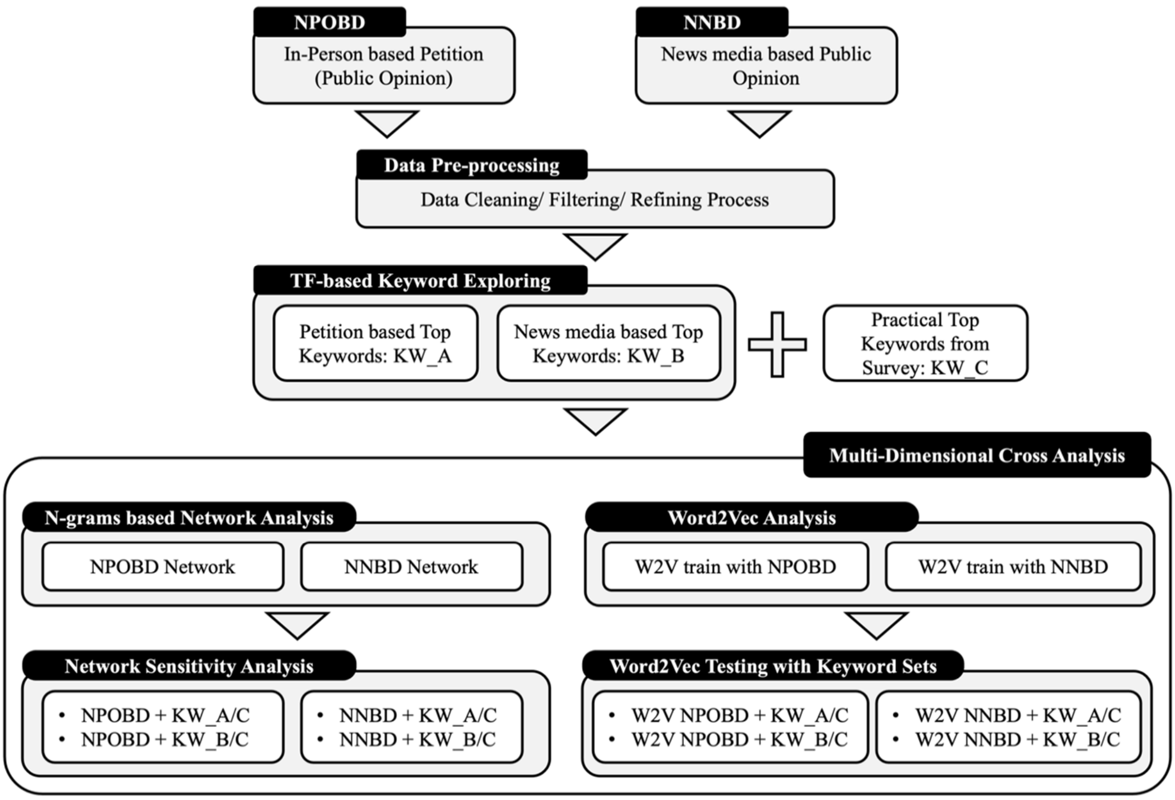
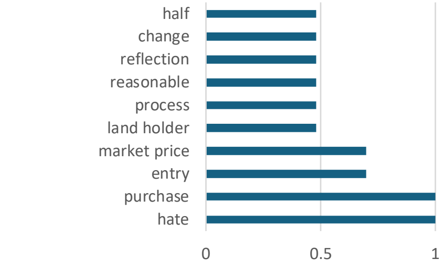
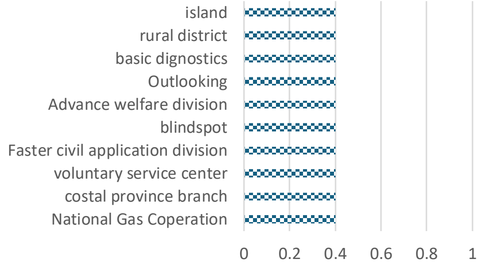
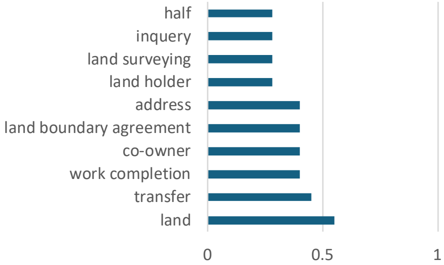
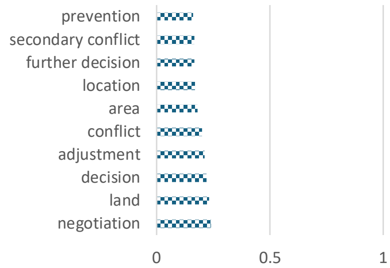
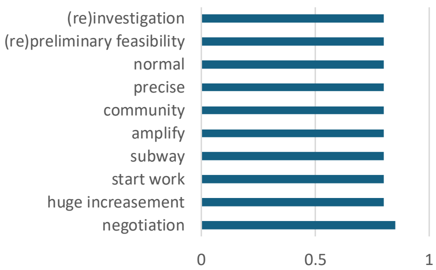
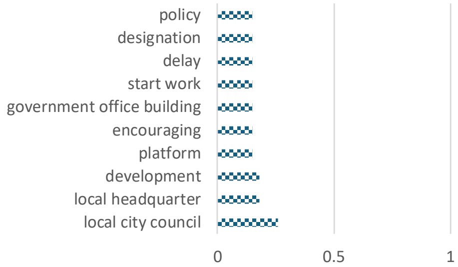
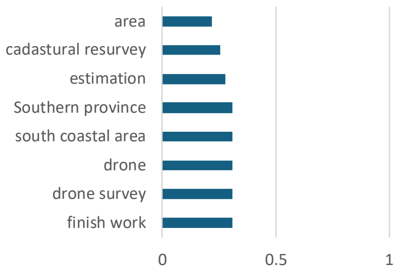
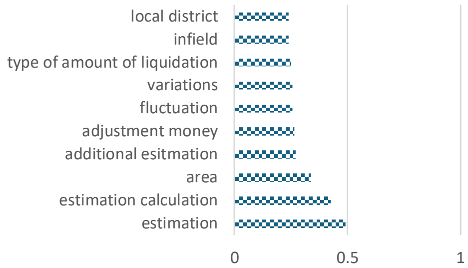


Fig. 1 A general flowchart for PRiSm in this study

TABLE II
THE RESULT OF SENSITIVITY ANALYSIS FROM EACH SOURCE AGAINST MAJOR KEYWORDS

<table>
<tr><th>Keywords</th><th>Sensitivity from NPOBD</th><th>Sensitivity from NNBD</th></tr>
<tr><td>Administration</td><td></td><td></td></tr>
<tr><td>Land boundary</td><td></td><td></td></tr>
<tr><td>Planning</td><td></td><td></td></tr>
<tr><td>Amount of liquidation</td><td></td><td></td></tr>
</table>

A. Gap Analysis between NPOBD and NNBD

The initial results in Table 2 highlight noticeable gaps between the NPOBD and NNBD regarding the same keywords. For example, the keyword 'administration' shows minor sensitivity within NNBD, whereas keywords like 'market price' and 'purchase' display greater sensitivity in NPOBD. Interestingly, the emotionally charged keyword 'hate' ranks higher in NPOBD, suggesting that some administrative processes are perceived negatively by the public. In contrast, keywords such as 'island,' 'rural district,' and 'branch in the province,' and 'blind spot' and 'welfare' are more prominent in NNBD, indicating diverse regional interests captured through public newspapers.

Another key term, 'amount of liquidation,' shows distinct reactions between the two datasets. In NPOBD, it is associated with terms like 'work finish,' 'precise (re)survey with drone/UAV,' and 'coastal province areas,' while in NNBD, it correlates more with financial terms such as 'estimation,' 'estimation calculation,' and 'variation/fluctuation.'

B. Graph Network-based Observations

Graph network analyses (Fig. 2 and 3) reveal further distinctions between NPOBD and NNBD. NPOBD displays a concise unity in centralities and word groups with only a few deviations from the major keyword clusters. Conversely, NNBD exhibits dispersed keyword groups with clearly distinguished centralities. This pattern mirrors the sensitivity differences observed; NPOBD presents diverse and higher sensitivities against primary keywords as Fig. 2 deficits, whereas NNBD shows relatively lower sensitivities and a broader keyword diversity (see Fig. 3).

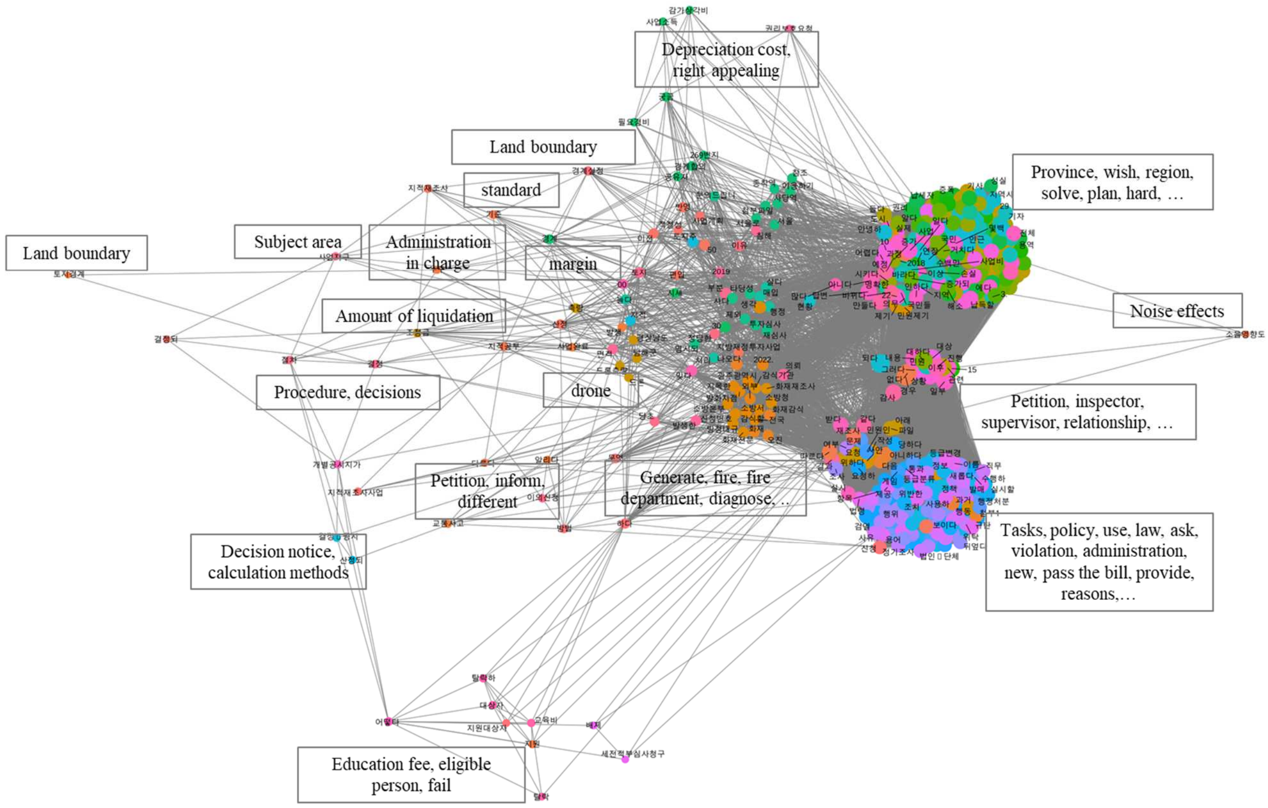


Fig. 2 Graph network visualization for NPOBD against major keyword set

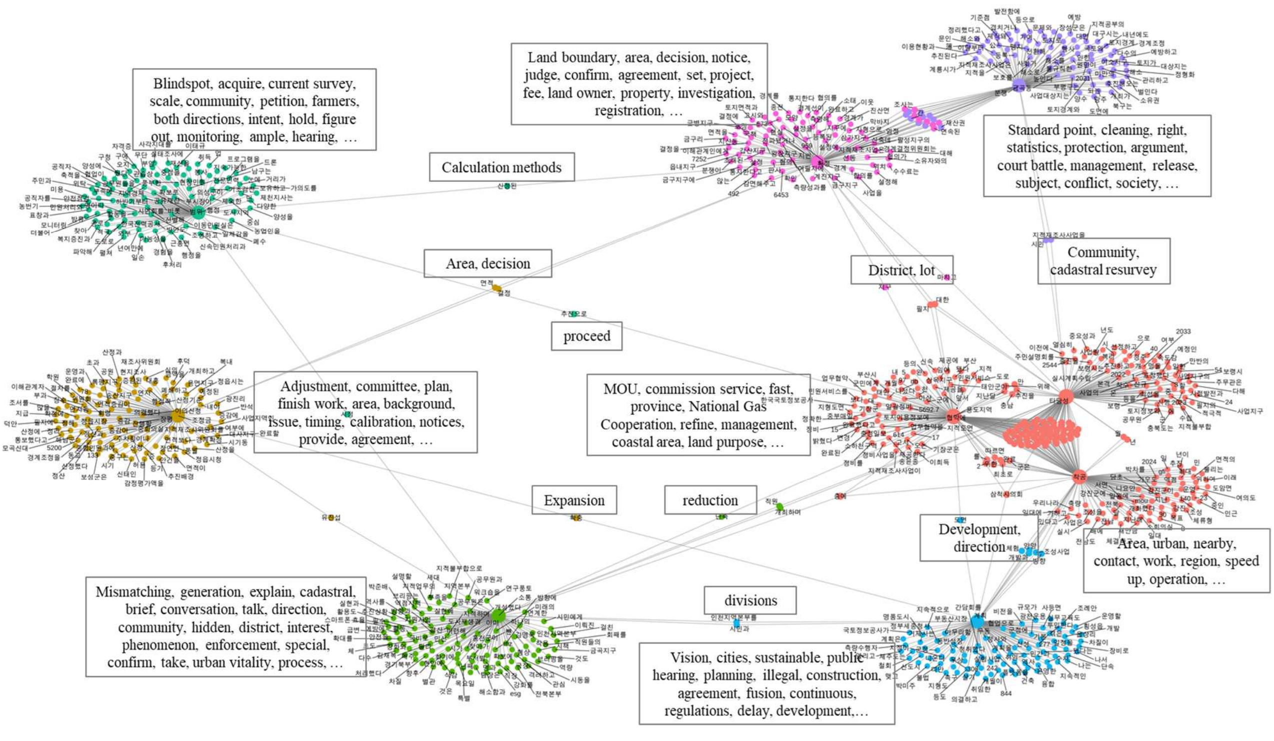


Fig. 3 Graph network visualization for NNBD against major keyword set

These differences likely stem from the distinct nature of in-person concerns versus nationwide issues and the differences in context between individual petitions and professionally written journalistic content. In the context-specific analysis, NPOBD's network visualizations are closely related to themes like wealth, land boundaries, costs, policies, and laws, suggesting that most petitions are concerned with personal interests. For NNBD, the dominant keyword groups relate to urban versus rural dynamics, memorandums of understandings (MOUs) among regions, institutions, and broader public hearing issues, reflecting regional distinctions and higher-level concerns. To be specific, distinguished keyword groups may represent each region, either urban or rural districts, and major keywords such as conflict, property, feasibility study, MOU, an objection, and subject land area led to those keywords as each unique group. NNBD touches on national or/and higher levels of concerns, while NPOBD implies property and issues from the personal level.

C. Word2Vec Outcomes

The outcomes from the W2V analyses are presented in Tables 3 and 4. Table 3 highlights the top five W2V results, designed to capture the most similar keywords from each dataset's top keyword group (i.e., train the W2V model using NPOBD and test its keyword in KW_A/C or train the model using NNBD and test KW_B/C). Table 4 shows the cross-testing results (i.e., train W2V model using NPOBD and test keyword from the cross data, KW_B/C, or train model using NNBD and test KW_A/C). Each table set of keywords was carefully selected based on their uniqueness and notable results. Some keywords, such as 'District/Zoning,' 'Planning,' and '(Re)survey,' can be tracked within the table and across tables to observe how their original W2V and cross-W2V outcomes differ or are similar.

TABLE III
SOME OF W2V RESULTS FOR TOP KEYWORDS IN EACH DATA SOURCE

NPOBD		NNBD	
Keyword	**Words Similarity Top5**	**Keyword**	**Words Similarity Top5**
(Re)survey	Resurvey Law Proceed Administration Change	(Re)survey	(Digital) Transform Digital cadastral Realization Digitalization Paper
Planning	Correlation Consistent Procedure Announcement Change	District/ Zoning	Subject area Land boundary Lot Abandoned area Petition
Land boundary	Target boundary Community Laws Means Budget	Land boundary	Fill information Renewal Long range plan Confirm Individual
Administration	Subject Resurvey Regulation Budget Case	Information	Development Geospatial Program/ System Data Expert

Note: Words are presented in similar order by W2V models.

In Table 3 each data source showed different similarity outcomes; in other words, the similarity words were extracted from the W2V model trained from each dataset by presenting the word that would be the most plausible next word in probability fashion. For instance, while NPOBD returned procedure-related and current-related terms like 'law,' 'proceed,' 'administration,' and 'budget' for keywords such as '(re)survey' and 'land boundary.' At the same time, NNBD suggested future-oriented terms for a cadastral survey like 'digital,' 'transform,' 'digital cadastral,' and 'realization.' For the rest of the words which have quite unique or similar words between the two data sources, the word groups that are highly correlated with policy, earning, and laws in the NPOBD case, while management and direction of cadastral survey word groups were returned in the NNBD case. This outcome shows that each context has unique characteristics that follow each interest. Petition-based NPOBD includes the most needed but hard-to-catch topics on a personal level: 'law,' 'administration,' 'procedure,' 'budget,' and 'cases.' On the other hand, NNBD holds several advice pieces of on which cadastral survey should move or/and change in terms of process and method, which are 'digital,' 'renewal,' 'system,' and 'data'.

TABLE III
SOME OF W2V RESULTS FOR CROSS-SENSITIVITY IN EACH DATA SOURCE

NPOBD		NNBD	
Keyword	**Words Similarity Top5**	**Keyword**	**Words Similarity Top5**
District/ Zoning	Task plan (Subject area) Basic survey (Land boundary) Cognition (Long-range plan) Announcement (Confirm) Duration (Individual)	Planning	Expand (Correlation) Result (Consistent) Revision (Procedure) Installation (Announcement) Deal (Change)
Information	Budget (Development) Institution (Geospatial) Urban (Program/System) Cognition (Data) Entire (Expert)	Subway	Receive (New town) Contrast (Administration investigation) Judgment (Service) Nominal (List) Rural/Urban (Change)
Drone/ UAV	Methods (Adopt) Resurvey (Future) Budget (Diversity) Petition (National) Answer (Development)	Grade	Hometown (district) Election (committee) Best (Needs) Public institution (Process) Software (Budget)
Subject	People (District) Inspection (Area) Context (Individual) Change (Proceed) Cognition (All subjects)	Eligible person	Reduce (Development) Inspection (Area) Process (Agreement) System (Shortage) Support (Registration)

Note: Words are presented in similar order by W2V models. Words in brackets are original similarity words from each data source

Table 4, showcasing cross-testing results, illustrates that each model yields somewhat similar or contrasting word groups across the tests. For example, the test word 'planning' elicited similar responses between NPOBD and NNBD, with terms like 'expand'-'correlation,' 'result'-'consistent,' 'revision'-'procedure,' and 'deal'-'change.' Note that the bracket word is the original similarity word from each data source. These results demonstrate how each data group extends discussions based on their interests. In petitions, NPOBD strongly correlates with personal concerns and financial implications, while NNBD reflects broader national and regional trends in community-level discourse.

IV. Conclusion

Cadastral surveys play a crucial role at multiple levels: they impact property rights on a personal scale, urban planning on a regional scale, and territorial management on a national scale. Despite their importance, efforts to guide, plan, and involve the public in cadastral surveys have been limited. This paper addresses the complex interplay between stakeholder achievements and community voices, which must often align smoothly.

Through this study, we have harnessed public sources such as petition data and news media records to capture a wide spectrum of public opinions. Petition data, oriented towards personal needs, reflects micro-level interests, while news media data encompasses macro-level issues pertinent to cadastral surveys. Our novel framework, the policy recommendation system in the cadastral survey (PRiSm), effectively utilizes these diverse data sources to explore and reconcile these varying interests. PRiSm highlights the diverse reactions across different interest groups and employs a cross-keyword-checking approach to uncover underlying patterns.

To underpin this framework, we adapted natural language processing techniques, including text mining with graph theory and the Word2Vec machine learning model. These methods have yielded significant insights, revealing word sensitivities and recommended terms that are invaluable for stakeholders, policy decisions, and individual preferences. PRiSm, therefore, stands as a powerful tool for preliminary studies of public opinion and for monitoring unexplored agendas and plans.

Furthermore, the machine learning models used in this study, such as the Word2Vec, offer the potential for extension into other applications, including Transformer-based models like GPT [39]. This research has also demonstrated the substantial computational demands of text mining processes, including those involving graph theory. We anticipate that future advancements in computational methods will enable quicker and more precise results, enhancing the effectiveness of frameworks like PRiSm in engaging with and interpreting complex public data [40].

References

[1] H. J. Lee. "A Study on the Designate of Responsible Agency for Cadastral Resurvey," *Journal of the Korean Society of Cadastre*, vol.39, no.3, pp. 131-114, 2023.

[2] G. M. Shin, "A Study on Legal Relationship When Land for Sale is a Cadastral Disagreement Land," *Real Estate Law Review*. vol.24, no.2, pp.1-29, 2020, doi: 10.32989/rel.2020.24.2.1.

[3] J. H. Choi and J. W. Lee, "Research on the Improvement Method for Liquidation Amount Calculation System of Cadstral Resurvey: Comparison of Expert Perceptions," *Appraisal studies,* vol.21, no.3, pp.33-57, 2022, doi: 10.23843/as.21.3.2.

[4] Y. J. Lee, "A Study on the Application Plan of RPA for Cadastral Resurvey Projects," *Journal of the Korean Society of Cadastre,* vol.37, no.3, pp.1-18, 2021, doi: 10.22988/ksc.2021.37.3.001.

[5] P. Krigsholm, K. Riekkinen and P. Ståhle. "The Changing Uses of Cadastral Information: A User-Driven Case Study," *Land*, vol.7, no.83, 2018, doi: 10.3390/land7030083.

[6] C. Daugbjerg and A. Kay, "Policy feedback and pathways: when change leads to endurance and continuity to change," *Policy Sciences,* vol.53, pp.253-268, 2020, doi:10.1007/s11077-019-09366-y.

[7] K. H. Lee and D. H. Kim, "Analysis of Research Trends in the Cadastral Field Using Topic Modeling," *Journal of the Korean Society of Cadastre*, vol.40, no.2, pp.71-82, 2024, doi:10.22988/ksc.2024.40.2.006.

[8] G. Maria, P. Chryssy and I. Charalabos, "A technical solution for 3D crowdsourced cadastral surveys," *Land Use Policy,* vol.98, 2020, doi:10.1016/j.landusepol.2019.104419.

[9] D. Paolo, "The usability of GNSS mass-market receivers for cadastral surveys considering RTK and NRTK techniques," *Geodesy and Geodynamics,* vol.10, no.4, pp. 282-289, 2019, doi:10.1016/j.geog.2019.04.006.

[10] Z. Sandra and N. Gerhard, "Cadastral Surveys Using Terrestrial Laser Scanning – Accuracy and Economy," *Geodetski Vestnik,* vol.67, no.4, pp. 473-486, 2023, doi: 10.15292/geodetski-vestnik.2023.04.473-486.

[11] K. H. Lee, C. Y. Baek and S. H. Kwon, "A Study on Cadastral Surveying of Forest Areas Using GNSS Methods, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry, and Cartography,* vol.42, no.4, 2024, pp.256-264, doi:10.7848/ksgpc.2024.42.4.253.

[12] G. Dogus, "3D modelling of subsurface legal spaces and boundries for 3D ladn administration," *Tunneling and Underground Space Technoogy*, vol. 152, 105956, 2024, doi: 10.1016/j.tust.2024.105956.

[13] S. M. Kim and J. Heo, "Development of 3D underground cadastral data model in Korea: Based on land administration domain model," *Land Use Policy,* vol.60, pp. 123-138, 2017, doi:10.1016/j.landusepol.2016.10.020.

[14] M. P. Jesper and P. Jenny, "Trends in 3D cadaster – A literature survey," *Land use Policy,* vol.131, 106716, 2023. doi:10.1016/j.landusepol.2023.106716.

[15] N. Vučić, M. Mađer, S. Vranić and M. Roić, "Initial 3D cadaster registration by cadastral resurvey in the Republic of Croatia," *Land Use Policy,* vol.98, 104335, 2020, doi:10.1016/j.landusepol.2019.104335.

[16] S. M. Kim and J. Heo, "Registration of 3D underground parcel in Korean cadastral system," *Cities*, vol. 89, pp.105-119, 2019, doi:10.1016/j.cities.2019.01.027.

[17] R. Hajji, H. E. Asri and C. E. Zriouli, "Upgrading to 3D cadaster in Morocco: Lessons learned from benchmarking of international 3D cadastral systems," *Land Use Policy*, vol 128, 106605, 2023, doi:10.1016/j.landusepol.2023.106605.

[18] G. Maria and P. Chryssy, "3D crowdsourced parametric cadastral mapping: Pathways integrating BIM/IFC crowdsourced adata and LADM," *Land Use Policy*, vol. 131, 106713, 2023, doi:10.1016/j.landusepol.2023.106713.

[19] A. Behnam, R. Abbas and O. Hamed, "Proposing a multi-jurisdictional framework for 3D digital cadaster in Australia and New Zealand," *Land Use Policy*, vol.131, 106714, 2023, doi:10.1016/j.landusepol.2023.106714.

[20] F. C. Yu, C. P. Chung, H. Y. Lee and C. P. Chen. "UAS Aerial Surveying in Resurvey Process of Cadastral Maps," 2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2023, doi:10.1109/ECICE59523.2023.10383170.

[21] C. potsiou, N. Doulamis, N. Bakalos, M. Gkeli, C. Ioannidis, and S. Markouizou. "A Prototype Machine Learning Tool Aiming to Support 3D Crowdsourced Cadastral Surveying of Self-Made Cities." *Land* ,vol.12, no.1, 2023, doi: 10.3390/land12010008.

[22] B. Tareke, M. Koeva and C. Persello. "Extracting Polygons of Visible Cadastral Boundaries Using Deep Learning," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, doi:10.1109/IGARSS52108.2023.10282644.

[23] V. Ostankovich and I. Afanasyev. "Illegal Buildings Detection from Satellite Images using GoogLeNet and Cadastral Map," 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 2018, doi: 10.1109/IS.2018.8710565.

[24] R. Petitpierre and P. Guhennec. "Effective annotation for the automatic vectorization of cadastral maps," *Digital Scholarship in the Humanities*, vol.38, no.3, 2023, doi: 10.1093/llc/fqad006.

[25] A. R. Song, S. A. Park and Y. I. Kim, "Updating cadastral maps using deep convolutional networks and hyperspectral imaging," *Paper presented at 40th Asian Conference on Remote Sensing: Progress of Remote Sensing Technology for Smart Future,* ACRS 2019, 2019.

[26] Y. J. Joo and D. H. Kim. "The Big Data Analytics Regarding the Cadastral Resurvey News Articles," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol.32, no.6, pp.651-659, 2014, doi: 10.7848/ksgpc.2014.32.6.651.

[27] T. I. Cho, B. G. Choi, Y. W. Na, Y. S. Moon, and S. H. Kim, "A Suggestion for Spatiotemporal Analysis Model of Complaints on Officially Assessed Land Price by Big Data Mining," *Journal of Cadastre & Land InformatiX,* vol.48, no.2, pp.79-98, 2018, doi:10.22640/lxsiri.2018.48.2.79.

[28] L. Ma, Y. Zhang. "Using Word2Vec to process big text data." In2015 IEEE International Conference on Big Data (Big Data), IEEE Oct. 29, pp. 2895-2897, 2015, doi: 10.1109/BigData.2015.7364114.

[29] S. Sivakumar, L.S. Videla, T.R. Kumar, J. Nagaraj, S. Itnal, and D. Haritha. "Review on word2vec word embedding neural net." 2020 international conference on smart electronics and communication (ICOSEC). IEEE, 2020, doi: 10.1109/ICOSEC49089.2020.9215319.

[30] Z. Wang, M. Wang, H. Shen, and Y. Han. "Application of Sentiment Classification of Weibo Comments Based on TextCNN Model." In 2022 International Conference on Computer Network, Electronic and Automation (ICCNEA), IEEE, pp. 223-228, 2022, doi:10.1109/ICCNEA57056.2022.00057.

[31] Y. Luo, F. Wang, F. Zhao, J. Guo, L. Wang, Y. Hao, and D.D. Zeng. "A Framework for Policy Information Popularity Prediction in New Media," 2019 IEEE International Conference on Intelligence and Security Informatics (ISI), Shenzhen, China, pp. 209-211, 2019, doi:10.1109/ISI.2019.8823415.

[32] A. Shuklin, D. Parygin, A. Gurtyakov, O. Savina and N. Rashevskiy, "Synthetic News as a Tool for Evaluating Urban Area Development Policies," 2022 International Conference on Engineering and Emerging Technologies (ICEET), Kuala Lumpur, Malaysia, pp. 1-6, 2022, doi: 10.1109/ICEET56468.2022.10007405.

[33] "Public Data Portal," data.go.kr, [Online]. Available: https://www.data.go.kr/en/index.do. [Accessed: Jul. 9, 2024].

[34] "Bigkinds," bigkinds.or.kr, [Online]. Available: https://www.bigkinds.or.kr. [Accessed: Jul. 9, 2024].

[35] P. Kalgotra, R. Sharda, and A. Luse. "Which similarity measure to use in network analysis: Impact of sample size on phi correlation coefficient and Ochiai index." *International Journal of Information Management*, vol.*55*, p.102229, 2020, doi:10.1016/j.ijinfomgt.2020.102229.

[36] T. Mikolov, K. Chen, G. Corrado, and J. Dean. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781, 2013, 10.48550/arXiv.1301.3781.

[37] X. Rong, "word2vec parameter learning explained." arXiv preprint arXiv:1411.2738, 2014, 10.48550/arXiv.1411.2738.

[38] Y. Goldberg, and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." arXiv preprint arXiv:1402.3722, 2014, 10.48550/arXiv.1402.3722.

[39] A. Khan, K. Khan, W. Khan, S. N. Khan, and R. Haq, "Knowledge-based Word Tokenization System for Urdu", Journal of Informatics and Web Engineering, vol. 3, no. 2, pp. 86–97, 2024. doi:10.33093/jiwe.2024.3.2.6.

[40] F. Zaman et al., "Intelligent Abstractive Summarization of Scholarly Publications with Transfer Learning", Journal of Informatics and Web Engineering, vol. 3, no. 3, pp. 256–270, 2024. doi:10.33093/jiwe.2024.3.3.16.