Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Horizon scanning in policy research database with a probabilistic topic model

Hyunuk Kim[a], Sang-Jin Ahn[b,*], Woo-Sung Jung[a,c,*]

[a] Department of Industrial and Management Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea
[b] Office of Strategic Foresight, Korea Institute of Science and Technology Evaluation and Planning (KISTEP), Seoul 06775, Republic of Korea
[c] Department of Physics, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

ARTICLE INFO

Keywords:
Foresight
Horizon scanning
Policy research database
Topic modeling
Latent Dirichlet allocation

ABSTRACT

National governments take advantage of collective intelligence when conducting foresight processes. They grasp emerging issues through expert reviews as well as public opinions. It raises national agendas and affects policy-making process. Therefore, by examining policy papers which contain societal issues, we can perceive past, current, and future environments. In this study, we exploit policy research database of Republic of Korea, which is a unique source that automatically collects all policy papers written by national research institutes, to extract latent topics and their trends over 10 years through a probabilistic topic model. Detected topics fairly correspond to expert-selected future drivers in national foresight report, implying that public discourse and policy agenda are coupled. We suggest to utilize open government data and text mining methods for building open foresight framework that various actors exchange their opinions on societal issues.

1. Introduction

The history of capitalism after World War II can be understood as a reconstruction process of economic system. Foresight led countries to build effective innovation system, focusing on the role of science and technology. Since 1970s when market matured, systemic view has been underlined in designing foresight process (Andersen and Andersen, 2014). It is hard to understand a dynamic environment without considering collective behaviors of actors. Hence, national governments include not only expert reviews but also public opinions to reflect social complexity in foresight. Collective intelligence helps governments to detect weak signals and to shape emerging issues. For successful national foresight, collaboration between departments, external knowledge sources, and credibility of evidences are also needed (Habegger, 2010).

To meet these requirements, we pay attention to open government which enhances transparency and collaborative capacity of national government (Lathrop and Ruma, 2010). Republic of Korea is taken as a subject of study because of its numerous foresight experiences and high open government availability (OECD, 2017). Foresight of Korea has two characteristics. First, it is coherent with STI public policy (Ahn, 2017), leading a success in catch-up development. Preliminary feasibility study is conducted by the government before launching a new national project. Based on *ex ante* evaluation, the Korean government decides whether they invest R&D budget in the project. Second, it heavily relies on expert opinions whereas participation of private stakeholders is limited (Andersen and Andersen, 2017). Experts who are invited by the Korean government run most of foresight activities so that the results could be biased.

Among various open government data of Korea, we exploit National Knowledge Information System (NKIS), a recently opened database that gathers all policy papers of national research institutes in online. As policy-making process is related to foresight (Da Costa et al., 2008; Havas et al., 2010), this unique policy database is suitable for extracting influential future drivers. A probabilistic topic model, Latent Diriclet allocation (LDA), is implemented to detect latent topics in policy research papers published during 2003–2015. We confirm that latent topics in NKIS match with expert-selected issues in national foresight report. It indicates that we can map the current environment as well as emerging issues through topic modeling in policy research database without conducting a series of expert surveys as before. Thus, policy research database can be used as an invaluable source of semi-automated horizon scanning that is built on topic models. Based on the results, we suggest to utilize open government data and text mining methods for developing open foresight platform that actors share their opinions on latent topics and mediate relevant policies.

This paper is organized as follows. Section 2 introduces previous studies on horizon scanning, foresight support system, and open

* Corresponding authors.
E-mail addresses: sein@kistep.re.kr (S.-J. Ahn), wsjung@postech.ac.kr (W.-S. Jung).

https://doi.org/10.1016/j.techfore.2018.02.007
Received 30 September 2016; Received in revised form 20 December 2017; Accepted 7 February 2018

foresight. Detailed information about NKIS, data retrieval procedure, and Latent Dirichlet allocation will be given in Section 3. Latent topics in policy research database and expert-selected issues in national foresight report will be compared in Section 4. Section 5 concludes the paper with implications on a web-based open foresight platform and policy-making process.

2. Background

2.1. Horizon scanning

As a part of foresight, horizon scanning identifies societal trends and potential issues (Schultz, 2006). Practitioners usually collect information by attending conferences and doing iterative expert surveys. They classify scanned issues into one of STEEP (Social, Technological, Economic, Ecological, and Political) fields depending on their characteristics (Habegger, 2010). STEEP framework makes it possible to analyze orientations of future signals (Saritas and Smith, 2011).

While these approaches are frequently used, they have limitations on selecting experts and detecting weak signals (Amanatidou et al., 2012). Foresight experts begin to consider web channels, such as blogs and Twitter, as new information sources (Pang, 2010). Web-based horizon scanning with data mining applications successfully catches weak signals (Palomino et al., 2012; Thorleuchter and den Poel, 2013, 2015; Yoon, 2012) and builds plausible scenarios (Kim et al., 2016).

2.2. Foresight support system

Foresight support system (FSS) is also one of the information and communication technology applications in foresight process. Bañuls and Salmeron (2011) define FSS as a collaborative computer-based system that aims to support communication, decision modeling, and rules of order in foresight processes. Function of statistical and quantitative analysis is also required for FSS (von der Gracht et al., 2015).

FSS can take a variety of forms. It utilizes web forum while it gives less structured information than formal databases (Woo et al., 2015). FSS at sector level (Durst et al., 2015; Keller et al., 2015) receives stakeholders' opinions to identify future drivers and to shape the future of domains. FSS that is built on combination of foresight method and prediction market promotes wisdom-of-crowds effect, achieving a good performance in macroeconomic forecasting (Prokesch et al., 2015).

Table 1
Policy paper category distribution drawn from National Knowledge Information System (NKIS). It is represented by sixteen policy categories and three time periods corresponding the Korean presidential terms. Each cell indicates the number of policy papers. The number of papers missing abstract information is in parentheses.

Category	2003–2007	2008–2012	2013–2015
(A) General Issue	124(2)	850(24)	223(65)
(B) Economy	409(49)	618(74)	649(78)
(C) Industry	214(16)	385(18)	286(18)
(D) Labor	417(24)	564(50)	375(21)
(E) Energy, Natural Resources	135(0)	320(6)	285(7)
(F) Land Development	240(9)	474(11)	369(32)
(G) Transportation	217(3)	338(12)	155(43)
(H) Broadcasting, Telecommunications	302(8)	350(3)	126(0)
(I) Social Problem	344(6)	586(18)	810(65)
(J) Healthcare, Social Welfare	231(7)	559(32)	330(51)
(K) Trade, Diplomacy, National Security	586(72)	655(49)	300(34)
(L) Public Administration, Social Security	543(4)	735(38)	587(17)
(M) Education	622(12)	837(16)	598(32)
(N) Science, Technology	229(4)	225(4)	210(1)
(O) Agriculture, Maritime, Fishery	271(2)	454(7)	493(5)
(P) Environment	179(1)	409(13)	320(3)
The number of papers in total	5063	8359	6166
The number of papers missing abstract	219	375	472
The number of papers having abstract	4844	7984	5694

2.3. Open innovation and open foresight

Open innovation (Chesbrough, 2006) that combines internal capacity and external knowledge for innovation, provides opportunities in business markets (Dodgson et al., 2006; van de Vrande et al., 2009) and public domains (Chan, 2013; Hilgers and Ihl, 2010). To facilitate open innovation, organizational boundaries should be broken up (Yun, 2015; Yun et al., 2016a,b) and macro phenomena originating from microscopic interactions have to be considered (Dougherty, 2017; Witt, 2017).

Based on the concept of open innovation, Daheim and Uerz (2006) suggest a collaborative foresight model 'Open foresight'. It consists of participatory structure, online openness, and incentives to participate (Miemis et al., 2012). Open foresight could be embedded in foresight support system, guaranteeing broader participation of actors. It fosters collective intelligence and enhances the quality of foresight by integrating various opinions (Keller et al., 2015).

3. Research methodology

National research institutes publish policy papers on important issues within society, country, and the world. We backtrack these issues through topic models in policy research database of Korea, National Knowledge Information System (NKIS). As the database exploits cross-organizational knowledge, we can find national interests which lie across many disciplines.

3.1. Data retrieval from NKIS

National Knowledge Information System (NKIS)[1] provides bibliometric information about policy papers written by national research institutes. For each policy research paper, we collect title, abstract, publication year, main author, policy category given by the government, institute, and other miscellaneous items from application programming interface of this database. As national policies are related to politics, in order to track policy changes, we divide the data set by the Korean presidential terms, 2003–2007, 2008–2012, and 2013–2015. Paper distributions by policy categories and presidential terms are in Table 1. Papers of missing abstract (~5%) are excluded in our analysis because topic modeling needs a set of words to infer latent topics. The number of target documents for three periods is 4844, 7984, and 5694. We extract nouns from abstracts, and form a corpus of national policy research through KoNLP package (Jeon, 2013) in statistical language R (R Core Team, 2014). General words, such as 'analysis', 'research', and 'problem', are excluded for obtaining clear topic structures.

3.2. Topic modeling

NKIS policy categories are broad to observe societal issues, which would be hidden and distributed over many disciplines. To get latent topics in policy research database, we implement a probabilistic topic model, latent Dirichlet allocation (LDA) (Blei et al., 2003), which has been used to reveal research trends (Blei, 2012), public agenda in Russian blogs (Koltsova and Koltcov, 2013), and software developers' interests (Barua et al., 2014).

LDA assumes that a text is made by the mixture of topic distribution of a document ($\sim Dir(\alpha)$) and word distribution of a topic ($\sim Dir(\beta)$), where *Dir* is Dirichlet distribution. It infers probability of topics per document ($\theta_{d,k}$) and probability of words per topics ($\phi_{w,k}$) for document d, topic k, and word w. The following conditions $\sum_k \theta_{d,k} = 1$, $\sum_w \phi_{w,k} = 1$ should be held. Topic k's influence, I_k, is estimated as the mean of $\theta_{d,k}$ over d, $I_k = \frac{1}{|D|} \sum_{d \in D} \theta_{d,k}$, where D is a document set. LDA doesn't give topic names so we manually characterize detected topics

[1] https://www.nkis.re.kr:4445/main.do (in Korean, accessed on Oct 31, 2017).

from their top topic words. In the results, latent topics are sorted in decreasing order of their influences.

Intra- and inter-period topical distances are measured through Jensen-Shannon divergence (JSD) over ϕ. JSD between two given distributions P and Q of the same size is calculated as $JSD(P\|Q) = H(0.5P + 0.5Q) - 0.5H(P) - 0.5H(Q)$, where H is Shannon entropy function. Small JSD value indicates that two target distributions are similar. As three predefined periods have different number of documents, we renormalize ϕ values of two successive periods for common words, when calculating inter-period topical distances. We project intra-period topic structure onto two-dimensional principal component space by following LDAvis methodology (Sievert and Shirley, 2014). Inter-period topic structure is drawn as a relation diagram which connects a topic in period 1(2) with another in period 2(3) of which JSD value is minimum among all possible topic pairs. Inter-period topic structure shows topic trends over a decade. Topic structures are drawn by R packages, 'ggplot2' (Wickham, 2009), 'ggrepel' (Slowikowski, 2017), and 'gridExtra' (Auguie, 2017).

To implement LDA, we have to decide hyperparameters (α, β) and the number of topics (K). We use $K = 28$ that is the same to the number of STEEP drivers in national foresight report, Korea Future Issues 2015. There is no consensus on how to set hyperparameters. We give $\alpha = 50/K$ and $\beta = 0.1$ as Griffiths and Steyvers (2004) used when finding research topics. Topic modeling analysis is conducted by *lda.collapsed.gibbs.sampler* function in R package 'lda' (Chang, 2015) with 1000 Gibbs sampler iterations.

4. Topic structure in policy research database

4.1. Intra-period topic structure

Topic structure in period 1 is given in Fig. 1. Education is one of the major topics of the Korean government. Relevant Topics, 1, 3, and 15, are clustered at the right of two-dimensional principal component space. Topics 4 and 18, which are located at the top of the plot, are about transportation planning and its effect on environment. International relation with China and North Korea is highlighted in Topics 11 and 13. Energy issues are found in Topics 24 and 26. At the center of the plot, there are many topics dealing with social and economic issues.

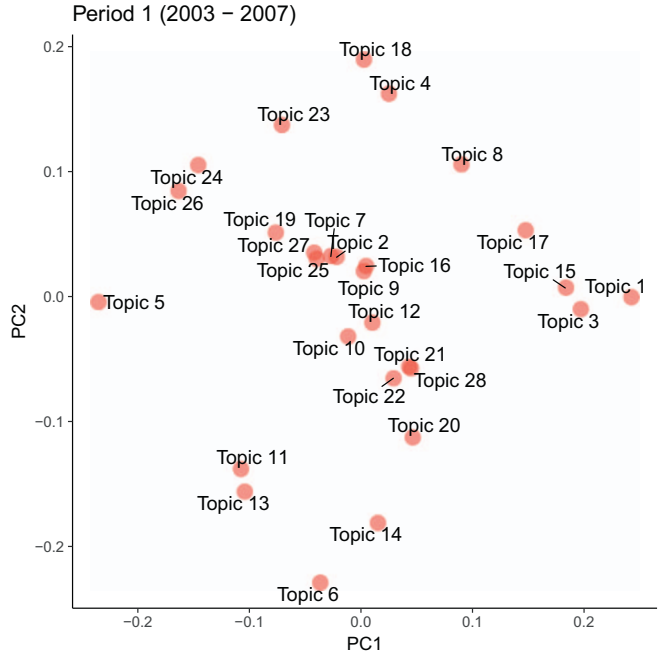
Compared to topic structure in period 1, those in period 2 (Fig. 2) is more dispersed in principal component space. Education (Topic 1) is still an important topic of the Korean government. Topics 10 and 19, at the bottom-left of the plot, are also about education, especially for youth and child. Multicultural family is newly found in Topic 8. Energy and climate change issues are revealed in Topics 6 and 15. Topics 16 and 17 describe geopolitical and economic relations with China and North Korea.

Fig. 3 shows topic structure in period 3. National interest in education is specialized toward early childhood education (Topics 1, 9, 13, 22). Geopolitical topics, Topics 6 and 8, include unification issues with North Korea. Legal issue (Topic 3) comes to the fore and it is entangled with topics of environment (Topic 10, 11), health (Topic 20), and regional development (Topic 21). Other topics are similar with previous policy interests.

4.2. Inter-period topic structure

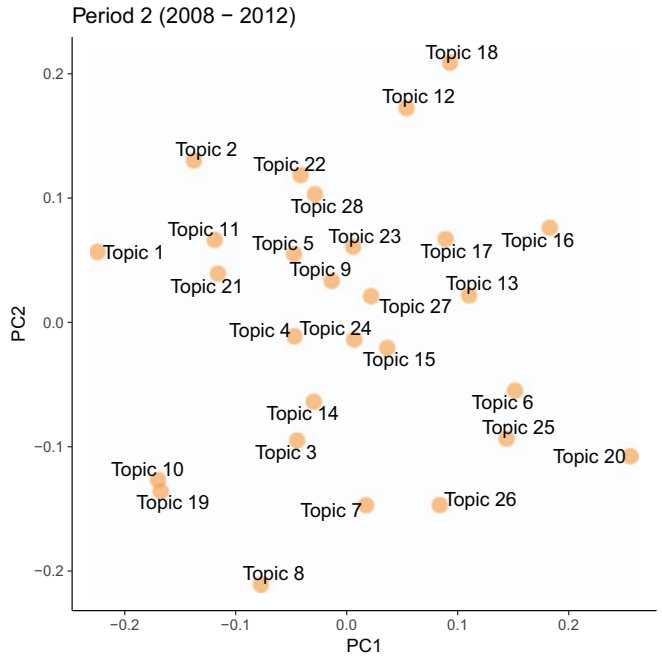
For visualization purpose, we connect a topic in period 1(2) with another in period 2(3) if their JSD topical distance is the smallest among possible pairs (Fig. 4). It means that topic in period 1(2) has only one counterpart in period 2(3). Even though this approach could eliminate some information, it helps to outline topical changes in national policy research. On the other hand, if we extract topic trends by a certain threshold, it is hard to get overall landscape of national interests due to multiple connections within finite topics.

From topic trends that cover two intervals, period 1–period 2 and period 2–period 3, we find emerging topics which are colored with blue. Topics of low rank in the previous period tend to be important in the next period. In period 2, national policy institutes focused on government (Topic 2), energy (Topic 6), family (Topic 8), social integration (Topic 9), information (Topic 11), and social security (Topic 14). In period 3, childcare (Topic 1) becomes the most dominant topic whereas it firstly appears in period 2 as Topic 10. Other emerging topics are about government (Topic 2), geopolitical issues with China and North Korea (Topics 6, 8), industry (Topic 7), education (Topic 9, 13), social welfare (Topics 12, 16), and regional development (Topic 21).



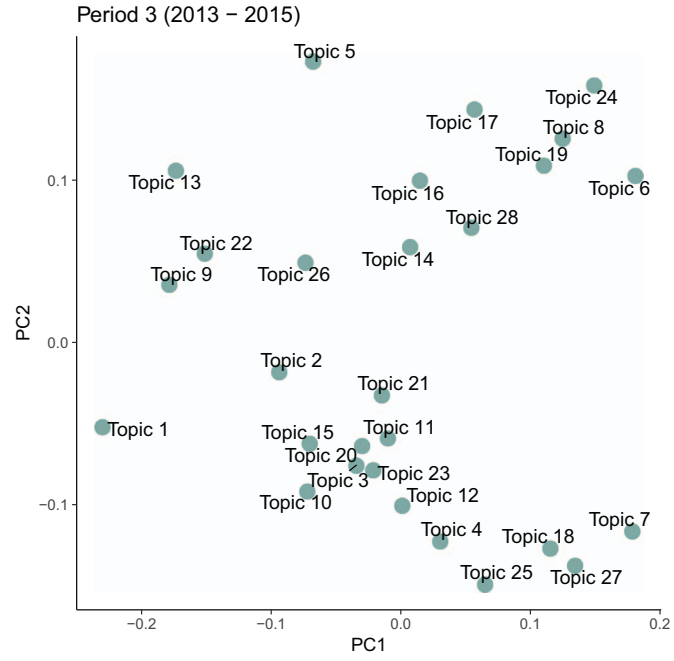
1	Education, School, Student
2	Effect, Estimation, Factor
3	Human Resource, Necessity, Foundation
4	Environment, Management, Effect
5	Industry, Economy, Increase
6	Socity, Civil, Philosophy
7	Employment, Worker, Training
8	Government, Management, System
9	Crime, System, Rule
10	Firm, Technology, Innovation
11	China, International, Country
12	Country, Development, Area
13	North Korea, Peace, Economy
14	Culture, Information, Space
15	University, Ability, Job
16	Region, City, Land
17	Data, Information, Statistics
18	Transportation, Planning, Logistics
19	Market, Service, Competition
20	Youth, Experience, Ratio
21	Female, Family, Discrimination
22	Participation, Organization, Governance
23	Regulation, Cost, Burden
24	Energy, Price, Electricity
25	Income, System, House
26	Agriculture, Increase, Consumption
27	Danger, Occurrence, Safety
28	Police, Activity, Topic

Fig. 1. Topic structure in period 1 (2003–2007). Twenty-eight latent topics are projected onto two-dimensional principal components. Top three words of each topic are listed in the table.



1	Education, School, Teacher
2	Management, Government, Necessity
3	Effect, Factor, Relation
4	Regulation, System, Law
5	City, Space, Planning
6	Energy, Greenhouse Gas, Decrease
7	Employment, Job, Labor Market
8	Female, Family, Multicultural
9	Cuture, Integration, Society
10	Youth, Nurture, Child
11	Data, Information, Statistics
12	Cooperation, Country, International
13	Industry, Firm, Innovation
14	Crime, Criminal, Danger
15	Environment, Climate, Management
16	China, Resource, Economy
17	North Korea, Politics, Economy
18	Growth, Development, Technology
19	Student, School, Achievement
20	Market, Economy, Increase
21	University, Human Resource, Specialty
22	Region, Farm, Connection
23	Fiscal, Government, System
24	Society, Social Welfare, Economy
25	Agriculture, Price, Market
26	Income, Cost, System
27	Transportation, House, Logistics
28	Expert, Topic, Presentation

Fig. 2. Topic structure in period 2 (2008–2012). Twenty-eight latent topics are projected onto two-dimensional principal components. Top three words of each topic are listed in the table.



1	Child, Nurture, Kindergarten
2	Management, System, Government
3	Law, Rule, System
4	Increase, Effect, Decrease
5	Participation, Cooperation, Activity
6	International, Country, China
7	Industry, Firm, Export
8	North Korea, Unification, Economy
9	Student, Learning, School
10	Environment, Effect, Sea
11	Climate, Management, Adoptation
12	Fiscal, Social Welfare, System
13	Education, Capability, Operation
14	Information, System, Supply
15	Employment, Family, Worker
16	Society, Public, Recognition
17	Culture, Value, Approach
18	Energy, Market, Supply
19	Technology, System, Innovation
20	Safety, Management, Health
21	Region, City, Space
22	Youth, Child, Prevention
23	Regulation, Cost, Crime
24	Economy, Development, Growth
25	Finance, Investment, Tax
26	Female, Family, Sex
27	Agriculture, Market, Farm
28	Presentation, Topic, Discussion

Fig. 3. Topic structure in period 3 (2013–2015). Twenty-eight latent topics are projected onto two-dimensional principal components. Top three words of each topic are listed in the table.

4.3. Comparison with STEEP drivers in Korea Future Issues 2015

To confirm the validity of topic modeling in societal issue detection, we compare latent topics in policy research database with expert-selected future drivers in national future report, Korea Future Issues 2015. Korea Future Preparatory Committee published the report in 2015 by surveying 1477 respondents, including professors, experts, and university students, to ask importance, impact, and likelihood of expert-selected drivers (Table 2) in the future. This survey reveals that Low Fertility & Super-Aging Society, Social Inequality, and Unstable Life of Future Generations would be influential in the future.

Korea Future Issues 2015 does not provide word lists which describe expert-selected future issues in detail, so that it is hard to show the relationships with LDA topics. Thus, as an alternative way, we draw a heatmap of future issues and latent topics in period 3 to see their co-incidences. To draw the heatmap of incidence, we firstly find policy papers of which abstract contains the name of future drivers in abstracts. Next, for these papers, we calculate mean θ over twenty-eight latent topics to get topical distributions.

Many future drivers clearly belong to single latent topic (Rows in Fig. 5). For example, Hyper-Connected Society, Aggravating Gender Inequality, and Food Safety match with Topics 16, 26, and 27,

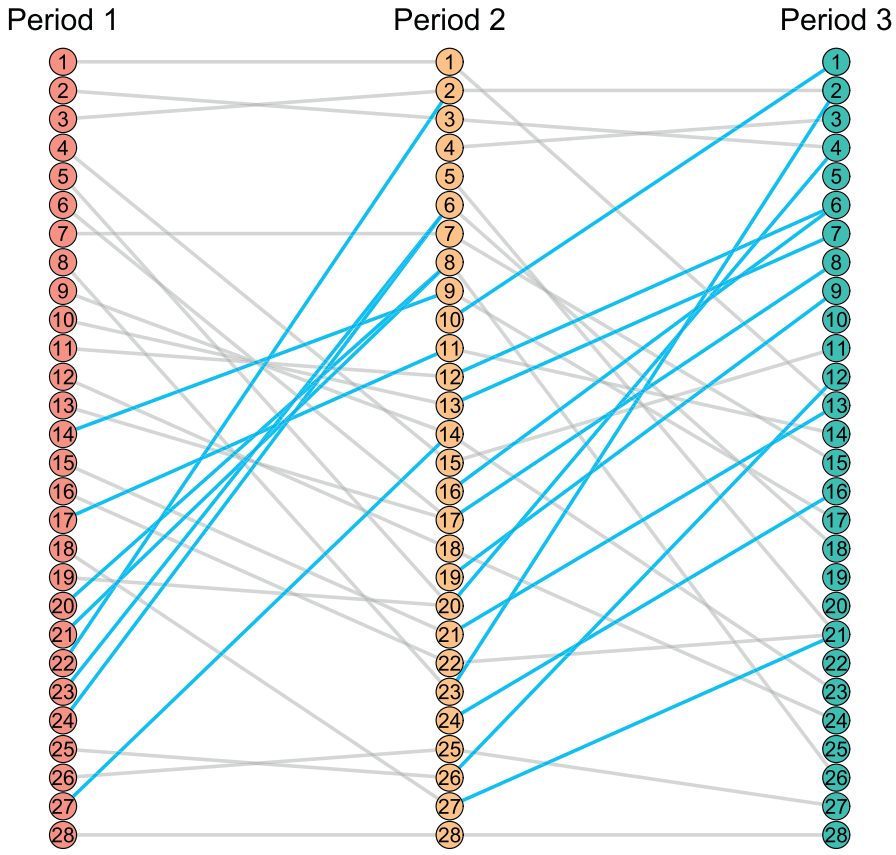


Fig. 4. Inter-period topic trends during 2003–2015. Topic relations which become dominant in the successive period (if a topic moves upward more than five ranks) are highlighted as blue lines. Topics of low rank in the previous period tend to be important in the successive period.

respectively. However, it is noted that some latent topics cover multiple future drivers. For instance, Topic 8 that is characterized by the words such as North Korea, Unification, and Economy is related with Unstable Life of Future Generations, Geopolitical Conflict with Neighboring Countries, and National Security/Unification. It would be due to overlapping nature of expert-selected issues and(or) low-resolution of topic models in this study, which only extracts twenty-eight latent topics that are insufficient to fully reflect the society. In addition, our procedure of drawing heatmap might also be problematic for grasping the overall landscape. It is interesting that urban planning (Topic 21) and taxation (Topic 25) issues are found in policy research database whereas they are not included in Korea Future Issues 2015. Topic modeling catches not only potential future issues but also missing topics in empirical foresight exercises.

5. Discussion and conclusion

In this study, we implement a probabilistic topic model to find important societal issues in policy research database. The effectiveness of topic model in horizon scanning is confirmed by comparing detected topics with expert-selected STEEP drivers in national foresight report. Topic modeling would complement the current foresight schemes, which often rely on manual surveys and reviews, by introducing potential future drivers in a semi-automated manner.

Topic model finds societal issues even though it is applied to three document sets of different periods. This flexibility allows foresight practitioners to implement topic models for specific periods of their interests. Moreover, depending on foresight purposes, we can examine detailed issues in a policy domain, focusing on documents that are classified into a latent topic. Through integration with bibliometric database of national S&T institutes, we can also evaluate national S&T policy and the extent of preparedness for emerging technologies.

We suggest to use topic modeling technique on policy research database for open foresight, a participatory and collaborative foresight concept, to facilitate active engagement of the public, stakeholders, and national government. It would break up the boundaries between actors.

Table 2
Future drivers in national future report of Korea, Korea Future Issues 2015. These drivers are selected by expert reviews and surveys.

Field	Issues
Society (10)	Low Fertility & Super-Aging Society, Social Inequality, Unstable Life of Future Generations, Emphasis on Quality of Life, Multiculturalism, Change in Traditional Family System, Credentialism & Excessive Competition in Education, Aggravating Gender Inequality, Fight Against Incurable Diseases (Homo-Hundred Era), Cyber Crime
Technology (15)	IoT, Big Data, Artificial Intelligence, Virtual Reality, Wearable Device, Stem Cell, Genetic Engineering & Molecular Biology, Synthetic Biology, Nano Material, 3D Printer, New Renewable Energy, GHG Reduction Technology, Energy Resource Recycle Technology, Space Exploration, Nuclear Energy Technology
Economy (6)	Hyper-Connected Society, Low Growth & Shift in Growth Strategies, Digital Economy, Job Insecurity, Manufacturing Revolution, Bipolarized Industrial Structure
Environment (7)	Disaster Risk, Energy Shortage & Resource Depletion, Climate Change & Natural Disasters, Growing Cross-border Environmental Impact, Nuclear Safety, Biodiversity Crisis, Food Safety
Politics (5)	Food Security, Geopolitical Conflict with Neighboring Countries, National Security/Unification, e-Democracy, Global Governance

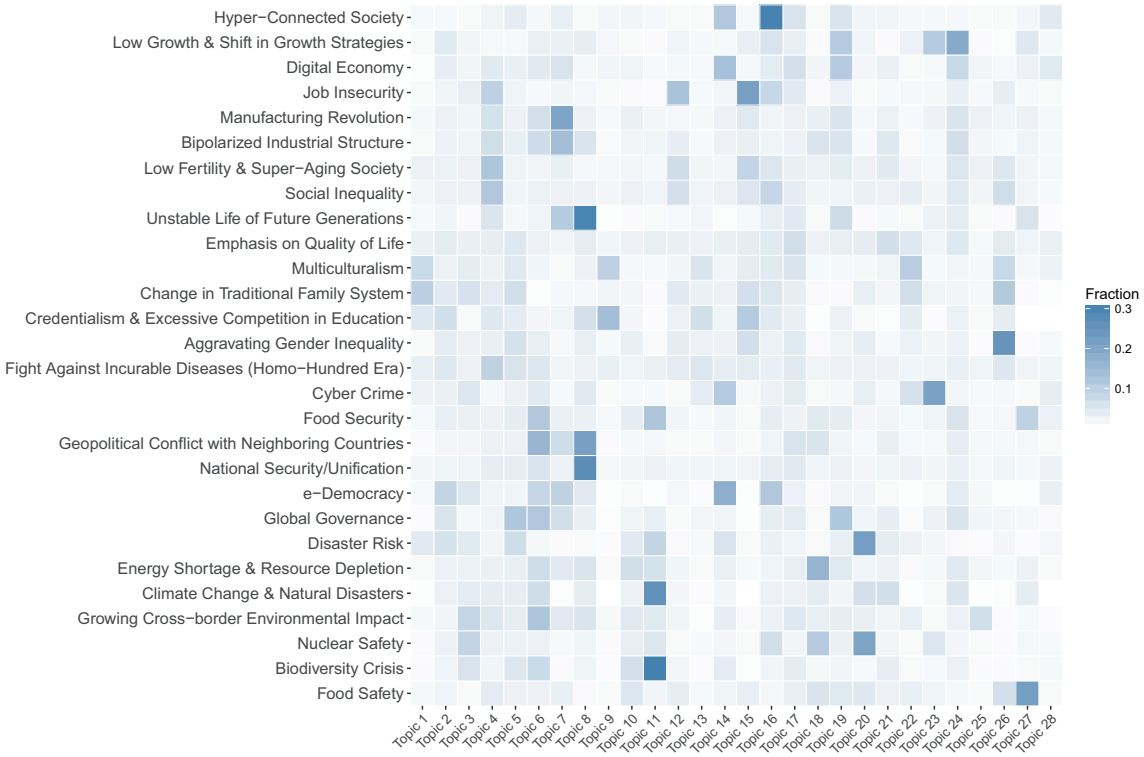


Fig. 5. Heatmap of incidence between latent topics and expert-selected future drivers.

Over latent topics in open policy research database, participants grasp what national interests are so that they could express and adjust their opinions with ease in open government system, affecting policy-making process. During the process, national government can also recognize important issues that they miss. It compensates the time delay of policy research that is late compared to scientific articles or social awareness (Molitor, 1977; Schultz, 2006). Iterative opinion exchanges between participants would detect emerging issues and wild cards. Open foresight platform at national level, which is built on policy research database and topic modeling, is expected to improve the quality of national foresight.

This paper also gives empirical support of agenda coupling between the public and policy elites. Information flows from the media affect policy makers and the public (Katz, 1957), resulting in policy agenda change which arises as a mixture of actor's behaviors, relationships, and politics (Birkland, 2007). While our study doesn't deal with the role of media, we can see the couplings between public discourse and policy agenda through incidences between expert-selected drivers and latent topics (Fig. 5). Public discourse is represented as expert-selected drivers, and policy agenda is detected as latent topics in policy research database. Long-term monitoring of public opinion and policy agenda changes is required to explore agenda-setting mechanisms which are deeply embedded between the public, institutes, and national government.

Acknowledgment

This work was conducted as a part of national foresight process in 2015 by Korea Future Preparatory Committee that receives support from Korea Institute of Science & Technology Evaluation and Planning (KISTEP). W.-S.J. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant no. NRF-2016R1D1A1B03932590). H.K. acknowledges NRF (National Research Foundation of Korea) Grant funded by Korean Government (NRF-2017H1A2A1044205-Global Ph.D. Fellowship Program).

References

Ahn, S.J., 2017. Institutional basis for research boom: from catch-up development to advanced economy. Technol. Forecast. Soc. Chang. 119, 237–245. http://dx.doi.org/10.1016/j.techfore.2016.05.022.

Amanatidou, E., Butter, M., Carabias, V., Könnölä, T., Leis, M., Saritas, O., Schaper-Rinkel, P., van Rij, V., 2012. On concepts and methods in horizon scanning: lessons from initiating policy dialogues on emerging issues. Sci. Public Policy 39, 208–221.

Andersen, A.D., Andersen, P.D., 2014. Innovation system foresight. Technol. Forecast. Soc. Chang. 88, 276–286. http://dx.doi.org/10.1016/j.techfore.2014.06.016.

Andersen, A.D., Andersen, P.D., 2017. Foresighting for inclusive development. Technol. Forecast. Soc. Chang. 119, 227–236. http://dx.doi.org/10.1016/j.techfore.2016.06.007.

Auguie, B., 2017. gridExtra: miscellaneous functions for "Grid" graphics. R package version 2.3.

Bañuls, V.A., Salmeron, J.L., 2011. Scope and design issues in foresight support systems. Int. J. Foresight Innov. Policy 7, 338–351.

Barua, A., Thomas, S.W., Hassan, A.E., 2014. What are developers talking about? An analysis of topics and trends in Stack Overflow. Empir. Softw. Eng. 19, 619–654.

Birkland, T.A., 2007. Agenda Setting in Public Policy. Handbook of Public Policy Analysis: Theory, Politics, and Methods. 125. pp. 63–78.

Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55, 77–84.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Chan, C.M.L., 2013. From open data to open innovation strategies: creating e-services using open government data. In: 2013 46th Hawaii International Conference on System Sciences, pp. 1890–1899. http://dx.doi.org/10.1109/HICSS.2013.236.

Chang, J., 2015. lda: Collapsed Gibbs Sampling Methods for Topic Models. R package version 1.4.2.

Chesbrough, H.W., 2006. Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business Press.

Da Costa, O., Warnke, P., Cagnin, C., Scapolo, F., 2008. The impact of foresight on policy-making: insights from the FORLEARN mutual learning process. Tech. Anal. Strat. Manag. 20, 369–387. http://dx.doi.org/10.1080/09537320802000146.

Daheim, C., Uerz, G., 2006. Corporate foresight in Europe: ready for the next step. In: Second international Seville seminar on future-oriented technology analysis: impact of FTA approaches on policy and decision-making, pp. 1–16.

Dodgson, M., Gann, D., Salter, A., 2006. The role of technology in the shift towards open innovation: the case of Procter & Gamble. R&D Manag. 36, 333–346. http://dx.doi.org/10.1111/j.1467-9310.2006.00429.x.

Dougherty, D., 2017. Taking advantage of emergence for complex innovation eco-systems. J. Open Innov. Technol. Mark. Complex. 3, 14. http://dx.doi.org/10.1186/s40852-017-0067-y.

Durst, C., Durst, M., Kolonko, T., Neef, A., Greif, F., 2015. A holistic approach to strategic foresight: a foresight support system for the German Federal Armed Forces. Technol. Forecast. Soc. Chang. 97, 91–104.

Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. Proc. Natl. Acad. Sci. 101, 5228–5235. http://dx.doi.org/10.1073/pnas.0307752101.
Habegger, B., 2010. Strategic foresight in public policy: reviewing the experiences of the UK, Singapore, and the Netherlands. Futures 42, 49–58.
Havas, A., Schartinger, D., Weber, M., 2010. The impact of foresight on innovation policy-making: recent experiences and future perspectives. Res. Eval. 19, 91–104. http://dx.doi.org/10.3152/095820210X510133.
Hilgers, D., Ihl, C., 2010. Citizensourcing: applying the concept of open innovation to the public sector. Int. J. Public Particip. 4.
Jeon, H., 2013. KoNLP: Korean NLP Package. R package version 0.76.9.
Katz, E., 1957. The two-step flow of communication: an up-to-date report on an hypothesis. Public Opin. Q. 21, 61–78. http://dx.doi.org/10.1086/266687.
Keller, J., Markmann, C., von der Gracht, H.A., 2015. Foresight support systems to facilitate regional innovations: a conceptualization case for a German logistics cluster. Technol. Forecast. Soc. Chang. 97, 15–28.
Kim, J., Han, M., Lee, Y., Park, Y., 2016. Futuristic data-driven scenario building: incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. Expert Syst. Appl. 57, 311–323. http://dx.doi.org/10.1016/j.eswa.2016.03.043.
Koltsova, O., Koltcov, S., 2013. Mapping the public agenda with topic modeling: the case of the Russian livejournal. Policy Internet 5, 207–227.
Lathrop, D., Ruma, L., 2010. Open Government: Collaboration, Transparency, and Participation in Practice. O'Reilly Media, Inc.
Miemis, V., Smart, J., Brigis, A., 2012. Open foresight. J. Futur. Stud. 17, 91–98.
Molitor, G.T., 1977. How to anticipate public-policy changes. SAM Adv. Manag. J. 42, 4–13.
OECD, 2017. Government at a Glance 2017. OECD Publishinghttp://dx.doi.org/10.1787/gov_glance-2017-en.
Palomino, M.A., Bardsley, S., Bown, K., Lurio, J.D., Ellwood, P., Holland-Smith, D., Huggins, B., Vincenti, A., Woodroof, H., Owen, R., 2012. Web-based horizon scanning: concepts and practice. Foresight 14, 355–373.
Pang, A.S.K., 2010. Social scanning: improving futures through web 2.0; or, finally a use for Twitter. Futures 42, 1222–1230 Global Mindset Change.
Prokesch, T., von der Gracht, H.A., Wohlenberg, H., 2015. Integrating prediction market and delphi methodology into a foresight support system – insights from an online game. Technol. Forecast. Soc. Chang. 97, 47–64. http://dx.doi.org/10.1016/j.techfore.2014.02.021.
R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Saritas, O., Smith, J.E., 2011. The big picture - trends, drivers, wild cards, discontinuities and weak signals. Futures 43 (3), 292–312 Special Issue: Future-oriented Technology Analysis.
Schultz, W.L., 2006. The cultural contradictions of managing change: using horizon scanning in an evidence-based policy context. Foresight 8 (4), 3–12.
Sievert, C., Shirley, K.E., 2014. LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63–70.
Slowikowski, K., 2017. ggrepel: Repulsive Text and Label Geoms for 'ggplot2'. R package version 0.7.0.
Thorleuchter, D., den Poel, D.V., 2013. Weak signal identification with semantic web mining. Expert Syst. Appl. 40, 4978–4985. http://dx.doi.org/10.1016/j.eswa.2013.03.002.
Thorleuchter, D., den Poel, D.V., 2015. Idea mining for web-based weak signal detection. Futures 66, 25–34. http://dx.doi.org/10.1016/j.futures.2014.12.007.
van de Vrande, V., de Jong, J.P., Vanhaverbeke, W., de Rochemont, M., 2009. Open innovation in SMEs: trends, motives and management challenges. Technovation 29, 423–437. http://dx.doi.org/10.1016/j.technovation.2008.10.001.
von der Gracht, H.A., Bañuls, V.A., Turoff, M., Skulimowski, A.M., Gordon, T.J., 2015. Foresight support systems: the future role of ICT for foresight. Technol. Forecast. Soc. Chang. 97, 1–6.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Witt, U., 2017. Capitalism as a complex adaptive system and its growth. J. Open Innov. Technol. Mark. Complex. 3, 12. http://dx.doi.org/10.1186/s40852-017-0065-0.
Woo, J., Lee, M.J., Ku, Y., Chen, H., 2015. Modeling the dynamics of medical information through web forums in medical industry. Technol. Forecast. Soc. Chang. 97, 77–90.
Yoon, J., 2012. Detecting weak signals for long-term business opportunities using text mining of web news. Expert Syst. Appl. 39, 12543–12550. http://dx.doi.org/10.1016/j.eswa.2012.04.059.
Yun, J.J., 2015. How do we conquer the growth limits of capitalism? Schumpeterian dynamics of open innovation. J. Open Innov. Technol. Mark. Complex. 1, 17. http://dx.doi.org/10.1186/s40852-015-0019-3.
Yun, J.J., Won, D., Jeong, E., Park, K., Yang, J., Park, J., 2016a. The relationship between technology, business model, and market in autonomous car and intelligent robot industries. Technol. Forecast. Soc. Chang. 103, 142–155.
Yun, J.J., Won, D., Park, K., 2016b. Dynamics from open innovation to evolutionary change. J. Open Innov. Technol. Mark. Complex. 2, 7. http://dx.doi.org/10.1186/s40852-016-0033-0.

Hyunuk Kim is a Ph.D. candidate at the Department of Industrial and Management Engineering, POSTECH, Korea.

Sang-Jin Ahn is a Research Fellow at KISTEP, Korea.

Woo-Sung Jung is an Associate Professor at the Department of Industrial and Management Engineering, POSTECH, Korea.