Relation Between News Topics and Variations in Pharmaceutical Indices During COVID-19 Using a Generalized Dirichlet-Multinomial Regression (g-DMR) Model

Jang Hyun Kim[1, 2], Min Hyung Park[1, *], Yerin Kim[1], Dongyan Nan[2], and Fernando Travieso[3, *]
[1] Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction,
Sungkyunkwan University, Seoul, 03063 Korea
[e-mail: alohakim@skku.edu (J.H.K.); pminhyung12@skku.edu (M.H.P.); yerinee95@skku.edu (Y.K.)]
[2] Department of Interaction Science/Department of Human-Artificial Intelligence Interaction,
Sungkyunkwan University, Seoul, 03063 Korea
[e-mail: ndyzxy0926@skku.edu (D.N.)]
[3] Department of Interaction Science,
Sungkyunkwan University, Seoul, 03063 Korea
[e-mail: nandotravieso@gmail.com (F. T.)]
*Corresponding author: Min Hyung Park and Fernando Travieso

Abstract

Owing to the unprecedented COVID-19 pandemic, the pharmaceutical industry has attracted considerable attention, spurred by the widespread expectation of vaccine development. In this study, we collect relevant topics from news articles related to COVID-19 and explore their links with two South Korean pharmaceutical indices, the Drug and Medicine index of the Korea Composite Stock Price Index (KOSPI) and the Korean Securities Dealers Automated Quotations (KOSDAQ) Pharmaceutical index. We use generalized Dirichlet-multinomial regression (g-DMR) to reveal the dynamic topic distributions over metadata of index values. The results of our analysis, obtained using g-DMR, reveal that a greater focus on specific news topics has a significant relationship with fluctuations in the indices. We also provide practical and theoretical implications based on this analysis.

---

1. Introduction

$\mathbf{C}$ommonly considered the most disruptive global event since WWII, the impact of COVID-19 on the 2020 stock market was beyond the market's expectations. At the beginning of the pandemic, the global market crashed. On March 9, 2020, Wall Street had a so-called "Black Monday," with the Dow Jones Industrial Average experiencing a loss of 2000 points [1]. However, supported by the government's expansionary policy, the market quickly recovered, and some industries even recorded new highs. The normalization of remote work boosted the stock prices of several tech-oriented companies, and the increased awareness regarding health issues raised the stakes for the entire health care industry.

Among the sectors involved in this historic volatility of the market, the pharmaceutical industry was at the core of the virus-induced turmoil. Stimulated by an expectation of vaccine development, stock values in the pharmaceutical sector have exhibited dramatic fluctuations during the pandemic. News on the actual development of vaccines by Pfizer, BioNTech, and Moderna lifted not just the related sector but the entire global market, according to VanEck's report [2].

In the midst of a global vaccine development race, Korean pharmaceutical companies also became involved in the development of COVID-19 vaccines and treatments. The media has covered stories on clinical trials involving plasma therapy, antibody treatments, DNA vaccines, synthetic antigen vaccines, and similar subjects on a daily basis.

Although several scholars have reported possible connections between news media trends and the stock market, the relationship has not been examined thoroughly. Therefore, we attempted to analyze this relationship in the pharmaceutical sector of the Korean financial market. To do this, we scraped news data from the news section of NAVER, the largest portal website in South Korea [3] from January 20 to November 20, 2020. News articles with the keyword "coronavirus" were collected. The text data included relevant nouns from these news articles. In addition, data on the Drug and Medicine index of the Korea Composite Stock Price Index (KOSPI) and data on the Pharmaceutical index of the Korean Securities Dealers Automated Quotations (KOSDAQ) were collected and normalized for use as metadata.

To analyze the relationship between the topic distributions and the metadata, we used a generalized Dirichlet-multinomial regression (g-DMR) model. By optimizing the polynomial equation form of a topic distribution function (TDF), this method captured the variations in the dynamic topic distributions in our two metadata types: (i) months in 2020 and (ii) the KOSPI and KOSDAQ indices.

The remainder of this paper is organized as follows. The literature review section (Section 2) provides an overview of prior research. Section 3 elaborates on our methods, including data collection, preprocessing, g-DMR, and details of model training. Then, Section 4 describes our results, presenting the topic modeling results in detail, with visual plots of the variations in the topic distributions in two metadata. Section 5 concludes the present work, summarizing our key findings and providing suggestions for future research.

2. Literature Review

At present, the use of machine learning for stock market analysis is no longer novel. By enabling computers to learn about and make predictions on a specific topic, machine learning can outperform the human brain in finding a particular solution [4]. Therefore, numerous scholars have applied different models to analyze diverse market movements. In 1990, Kimoto and Asakawa presented a model that was able to find the appropriate timing for a market trade

transaction using modular neural networks [5]. In 2013, Wang et al. proposed a hybrid model that used decision trees and a support vector machine to predict stock futures [6]. Chen et al. [7] expanded the scope of application to the Chinese stock market using a long short-term memory network for stock return prediction. By contrast, the use of text data for market analysis has a relatively short history. Schumaker and Chen [8] made one of the earliest attempts by using breaking financial news articles and quotes related to S&P 500 stocks. Studies that utilized text data to analyze stock market movements exhibited a clear trend of using sentiment analysis. This is because sentiments can have a profound impact on people's behavior and decision-making [9, 10, 11]. Mittal and Goel [12] examined the correlation between public sentiment and market sentiment using Twitter data. Moreover, Shah et al. [13] used sentiment analysis to predict stock prices and achieved an accuracy of 70.59% in predicting the direction of stock price movement. While these two studies both used text data and sentiment analysis for market prediction, the latter extracted sentiments from news data.

In addition to this trend, Nguyen and Shirai [14] incorporated latent Dirichlet allocation (LDA) topic modeling [15] with sentiment analysis. Among the various topic models used in studies, such as crowd detection [16], bug prediction [17], and text classification [18], LDA is one of the most prominent topic modeling techniques in natural language processing. Nguyen and Shirai proposed topic sentiment LDA in particular. They analyzed texts from social media and achieved an accuracy of 56% in stock movement prediction. This accuracy was significant in that the study increased accuracy by 6.07% compared to a previous model that did not use news data. LDA's potential utility in stock market analysis is due to its ability to recognize changing topics. Originally, LDA did not have a specific style for visualizing topic distribution based on time. Nevertheless, Hall et al. [19] conducted a post hoc calculation for topic probabilities based on the year of publication. They used LDA to estimate the probability of a paper covering a certain topic in a given year. Additionally, the DMR topic model [20] extended the application of LDA by reflecting a document's metadata, such as the author and date of publication.

The historic market volatility evoked by the COVID-19 pandemic throughout 2020 and presumably beyond has newly accelerated the trend of machine learning-based market prediction. Owing to the pandemic's massive effect on the market, related data, such as the daily number of COVID-19 cases and social media information on COVID-19, were included as variables in many studies. Baek et al. [21] focused on how volatility increased during the pandemic by using a Markov-switching autoregressive model and concluded that fluctuations were more sensitive to virus-related news than to economic indicators. Biswas et al. [22] used both news and social media data related to COVID-19 and predicted the Indian stock market. Finally, in terms of research that used topic modeling, Agade and Balpande [23] explored the impact of COVID-19 on industries such as business and finance using LDA and nonnegative matrix factorization. The research highlighted the vast impact of COVID-19 and the consequent damage to many industries, as well as to the business and finance sector.

3. Method

3.1. Data collection and preprocessing

In this study, the KOSPI and KOSDAQ indices related to the vaccine and news data on coronavirus were used. We scraped articles that included the keyword "coronavirus" from the NAVER website, which is the largest web portal site in South Korea [3] and has been utilized for predicting stock market in prior studies [24, 25]. The number of articles crawled each day

varied from 500 to 1,000. We used a total of 223,467 news articles. In terms of the indices, our study employed two market indices that have distinct traits in the Korea Exchange: the KOSPI and the KOSDAQ. KOSPI is a representative list of the companies publicly traded in Korea, and KOSDAQ includes approximately 1,000 small and medium-sized businesses and start-ups [26]. Koo et al. [27] reported that investing in KOSDAQ is considered riskier than in KOSPI. Among the many industry indices for each index, the KOSPI Drug and Medicine index and the KOSDAQ Pharmaceuticals index were collected from the Korean Statistical Information Service. The data collection period was from January 20 to November 20, 2020. The index data were normalized.

We used Khaiii [28], an open-source library for processing Korean language. However, as we extracted words with high occurrences, we found several new compound words related to the COVID-19 that were not properly split by Khaiii tokenizer (e.g., social distancing, QR code, work from home, untact, KF94). To resolve this issue, we repeated the extraction of the frequent words and manually added the words to the dictionary. With the customized dictionary, we tokenized the data and conducted part-of-speech-tagging and lemmatization.

3.2. Generalized Dirichlet-multinomial regression (g-DMR)

As might be expected, the g-DMR model is a generalized form of the DMR model [29]. In DMR, ψ_d, i.e., the metadata of a document, becomes one-hot encoded. A feature vector (λ_t), which varies with topic t and with the metadata of the document, is set with values from $N(0, \sigma^2)$. ψ_d and λ_t have linear combinations, as expressed in equation (1). As described in equation (2), topic distribution ($\alpha_{d,t}$) is derived from the output of the linear combination between ψ_d and λ_t. Consequently, α_d, i.e., the stacked vector of $\alpha_{d,t}$, determines the document-topic distribution vector for document d (θ_d), as shown in equation (3). This method is considered to be useful in that it supports implementation of topic modeling with the given metadata of documents. However, using only the linear combination is a limitation of the model as this approach cannot explain the dynamic changes in topic distributions.

$$\Phi_t(\psi_d) = \lambda_t \psi_d \tag{1}$$

$$\alpha_{d,t} = \exp\left(\Phi_t(\psi_d)\right) \tag{2}$$

$$\theta_d = Dir(\alpha_d) \tag{3}$$

However, in g-DMR, the form of $\Phi_t(\psi_d)$ is generalized so that the model having adaptable $\alpha_{d,t}$ can capture more complex variations in topic distributions in the metadata. Owing to this change, the model can accept the metadata of continuous variables as well. The generalized form of function $\Phi_t(\psi_d)$ is noted as a TDF in g-DMR, and it can be adapted to approximate the dynamic topic distributions for the metadata of the document.

Among many approximation methods (e.g., simple polynomial approximation, Fourier approximation, and Hermite approximation), g-DMR utilizes a shifted Legendre polynomial (SLP) based on stability in convergence and efficiency in computation. An SLP with order i is defined as follows:

$$L_i(\psi) = \sum_{k=0}^{i}(-1)^{i+k}\binom{i}{k}\binom{i+k}{k}\psi^k \tag{4}$$

Furthermore, we apply this method here by using more than one metadata with a mutating form of TDF. Using two metadata types (months in 2020, and either the KOSPI or the KOSDAQ index) with topic modeling, we employed the form of two-dimensional Legendre TDF, which can be expressed as

$$\Phi_t(\psi) = \sum_{i=0}^{I}\sum_{j=0}^{J}\lambda_{t,ij}L_i(\psi_1)L_j(\psi_2) \tag{5}$$

where terms I and J are the order of the SLP to be used for our two metadata. We set $I = 4$ and $J = 3$. Our TDF is expressed as follows:

$$\Phi_t(\psi) = \sum_{i=0}^{4}\sum_{j=0}^{3}\lambda_{t,ij}L_i(\psi_1)L_j(\psi_2) \tag{6}$$

3.3 Training information

We utilized the (i) months and (ii) KOSPI or KOSDAQ index values as metadata in g-DMR training. Therefore, the metadata range was $months \in [1, 11]$ and $index \in [0, 1]$. Several hyperparameters were found empirically during the training. We adjusted the values until each topic was composed of consistent words. The topic number, k, was set as 9 (we tested $k \in [5, 13]$). The degrees for the two metadata types (months in 2020 and KOSPI or KOSDAQ Index) were set as 4 and 3, respectively, which implies the order of the equation for approximation of the two. We used 0.01 for the exponential value of the mean in the normal lambda distribution. The standard deviation value of the normal distribution was set to 1.0 for non-constant lambda terms and to 3.0 for constant lambda terms. The model was trained with 500 iterations, at which point the log-likelihood stopped decreasing.

4. Results

4.1 Topics and words

As a result of the topic modeling using g-DMR with $k = 9$, eight common topics were shared by both the KOSPI and KOSDAQ (**Table 1**) and one topic was not (**Table 2**).

Table 1. Eight topics and words shared by the KOSPI and KOSDAQ

Topic name	Words
Politics	President, government, committee, political response, prime minister, minister, North-Korea, member of national assembly, nation, chairman, etc.
Vaccine R&D	Vaccine, development, virus, research, clinical research, medicine, trial, United States, antibody, effectiveness, approval, vaccination, diagnosis, cure, etc.
Quarantine System	Prevention, mask, facility, measure, wearing (mask), social distancing, sterilize, prohibition, etc.
Domestic Spread of Virus	Confirmation, coronavirus, corona-test, contact, Seoul, the confirmed, positive, infection, patient
World Spread of Virus	United States, China, Japan, fatality, government, world, infection, cumulative confirmed person, etc.
Emergency Aid	Finance, economy, company, hire, government, loan, small business owner, provision, fund, project, crisis, bank, application, emergency, work, income, etc.
Social Change	Student, school, online, overcome, education, class, cancellation, delay, campus, culture, performance, participation, donation, etc.
Economy	Decrease, quarter, increase, market, sales, economy, dollars, prospect, export, growth, selling, business, fall, etc.

Table 2. Unique topic for the KOSPI and KOSDAQ

Topic name	Words
Change in Work (KOSPI)	Company, responsiveness, industry, business, factory, service, online, production, digital, untact, etc.
Pandemic in Asia (KOSDAQ)	Japan, China, spread, infection, Tokyo, Korea, confirmation, occurrence, the authorities, etc.

4.2 Topic Trends with metadata

In this section, the plots of our results are presented by topics. The x-axis of the plot refers to the months in 2020, and the y-axis represents the normalized KOSPI or KOSDAQ index values. The plot presents the trends of importance of a specified topic by metadata. The color indicates the significance of the topic.

4.2.1 Topic trends with the KOSPI Drug and Medicine index

In **Fig. 1**, the topic of "Politics" becomes relatively significant in October and November, but the value of the importance itself is approximately 0.25. The results show that the index value was also significantly high in this section. The topic importance of "Vaccine R&D" shows a similar tendency to that of "Politics" with regard to our two metadata (**Fig. 2**). However, the maximum value of the importance of "Vaccine R&D" is higher than that of the former topic.
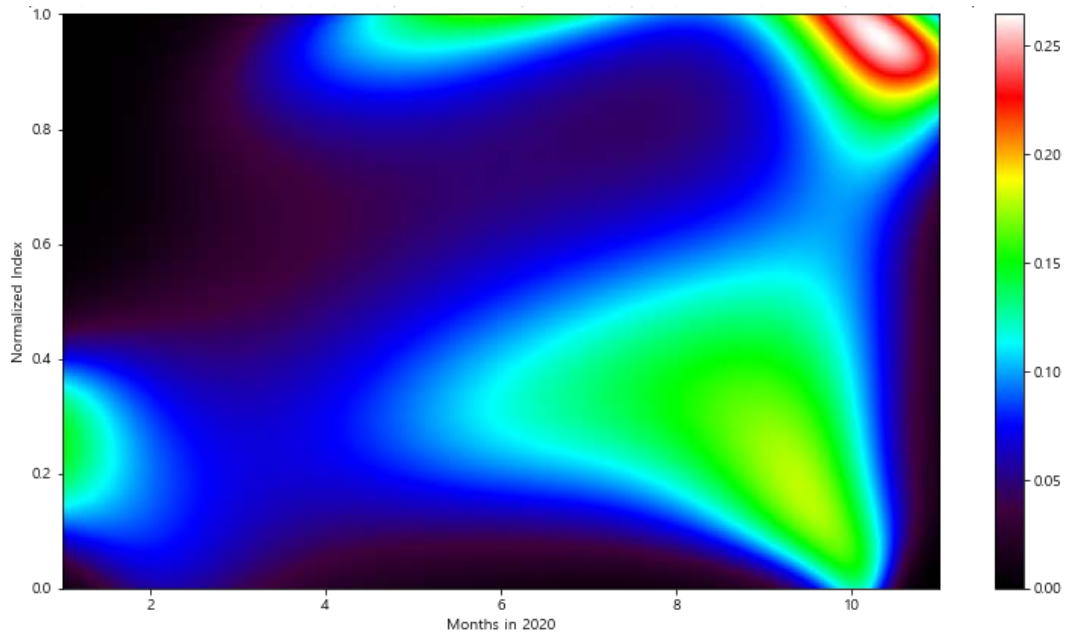
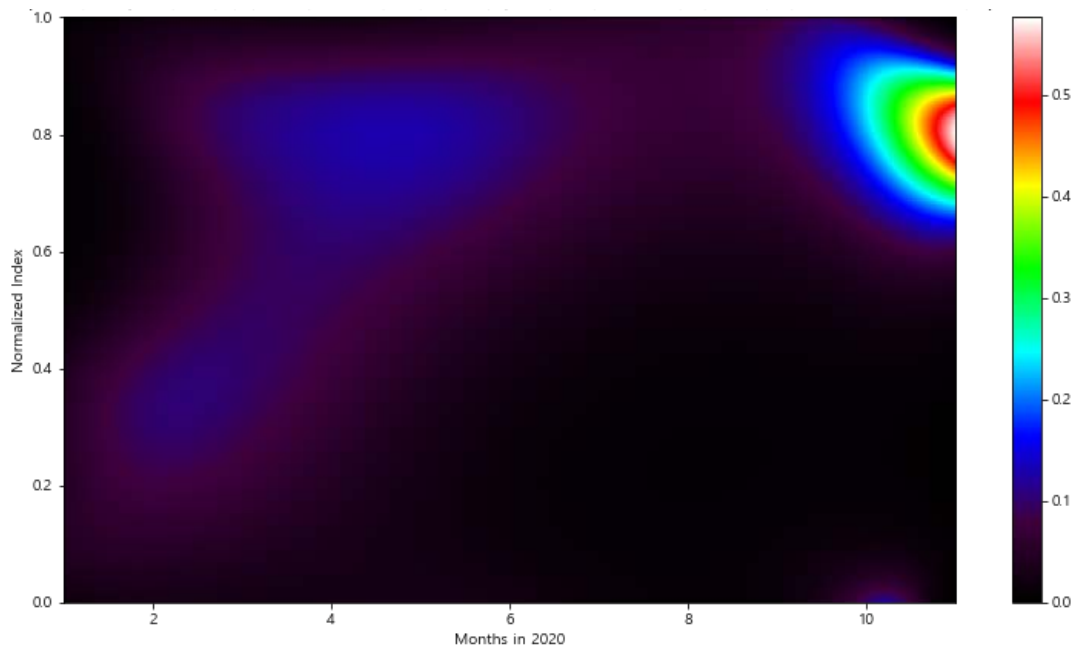**Fig. 1.** Trend of the topic of "Politics" with KOSPI

Fig. 2. Trend of the topic of "Vaccine R&D" with KOSPI

The topic of "Quarantine System" is found to have similar significance over the months but with a higher index value for the higher significance of the topic (**Fig. 3**). When the topic of "Domestic Spread of Virus" is highly concentrated in the period near November, the KOSPI index is low (**Fig. 4**).
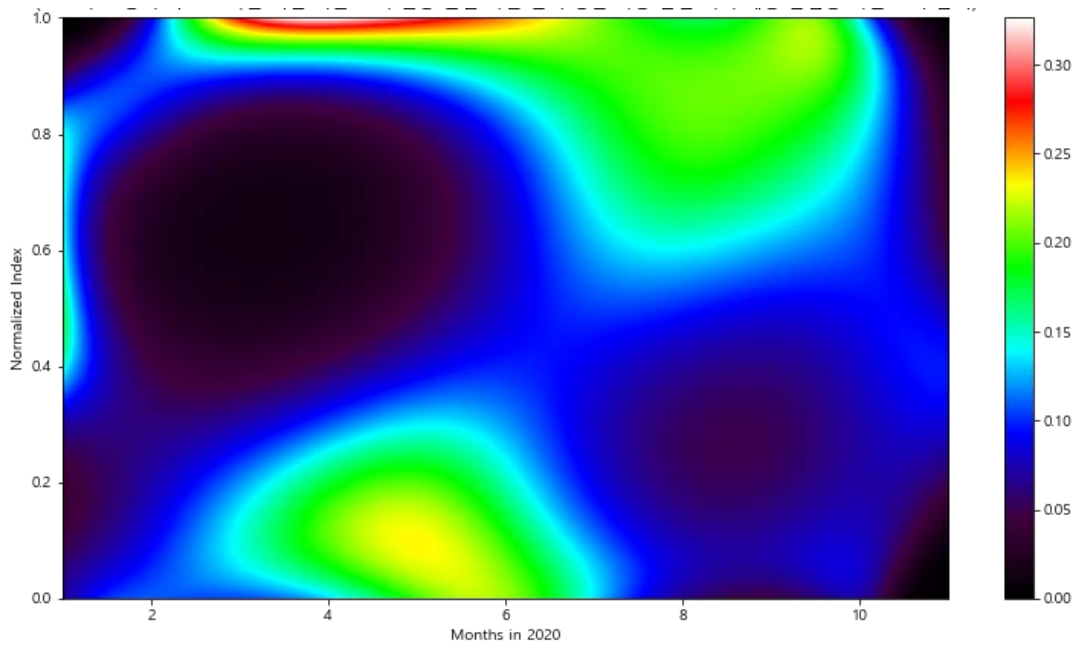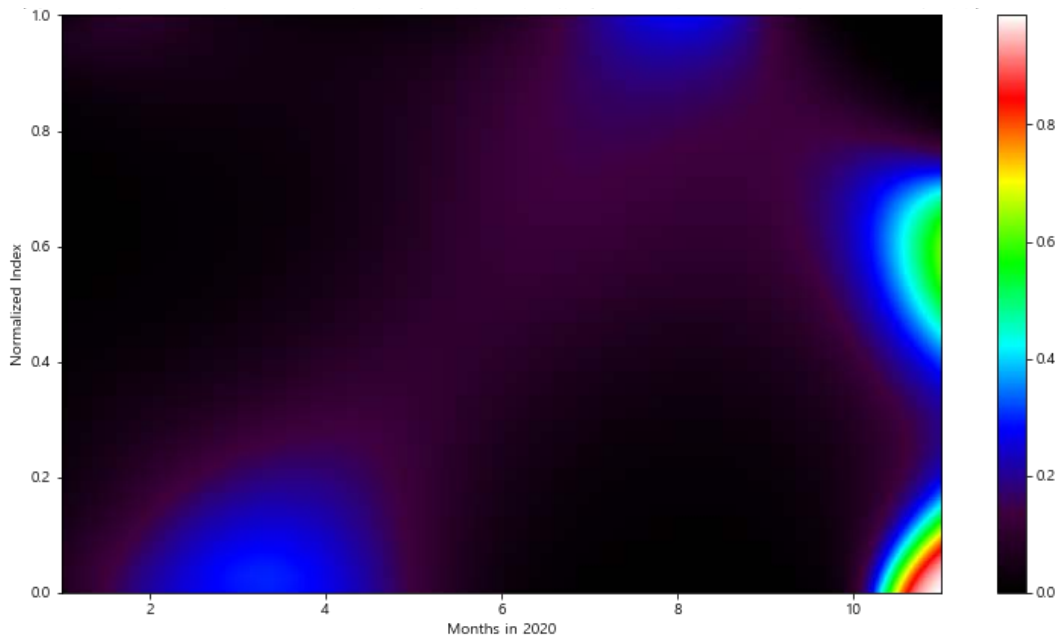
Fig. 3. Trend of the topic of "Quarantine System" with KOSPI

Fig. 4. Trend of the topic of "Domestic Spread of Virus" with KOSPI

As shown in **Fig. 5**, the topic of "World Spread of Virus" occupied an exceedingly higher portion in January than in other months. The intense concentration of the topic is also accompanied by a high index. The appearance of the topic of "Emergency Aid" is strong from July to September but a low index value follows the topic (**Fig. 6**).
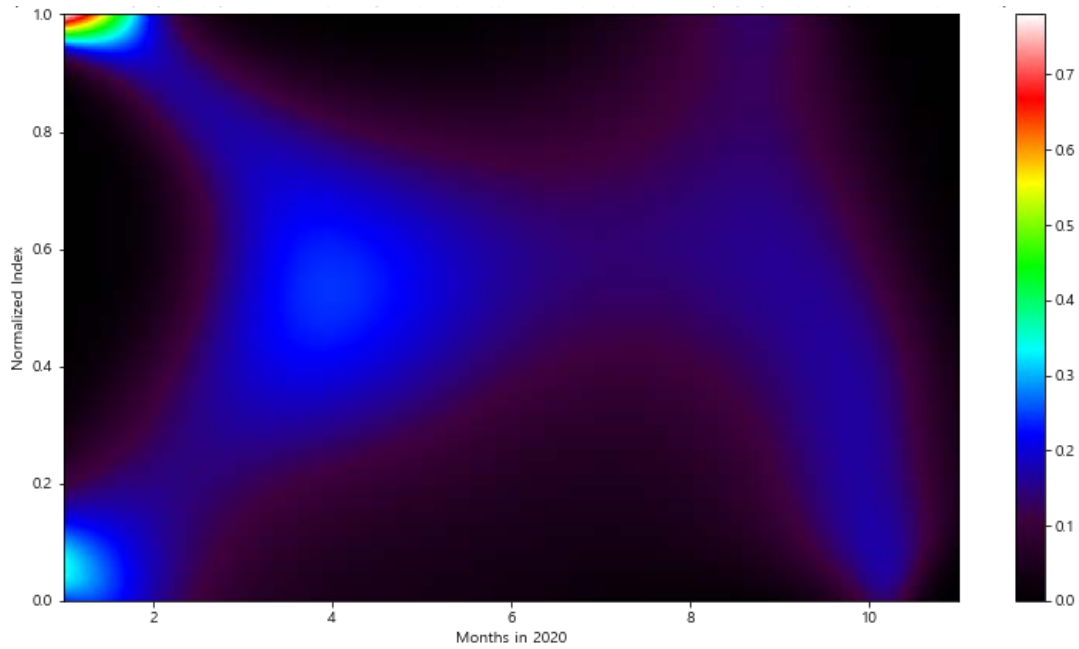
Fig. 5. Trend of the topic of "World Spread of Virus" with KOSPI
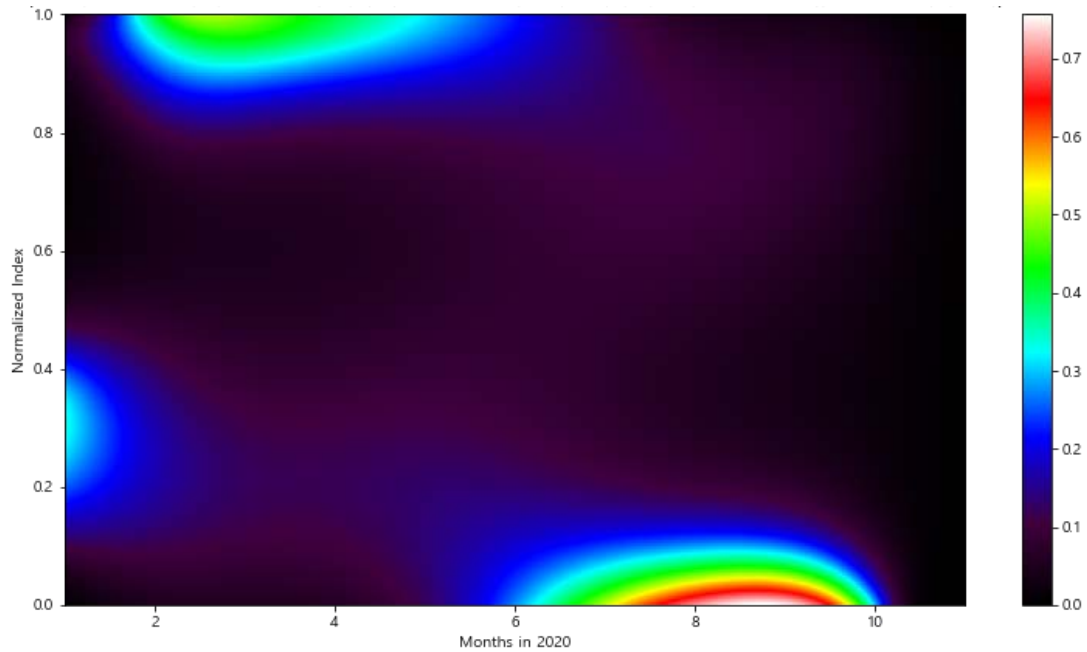
Fig. 6. Trend of the topic of "Emergency Aid" with KOSPI

The topic of "Social Change" emerges more from July to September in **Fig. 7**. The KOSPI index exhibits lower values with a higher intensity of the topic. The importance of the topic of "Economy" is found to be high in the early stages of the pandemic, with an index value of approximately 0.6 (**Fig. 8**).
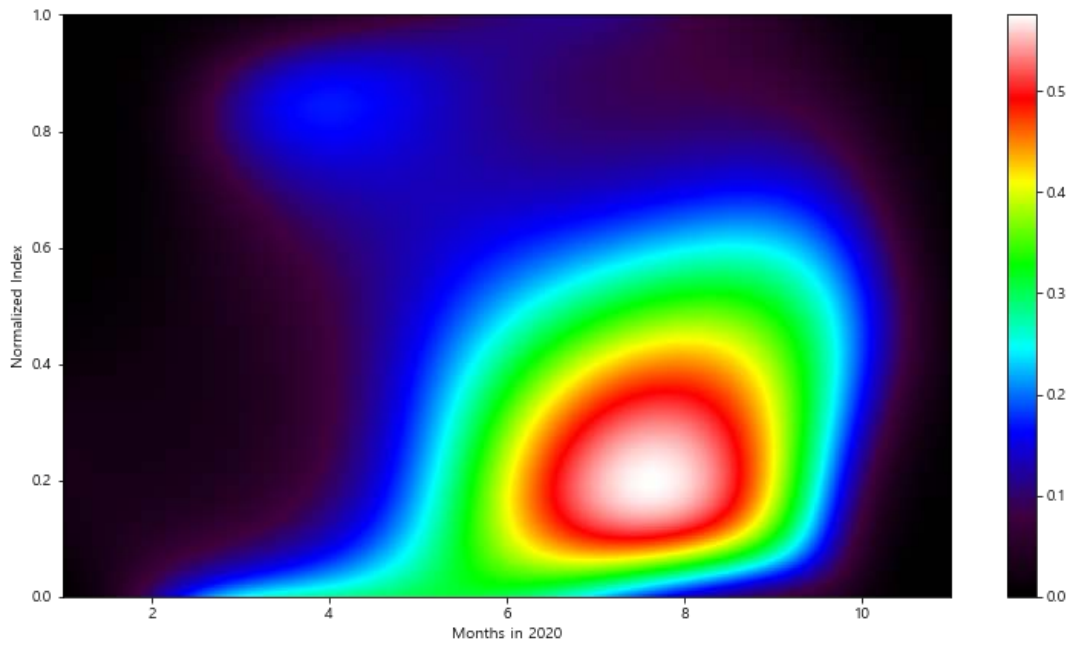
Fig. 7. Trend of the topic of "Social Change" with KOSPI
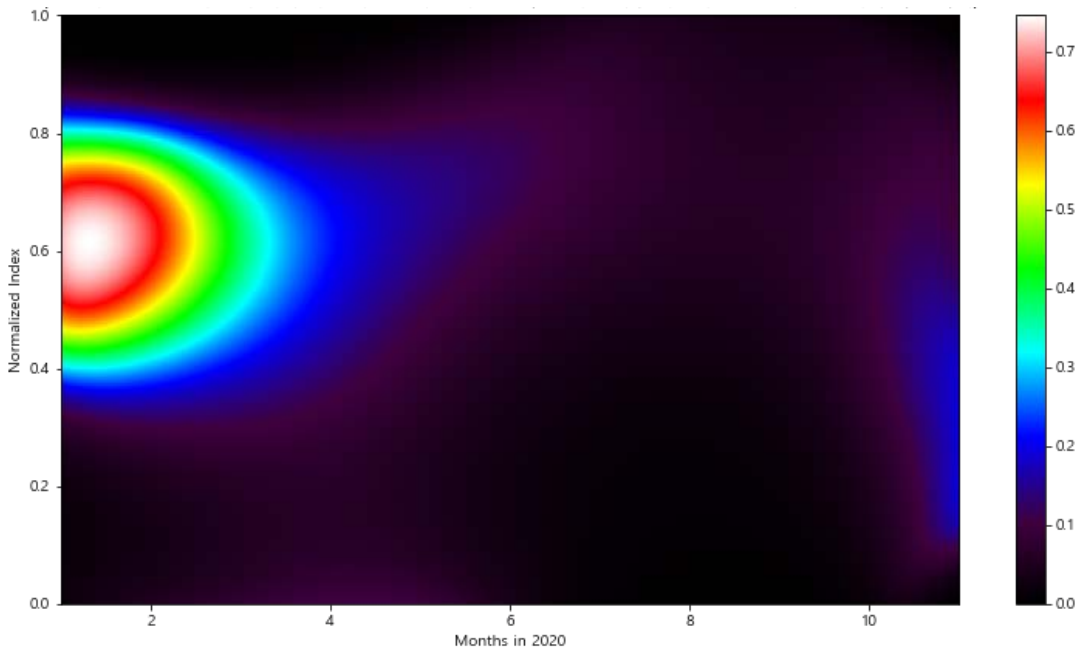
Fig. 8. Trend of the topic of "Economy" with KOSPI

The topic of "Change in Work" remarkably emerges more in November (**Fig. 9**). Higher index value is accompanied by higher emergence of the topic.
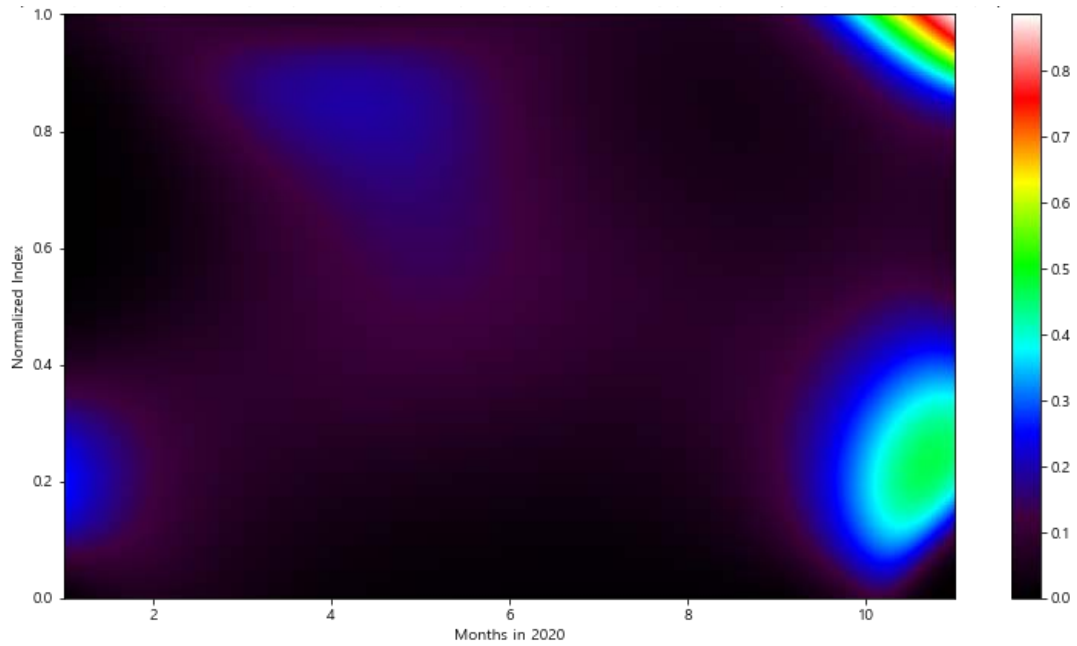
Fig. 9. Trend of the topic of "Change in Work" with KOSPI

4.2.2 Topic trends with the KOSDAQ Pharmaceuticals index

As in **Fig. 10**, the topic of "Politics" is slightly more concentrated between March and May, but the degree of importance is relatively low. The more the topic is shown, the higher the KOSDAQ index value. "Vaccine R&D" exhibits a similar trend to that of "Politics" (**Fig. 11**).
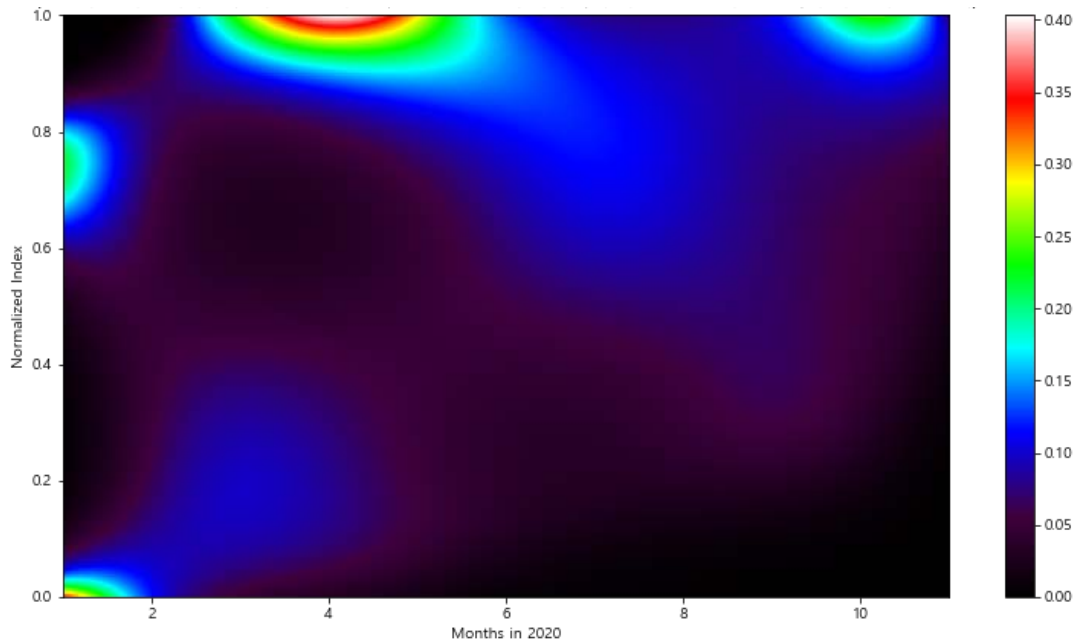

Fig. 10. Trend of the topic of "Politics" with KOSDAQ
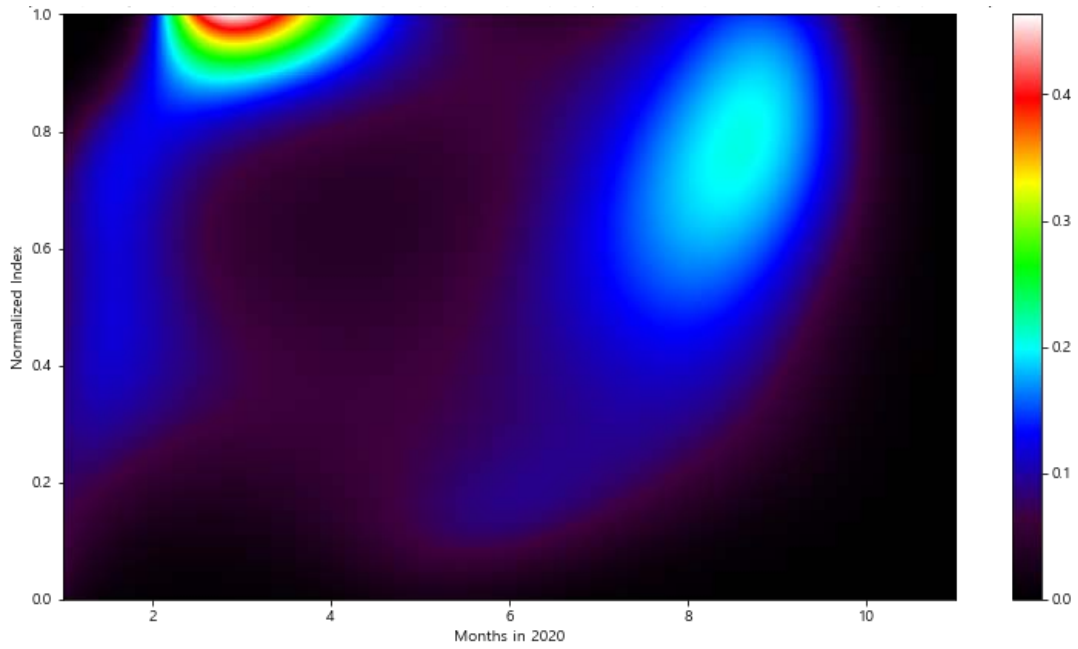
Fig. 11. Trend of the topic of "Vaccine R&D" with KOSDAQ

The topic of "Quarantine System" is more focused from September to November, as shown in **Fig. 12**. The topic importance is high, regardless of the index value. The topic of "Domestic Spread of Virus" has higher importance in November and a high index value (**Fig. 13**).
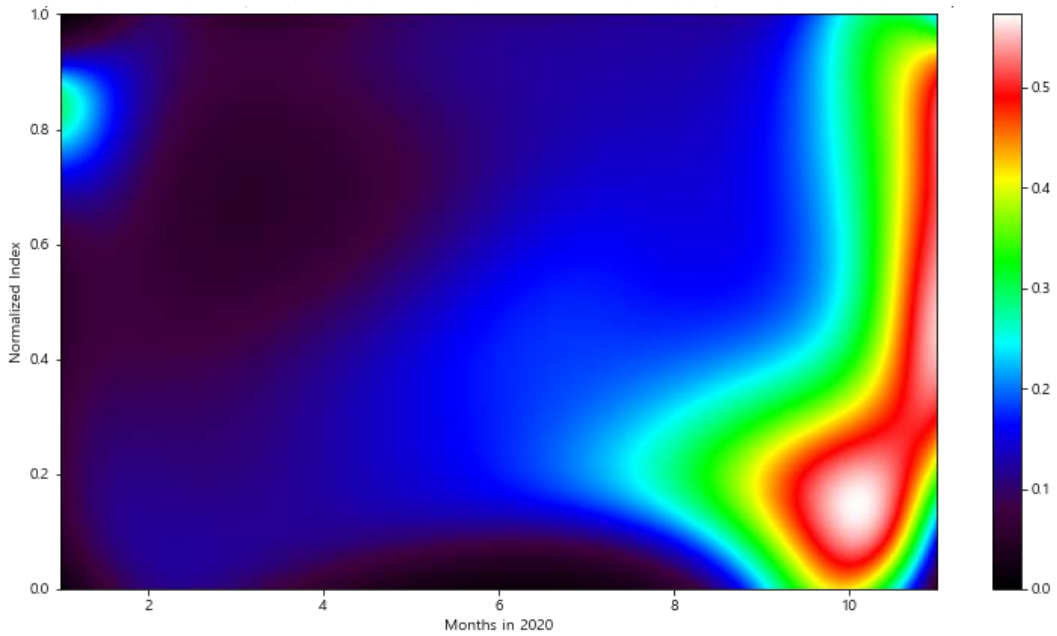

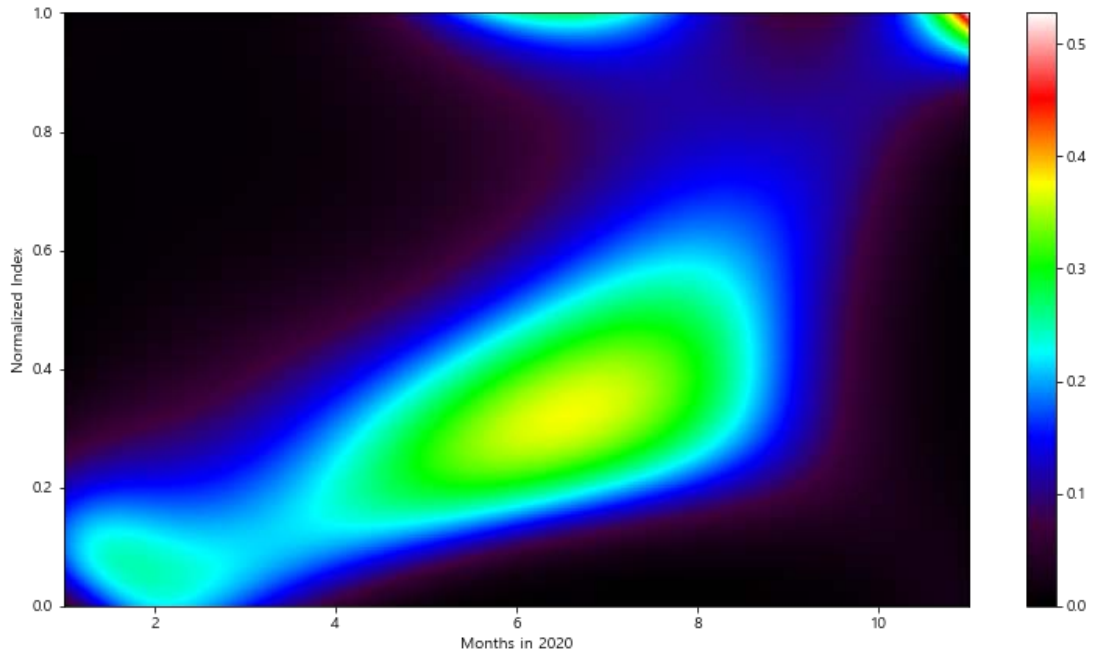**Fig. 12.** Trend of the topic of "Quarantine System" with KOSDAQ

Fig. 13. Trend of the topic of "Domestic Spread of Virus" with KOSDAQ

As revealed in **Fig. 14**, the topic of "World Spread of Virus" emerges more in both ends of the period. A KOSDAQ index value of approximately 0.6 is linked with the intensity of the topic. As shown in **Fig. 15**, "Emergency Aid" is rather more important in January than in other months. The intense topic appearance is related to a higher value of the KOSDAQ index value.
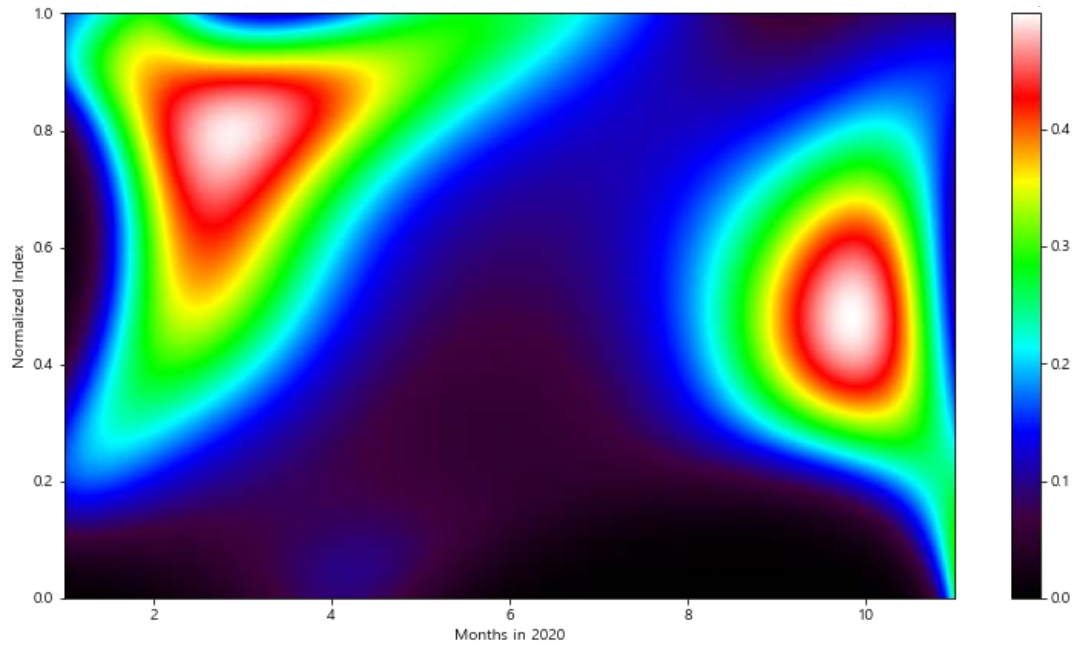

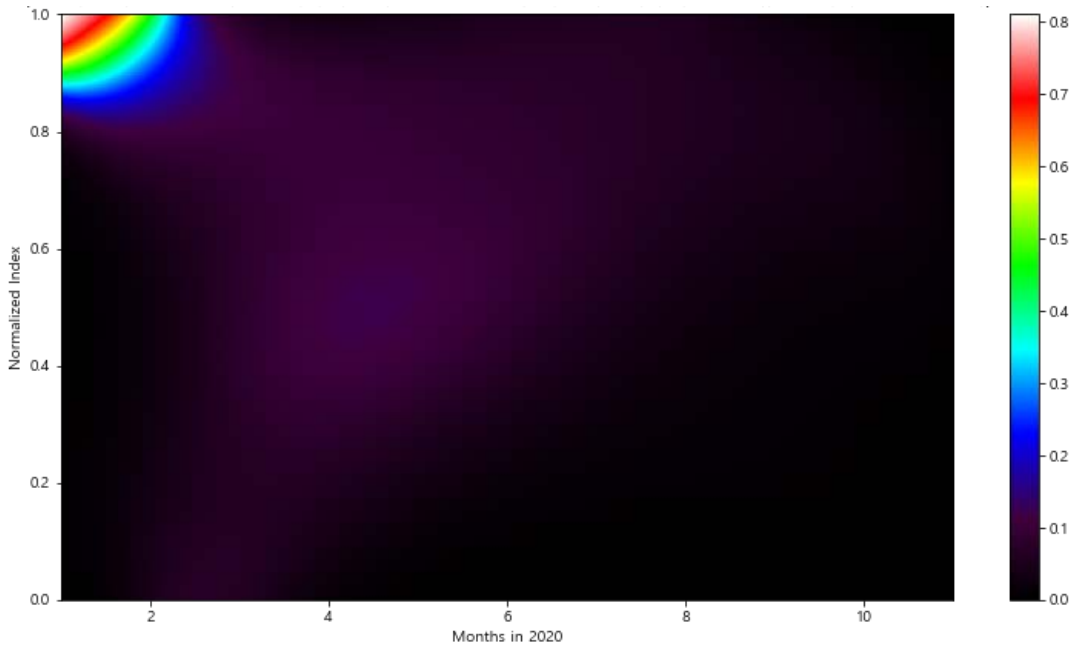**Fig. 14.** Trend of the topic of "World Spread of Virus" with KOSDAQ

Fig. 15. Trend of the topic of "Emergency Aid" with KOSDAQ

The topic of "Social Change" exhibits higher importance in November. A low KOSDAQ value is linked with the higher importance of the topic (**Fig. 16**). The topic of "Economy" is concentrated between March and June, and the index value is approximately 0.6 for the state, which is indicated in **Fig. 17**.
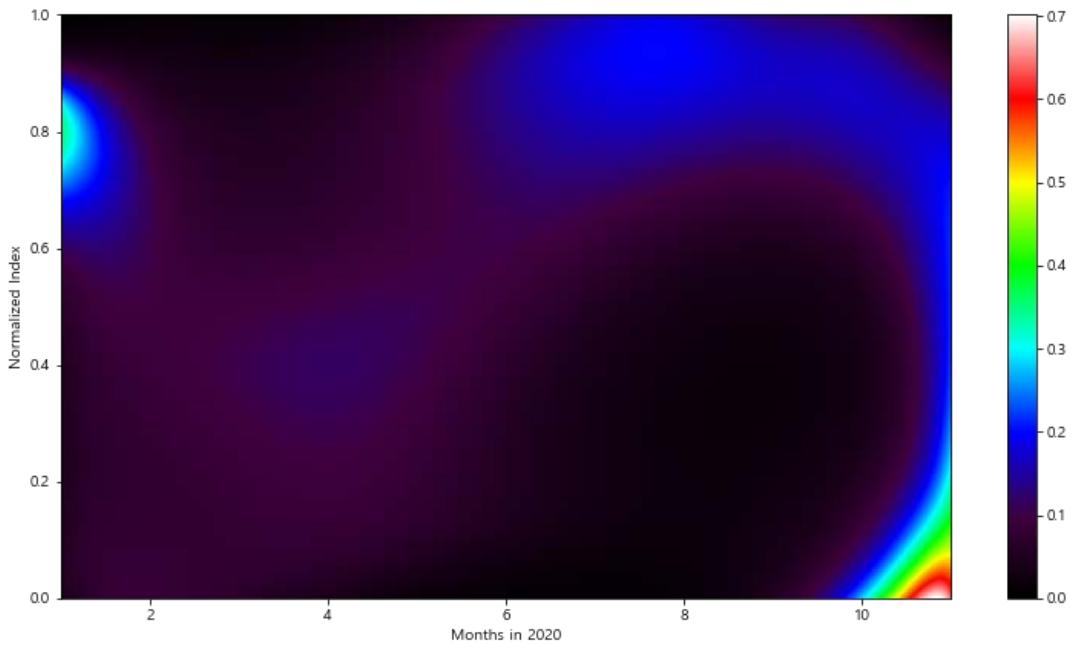

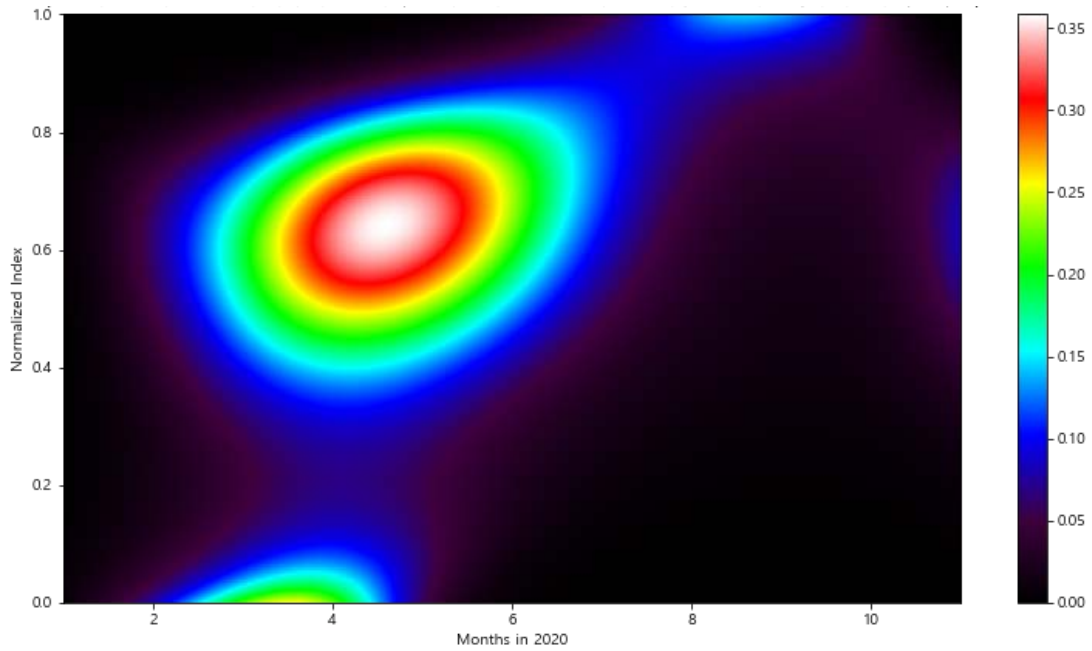**Fig. 16.** Trend of the topic of "Social Change" with KOSDAQ

Fig. 17. Trend of the topic of "Economy" with KOSDAQ

The topic of "Pandemic Spread in Asia" is considered to be more important between May and September. In addition, the low KOSDAQ index is related to the high importance of the topic (**Fig. 18**).
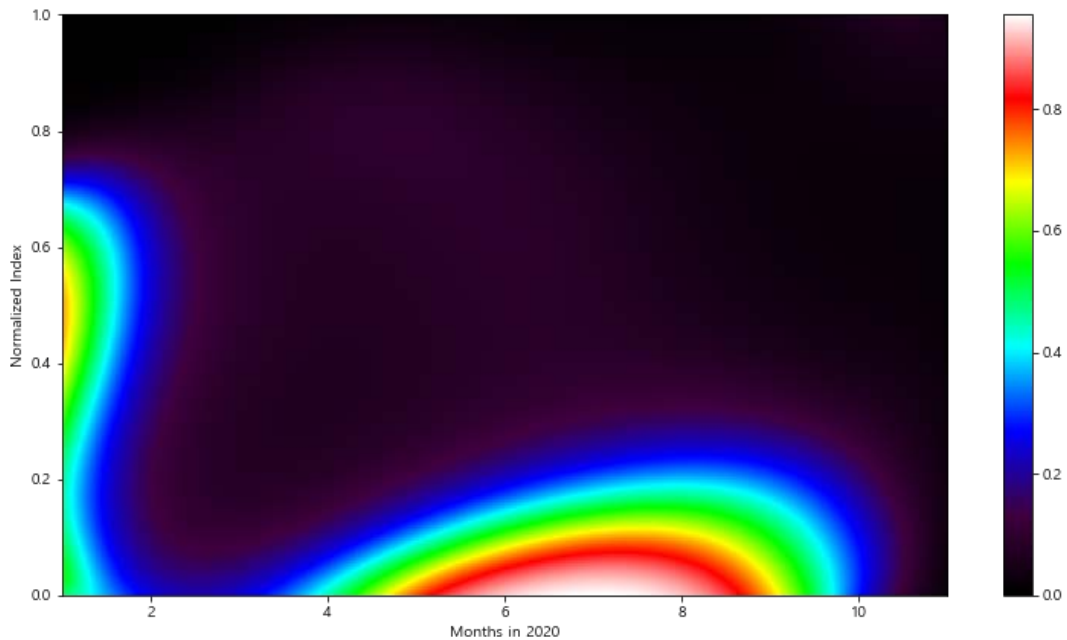

Fig. 18. Trend of the topic of "Pandemic Spread in Asia" with KOSDAQ

4. Conclusion and limitations

Our research is one of the first attempts to explore the relationship between specific news topics and pharmaceutical indices in Korea by employing the g-DMR method, which captures the dynamic topic distributions of multiple continuous metadata. Using more than one metadata, the research provides a more comprehensive understanding of the link between news content and pharmaceutical market in the situation of major health crisis. Moreover, by incorporating a big data based quantitative approach and a qualitative analysis of each topic, the research is able to provide a more evidence-based and concrete illustration. Our main findings, presented below, can serve as a guideline for decision-making by some stakeholders in the pharmaceutical field.

First, the results reveal that when the importance of the topic of "Vaccine R&D" increases, both the KOSPI Drug and Medicine index and the KOSDAQ Pharmaceuticals index tend to be high. Therefore, the impact of news is directly linked to the stock prices of relevant companies. This finding supports the strong effect theory of mass media. The result may be at least partially caused by the agenda-setting power of mass media [30].

Second, the higher importance of "Social Change" was associated with lower KOSPI Drug and Medicine index and KOSDAQ Pharmaceuticals index. It could have been expected that higher importance of "Social Change" will coincide with higher indices as they might share link with the severity of the pandemic. However, the current research presented the opposite result. This implies that people's attention to the "Social Change" was more focused on "social resilience" (e.g., online learning) [31], rather than the severity of the pandemic. Such belief for resilience was reflected in other financial sectors' recovery, according to prior research [32, 33]. Based on the inverse movement between pharmaceutical sector and others under the COVID-19 situation, the belief for overcoming the pandemic led to the lower pharmaceutical indices.

Third, when news topics related to the "Domestic Spread of Virus," "Emergency Aid" are high in importance, the KOSPI Drug and Medicine index and the KOSDAQ Pharmaceuticals index tend to be low and high, respectively. This is an interesting finding. As investors of KOSPI are relatively more risk averse, it may be the case that these investors liquidated their assets when the domestic COVID-19 situation worsened. Conversely, KOSDAQ investors, who have a higher risk tolerance, anticipated that the market would rebound; hence, they may have contributed to the upward trend of the index.

Overall, these findings might indicate that news topics related to "Vaccine R&D," "Social Change," "Domestic Spread of Virus," and "Emergency Aid" are significantly related to the pharmaceutical index.

We believe this method will contribute to the examination of the relationship between news topics and industry indices in other sectors of the market as well. Our study provides several meaningful implications, as addressed above. However, there is still room for improvement, which should be explored in future studies. We suggest a quantitative measurement of the trends between metadata and topic importance for future research. Furthermore, as "NAVER" data may have a data selection bias, a verification of our methods using the data from different media platform constitutes a meaningful attempt.

References

[1] L. Bayly, "Dow closes with decline of 2,000 points, almost ending 11-year bull market," 2020. [Online]. Available: https://www.nbcnews.com/business/markets/dow-set-open-decline-1-300-points-oil-war-adds-n1152941

[2] VanEck, "Pharma Sector Steps up as a Portfolio Prescription," 2020. [Online]. Available: https://www.vaneck.com/blogs/thematic-investing/pharma-sector-steps-up-as-a-portfolio-prescription/?country=us

[3] J. Kim, and N. Han, "The development of Korean online dictionaries: a case study of Naver dictionary services," *Lexicography*, vol. 3, no. 1, pp. 19-37, 2016. Article (CrossRef Link)

[4] J. Song, K. T. Kim, B. J. Lee, S. Y. Kim, and H. Y. Youn, "A novel classification approach based on Naïve Bayes for Twitter sentiment analysis," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 6, pp. 2996-3011, 2017. Article (CrossRef Link)

[5] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in *Proc. of IEEE IJCNN*, San Diego, CA, USA, pp. 1-6, 1990.
Article (CrossRef Link)

[6] D. Wang, X. Liu, and M. Wang, "A DT-SVM strategy for stock futures prediction with big data," in *Proc. of IEEE CSE*, Sydney, NSW, Australia, pp. 1005-1012, 2013. Article (CrossRef Link)

[7] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. of IEEE BigData*, Santa Clara, CA, USA, pp. 2823-2824, 2015.
Article (CrossRef Link)

[8] R. P. Schumaker, and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Trans. Inf. Syst.,* vol. 27, no. 2, pp. 1-19, 2009.
Article (CrossRef Link)

[9] Y. Chen, Y. Lin, and W. Zuo, "Phrase-based topic and sentiment detection and tracking model using incremental HDP," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 12, pp. 5905-5926, 2017.
Article (CrossRef Link)

[10] D. Nan, Y. Kim, M. H. Park, and J. H. Kim, "What Motivates Users to Keep Using Social Mobile Payments?," *Sustainability*, vol. 12, no. 17, 2020, Art. no. 6878. Article (CrossRef Link)

[11] J. Jung, P. Petkanic, D. Nan, and J. H. Kim, "When a girl awakened the world: A user and social message analysis of Greta Thunberg," *Sustainability*, vol. 12, no. 7, 2020, Art. no. 2707.
Article (CrossRef Link)

[12] A. Mittal, and A. Goel, "Stock prediction using twitter sentiment analysis," Stanford Univ., Stanford, CA, USA, 2012. Article (CrossRef Link)

[13] D. Shah, H. Isah, and F. Zulkernine, "Predicting the effects of news sentiments on the stock market," in *Proc. of IEEE BigData*, Seattle, WA, USA, pp. 4705-4708, 2018. Article (CrossRef Link)

[14] T. H. Nguyen, and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proc. of ACL-IJCNLP*, Beijing, China, pp. 1354-1364, 2015.
Article (CrossRef Link)

[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993-1022, 2003. Article (CrossRef Link)

[16] X. Huang, W. Wang, G. Shen, X. Feng, and X. Kong, "Crowd activity classification using category constrained correlated topic model," *KSII Trans. Internet Inf. Syst.*, vol. 10, no. 11, pp. 5530-5546, 2016. Article (CrossRef Link)

[17] G. Yang, K. Min, J. W. Lee, and B. Lee, "Applying topic modeling and similarity for predicting bug severity in cross projects," *KSII Trans. Internet Inf. Syst.*, vol. 13, no. 3, pp. 1583-1598, 2019.
Article (CrossRef Link)

[18] J. Ma, Y. Zhang, Z. Wang, and B. Chen, "A new fine-grain SMS corpus and its corresponding classifier using probabilistic topic model," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 2, pp. 604-625, 2018. Article (CrossRef Link)

[19] D. Hall, D. Jurafsky, and C. D. Manning, "Studying the history of ideas using topic models," in *Proc. of EMNLP*, Honolulu, HI, USA, pp. 363-371. Article (CrossRef Link)

[20] D. Mimno, and A. McCallum, "Topic models conditioned on arbitrary features with dirichlet-multinomial regression," *arXiv preprint arXiv:1206.3278*, 2012. Article (CrossRef Link)

[21] S. Baek, S. K. Mohanty, and M. Glambosky, "COVID-19 and stock market volatility: An industry level analysis," *Financ. Res. Lett.*, vol. 37, 2020, Art. no. 101748. Article (CrossRef Link)

[22] S. Biswas, I. Sarkar, P. Das, R. Bose, and S. Roy, "Examining the effects of pandemics on stock market trends through sentiment analysis," *J. Xidian. Univ.*, vol. 14, no. 6, pp. 1163-1176, 2020. Article (CrossRef Link)

[23] A. Agade, and S. Balpande, "Exploring the non-medical impacts of Covid-19 using natural language processing," *Preprints*, 2020. Article (CrossRef Link)

[24] K. Nam, and N. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market," *Decis. Support. Syst.*, vol. 117, pp. 100-112, 2019. Article (CrossRef Link)

[25] Y. Kim, S. R. Jeong, and I. Ghani, "Text opinion mining to analyze news for stock market prediction," *Int. J. Advance. Soft Comput. Appl.*, vol. 6, no. 1, 2014. Article (CrossRef Link)

[26] Haps, "Investing in the Korean market," 2011. [Online]. Available: https://www.hapskorea.com/investing-korean-market/

[27] B. Koo, J. Chae, and H. Kim, "Does Internet Search Volume Predict Market Returns and Investors' Trading Behavior?," *J. Behav. Financ.*, vol. 20, no. 3, pp. 316-338, 2019. Article (CrossRef Link)

[28] https://github.com/kakao/khaiii

[29] M. Lee, and M. Song, "Incorporating citation impact into analysis of research trends," *Scientometrics*, vol. 124, pp. 1191–1224, 2020. Article (CrossRef Link)

[30] M. E. McCombs, and D. L. Shaw, "The agenda-setting function of mass media," *Public Opin. Q.,* vol. 36, no. 2, pp. 176-187, 1972. Article (CrossRef Link)

[31] Global Resilience Institute, "Community resilience amidst COVID-19: Virtual learning platforms," 2020. [Online]. Available: https://globalresilience.northeastern.edu/community-resilience-amidst-covid-19-virtual-learning-platforms/

[32] R. Albuquerque, Y. Koskinen, S. Yang, and C. Zhang, "Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash," *The Review of Corporate Finance Studies*, vol. 9, no. 3, pp. 593-621, 2020. Article (CrossRef Link)

[33] M. Pagano, C. Wagner, and J. Zechner, "Disaster resilience and asset prices," *arXiv preprint arXiv:2005.08929*, 2020. Article (CrossRef Link)

Jang Hyun Kim is an associate professor at Interaction Science, Sungkyunkwan University and his research focuses on social/semantic data analysis, social media, and future media

Min Hyung Park is a Master's student of Human-AI Interaction at Sungkyunkwan University in Seoul. His major is Applied Artificial Intelligence. His research area is development of deep learning models in natural language processing.

Yerin Kim is a Master's student of Applied Artificial Intelligence at Sungkyunkwan University in Seoul. Before, she majored in International Relations. She is interested in multidisciplinary approach to international relations theories and artificial intelligence.

Dongyan Nan is a PhD student at Department of Interaction Science, Sungkyunkwan University. His research focuses on user experience, big data, and technology management.

Fernando Travieso received a doctorate from Department of Interaction Science, Sungkyunkwan University. His research focuses on human-computer interaction.