Qual Quant (2018) 52:653–667
https://doi.org/10.1007/s11135-017-0582-8

Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence

Young-Joo Lee[1] **· Ji-Young Park**[1]

Published online: 12 October 2017

Abstract To foresee the advent of new technologies and their socio-economic impact is a necessity for academia, governments and private enterprises as well. In the future studies, the identification of future signal is one of the renowned techniques for analysis of trends, emerging issue, and gaining future insights. In the Big Data era, recent scholars have proposed using a text mining procedure focusing upon web data such as new social media and academic papers. However, the detection of future signals is still under a developing area of research, and there is much to improve existing methodology as well as developing theoretical foundations. The present study reviews previous literature on identifying emerging issue based on the weak signal detection approach. Then the authors proposed a revised framework that incorporate quantitative and qualitative text mining for assessing the strength of future signals. The authors applied the framework to the case study on the ethical issues of artificial intelligence (hereafter AI). From EBSCO host database, the authors collected text data covering the ethical issues in AI and conducted text mining analysis. Results reveal that emerging ethical issues can be classified as strong signal, weak signal, well-known but not so strong signal, and latent signal. The revised methodology will be able to provide insights for government and business stakeholders by identifying the future signals and their meanings in various fields.

Keywords Future signal · Artificial intelligence · Ethical issue · Text mining · Data-driven foresight

✉ Young-Joo Lee
billieyz@gmail.com

Ji-Young Park
jiyoung.park@nia.or.kr

[1] Future Strategy Center, National Information Society Agency (NIA), 53, Cheomdan-ro, Dong-gu, Daegu 40168, Korea

1 Introduction

To foresee the advent of new technologies and their socio-economic impact is a necessity for academia, governments and private enterprises. Environmental scanning, also known as horizon scanning, is a traditionally developed foresight methodology (Choo and Auster 1993). Environmental scanning can help management to plan the future event by analyzing information from a wide field of the external environment (Choo 1999). The process typically involves identification of events, trends and their relationships by analyzing information from the environments. The identification of weak signals is one of the renowned techniques for analysis of trends, emerging issue, and gaining future insights (Ansoff 1975; Coffman 1997; Kamppinen et al. 2002). Weak signal is defined as 'the future-oriented information behind future trends, changes and emerging phenomena' (Hiltunen 2008). Weak signal can be detected by unusual, interrupting events (Dator 2002) and is related to discontinuous transformations (Hiltunen 2006). Coffman (1997) argue weak signal has seven characteristics: 'an idea or trend that will affect the business', 'the new and surprising the signal', 'sometimes difficult to track down amid other signals', 'a treat or opportunity', 'often scoffed at by people', 'usually has a substantial lag time before it will become mainstream', and 'opportunity to learn, grow and evolve'. As Future changes can be seen through discontinuity or disconnections of new behavioral patterns that are entirely different from those that existed in the past (Ansoff 1975), detection of the signs or indication of the discontinuity is very important to predict the future. The analysis of weak signal is widely used in the research field including technology management and R&D planning (Hong et al. 2009; Thorleuchter and Van den Poel 2013c; Julien et al. 2004). The procedure for foresight based on weak signal detection consists of four steps: (1) scanning weak signals, (2) assessing weak signals, (3) transforming the signals into issues, and (4) interpreting the issues for new futures (Maurits Butter et al. 2010).

In traditional future studies including weak signal detection method, the scanning of future signal has been relied heavily upon the intuitive insight of experienced-experts; while costly and time-consuming, holds the possibility to provide different results based on the analyzer's perspectives. In the Big Data era, as the number and size of information for scanning increases exponentially, it is almost impossible for a group of experts to effectively scan topics from various information sources. Therefore recent scholars have proposed using a text mining of information sources from the internet, focusing upon web data such as new social media and academic papers (e.g. Yoo 2009; Yoon 2012; Thorleuchter and Van den Poel 2013a, b, c; Smith 2015; Hong et al. 2016). However, the detection of future signal is still under a developing area of research, and there is much to improve existing methodology as well as developing theoretical foundations.

Accordingly, the present study tries to extend and expand the seminal work of Hiltunen (2008), Yoon (2012), and other related literature to develop a revised method for detecting future signal based on the text mining techniques. In Sect. 2, the authors discuss related theories, methodologies, and limitations. Then we continue to establish a revised analytical framework that combines quantitative and qualitative analysis. In Sect. 3, we apply our new framework in the case study of issue spreading of ethical problems of AI and robotics. Section 4 follows by discussing methodological and managerial implications.

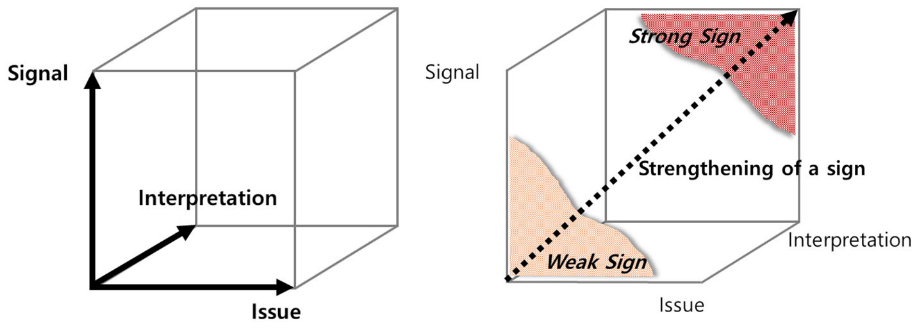

Fig. 1 Thre dimensional space of future sign (Hiltunen 2008)

2 Analysis framework for detecting future signals with text mining

2.1 Future sign and its three dimensions

A futurist Hiltunen (2008) proposed three-dimensional space to understand the signal amplification process. In this three-dimensional space, each axis called the signal, the issue, and the interpretation, respectively. The signal is defined as 'the number and/or visibility of signals', the issue is defined as 'number or variety of events', and the interpretation is defined as 'the receiver's understanding of the future sign's meaning' (Fig. 1).

The magnitude of the future sign can be measured by a combination of these three dimensions (Kuusi and Hiltunen 2011). A weak sign is a future sign which has low level of signal, issue, and interpretation, meaning that a presentation of the future issue is about to diffuse to some extent but yet to be much interoperated by receivers. A weak sign becomes a strong sign when interpretation of the meaning of the sign is becoming clear.

Strong sign has a potential to become a trend or mega trend. In this sense a weak sign is not regarded as a part of a future trend, but a clue for detecting how trend will be shaped in the future. Although Hiltunen (2008) developed the theoretical basis for detecting the weak signal,[1] she has left us room for further research on how to quantitatively and qualitatively evaluate the strength of each dimension.

2.2 Detecting future signal using text mining techniques

Yoon (2012) exploited term occurrence frequency in documents as a means to detect this patterns and measure the changes and strength of the future signals. To analyze the pattern, Yoon (2012) suggested text mining method from web news to measure signal and issue in the approach of Hiltunen (2008). After extracting keywords frequency in the corpus, he classified keywords with low frequency but high increase rate as weak signals, and keywords with high frequency and high increase rate as strong signals. More specifically, he proposed degree of visibility (DoV) as an indicator for signal and degree of diffusion (DoD) as an indicator for issue. The formula to calculate each indicator is as follows:

[1] In the literature the term 'signal' and 'sign' were used interchangeably, we consider Hiltunen (2008)'s weak sign to be a synonym with weak signal for preventing confusion.

$$DoV_{ij} = \left(\frac{TF_{ij}}{NN_j}\right) \times \{1 - tw \times (n - j)\} \quad DoD_{ij} = \left(\frac{DF_{ij}}{NN_j}\right) \times \{1 - tw \times (n - j)\}.$$

Degree of visibility (DoV) is the statistical representation of the 'signal' level of a future sign that measures the degree of the frequency of a defined keyword in a set of documents. In the formula, *TFij* is the total occurrence frequency of a keyword i in the period j, *NNj* is the total number of news articles in the period j, *n* is the number of periods (Yoon 2012, p. 12547). Likewise, in terms of degree of diffusion (DoD), the document frequency of each keyword is directly related to 'issue' level of a future sign. In the formula, *DFij* is the number of the document frequency of keyword *i* in the period *j*, *NNj* is the total number of news articles in the period *j*, *n* is the number of periods, and *tw* is a time-weight (Yoon 2012, p. 12548).

Next, he schemed out a keyword emergence map (KEM) using TF and DoV, and keyword issue map (KIM) using DF and DoD. These two map is a baseline portfolio to detect weak signal and strong signal. In the case of KEM, the TF is mapped onto the X axis and the geometric average of the DoV for each keyword is mapped onto the Y axis. This is mirrored within KIM, DF and DoD. In each map, the median value in axis X and Y divides the four quadrants and classifies the keywords by the location in each quadrant.

According to Hiltunen (2008) future signs that have a possibility of being weak signals have closely related to topics that have an abnormal pattern but are rarely exposed. Applying this concept to Yoon's (2012) keyword portfolio maps, the 'abnormal pattern' can be detected if related keywords have an unusual fluctuation in the increase of occurrence frequency, and the 'rarely exposed pattern' can be detected if absolute occurrence frequency is lower than other keywords. Conversely, related keywords which exhibit a high level of increase rate and absolute frequency of occurrence would be a strong signal.

Therefore, in the Fig. 2, it might be deducted that keywords in the quadrant A are *weak signals*, which are below average in share but growth rate is above average. These signals are not yet strong, but are expected to rise in the future. The keywords in the quadrant B are *strong signals*, which show more than average in share and increase rate and is gaining much awareness by majority. Finally, if you select keywords that locate in the same quadrant at both KEM and KIM, those are clearer weak or strong signals.

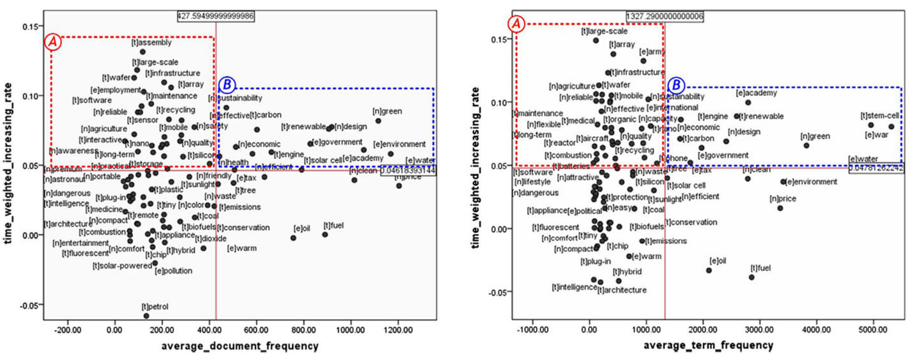

Fig. 2 Examples of keyword portfolio maps (KIM, KEM) by Yoon (2012)

2.3 Revised framework for detecting future sign

Although Yoon (2012) have improved a traditional methodology by linking detecting future signals with quantitative analysis method, there still remain limitations in terms of theory and methodology.

One limitation of the method proposed by Yoon (2012) is that there is no way to interpret and distinguish the signals when keywords classifications are mixed at each map. For example, when a keyword is located in one quadrant in KEM but different quadrant in KIM, analyzers may have to decide to discard the keyword because it shows confusing signal or just choose one quadrant by their intuition. Then we may lose objectivity of the analysis which were significant merit in the Yoon (2012)'s study.

Another methodological problem is that we cannot identify the inherent meaning of future signal with just a single keyword. It is related to the level of abstractedness of each keyword. If individual keywords contain multiple numbers of meaning, it may not be able to distinguish issues at appropriate level for the purpose of research because it covers various sub-issues. On the other hand, very specific keywords often refer to individual events, which may not be able to capture the high-level issues.

Finally, as Hiltunen (2008) also argued, the issue and signal dimension are measurable which are objective matter of the future sign, the interpretation dimension is a subjective matter, which is contextually dependent. It is still unclear on how to measure or assess the degree of interpretation to determine whether a signal would become *weaker* or *stronger*.

Synthesizing abovementioned weakness, this study added a procedure to extract issue topics based on the semantic analysis of keywords, which help us find meaning from a group of words that share context rather than individual word (Thorleuchter and Van den Poel 2013b; Murtagh et al. 2016). This additional procedure includes analysis of the co-occurrence relation of keywords appearing in each quadrant in KIM and KEM to reveal central, salient meaning (Doerfel 1998). The frequency and co-occurrence relationship of the words can show the symbolism that is emphasized in the whole buzz and the tendency to create a specific meaning through the combination of words (Hellsten et al. 2010; Hsu et al. 2013; Wang and Rada 1998). It is a sense-making process that complement quantitative analysis with qualitative or deductive analysis, which in turn help us to identify mutually related events (Kamppinen et al. 2002) and the potential drivers of the change (Kuosa 2010). Suggested analysis framework is shown in Fig. 3.

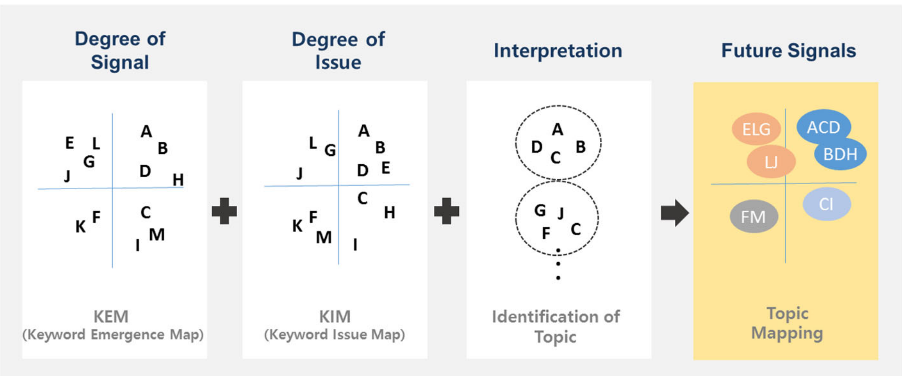

Fig. 3 Revised analysis framework

There remains additional arguments on how to interpret signals that locates in the remaining dimensions; low-left and low-right quadrant. Based on the attribute of occurrence frequency of the keywords in each quadrant, we defined keywords in the low-left quadrant as a *latent signals*, which show below average in share and growth rate. Likewise keywords in the low-right quadrant is defined as a *well know but not so strong signals*, which show above average in share but low growth rate. Detailed discussion on each type of future signals will be continued in the following case study.

3 Case study: ethical issues on artificial intelligence

Artificial intelligence technology and robotics are spreading into everyday areas such as leisure, sports, and traffic, which were before mainly used in specific fields such as military and finance. For example, major IT companies such as Google and Tesla competitively develop and test autonomous vehicles and global logistics companies such as Amazon are actively promoting various services such as delivery service using drones or other type of unmanned aircraft vehicle (UAV). Moreover, Artificial intelligence technology is starting to compete in a unique area of human beings beyond merely technical utilization. Recent developments in AI such as IBM's Watson and Google's AlphaGo[2] have dismayed the public; wherein AI programs outperformed the best human experts in quiz and chess games in terms of information retrieval and processing.

However, there are also cases where services or products incorporating AI technologies have resulted in social problems. Some noteworthy incidents involve a child injury by security robot at Palo Alto shopping mall,[3] malicious behavioral learning by Microsoft chat bot 'Tay',[4] and the death of an individual involving Tesla's self-driving car.[5] These series of events have raised concerns and worries over the positive expectations of AI technology. The issue commences from questions in relation to safety and accountability arising from accidents caused by or involving autonomous decision making by AI/robots, extending to job competition between humans and AI/robots, and further concerning threats to the human species.

Most recently, academia and field experts have responded in argument that humanity as a whole should prepare for socioeconomic side effects and normative issues arising from the proliferation of AI technology and robotics. The Western world has raised the need for creating social and institutional debate about the risk of AI. The US government through the White House released a report suggesting the need for preparation for the economic effects which may result from the proliferation of AI.[6] The EU has also debated the

[2] Sam Byford, 'Google's AlphaGo AI beats Lee Se-dol again to wi Go series 4-1' on The Verge (15 March 2016), http://www.theverge.com/2016/3/15/11213518/alphago-deepmind-go-match-5-result.

[3] Lilian Kim, 'Parents Upset after Palo Alto Security Robot Injures Child' on ABC7 (11 July 2016), http://abc7.com/news/parents-upset-after-norcal-security-robot-injures-child/1424537/.

[4] Samuel Gibbs, 'Microsoft's racist chatbot returns with drug-smoking Twitter meltdown' on The Guardian (30 March 2016), https://www.theguardian.com/technology/2016/mar/30/microsoft-racist-sexist-chatbot-twitter-drugs.

[5] Neal Boudette, 'Tesla's Self-Driving System Cleared in Deadly Crash' on The New York Times (19 January 2017), https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html?_r=0.

[6] Executive Office of the President of the United States, *Artificial Intelligence, Automation and the Economy* (Executive Office of the President, 2016), https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf.

possibility of affording robots legal personhood.[7] The global technical professional association, the IEEE, has also published the first version of a framework on how to ethically design AI system.[8]

Considering that AI technologies are still within their initial commercialization stage, it is necessary to predict what kind of ethical issues may arise from their use in the near future. So far, specifically in South Korea, forecasts and responses to the issues are mostly being led by a small group of experts such as scientists and institutions, with little government involvement. The lack of analysis of public awareness regarding ethical problems of AI/robots motivated the need to conduct the current study. While policymakers require objective evidence for the following questions, related information is still insufficient. Such questions include:

1. How can we evaluate current status of the AI related ethical issues in our society?
2. Which issues will be under spotlight to the extent that policy intervention is needed and when will the social consensus be formed?
3. Among the ethical issues related to AI that will emerge in the future, which one should we put policy priority?

Therefore, the purpose of the case study is to track the flow of social awareness on the ethical issues of AI, and draw implications for future policy directions. Firstly, we attempted to identify types of ethical issues related AI based on the future signal detecting method proposed in section II. We then sought to analyze which issues are under spotlight now and to emerge in the near future (Fig. 4).

3.1 Collection of web articles

Text data from online publication was utilized to search for future signals. The selected available data source is the EBSCO host database, which provides full texts from major English publications all over the world. Using this database, we conceived an article covering the ethical issues of AI and robots between 1994 and 2016. A total of 40 search keywords with high relevance to ethical issues were selected through preliminary search and pilot search test. Examples of search keywords are: ethics, humanity, algorithm, safety, kill, vehicle, etc. We created a search syntax with the combination of these keywords with ‘Artificial Intelligence’ and ‘robot’ using AND operator. An example of a search phrase is: [(“Artificial Intelligence” OR “A.I”) OR “Robot”] AND (“human rights” OR “ethic”).

A total of 708 pieces of text data were obtained after reviewing the relevancy to the research context. The number of documents counted separately for each year is shown in the following graph. As the number of documents prior to the 1990s was too small to find meaningful patterns, those period was regarded as one period. Thus a total of 18 periods were set from the 1990s to 2016 (Fig. 5).

As shown in the graph, documents dealing with the ethical issues of AI and robots appeared steadily until 2013, but the number of documents has dramatically increased since 2014. This increase is related to the start of competitive development of AI by IT companies such as IBM in earnest towards commercialization.

[7] Alex Hern, ‘give Robots Personhood Status, EU Committee Argues’ on The Guardian (12 January 2017), https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues.

[8] IEEE, ‘Robot Ethics’ on IEEE (2017), http://www.ieee-ras.org/robot-ethics.

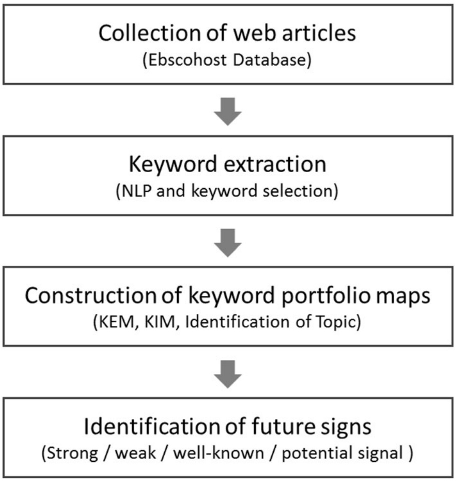

Fig. 4 Text mining procedure for detecting future signs

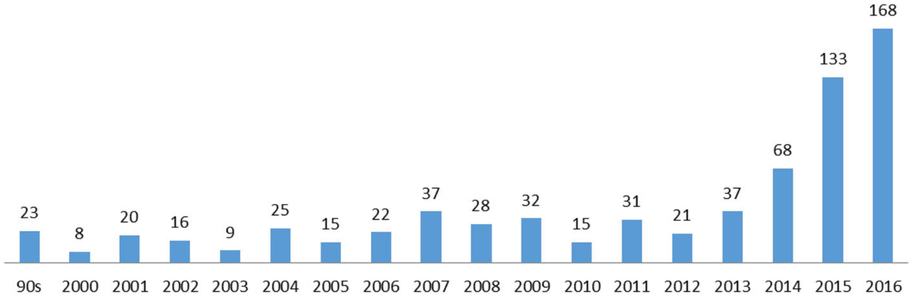

Fig. 5 The number of web documents

3.2 Keyword extraction

In the pre-processing procedure, we extracted morpheme from collected raw text using R programing software. Verbs, adjectives, and nouns were extracted as morphemes using Stanford NLP Engine to identify English words. Afterwards, TF and DF were calculated for each year. The total number of extracted keywords is 15,925. It would be not efficient to interpret all keywords in the data set as noise increase by keywords that are not relevant to the analysis context. Thus we set a dropout criterion to eliminate unnecessary keywords. First, we eliminated keywords that did not appear in succession over 4 years, reducing them to 1312. Next, the final 500 keywords were left as the target of analysis after manual examination and selection of keywords which are relevant to the analysis context.

3.3 Construction of keyword portfolio maps

DoV and DoD were calculated using the DF and TF values of the finally selected 500 keywords. According to the method of Yoon (2012), the keyword portfolios using KEM and KIM was constructed. The result is shown in Figs. 6 and 7.

As mentioned in the Sect. 2, future signals cannot be extracted at an appropriate level when the keywords classified by KIM and KEM are analyzed separately, For the

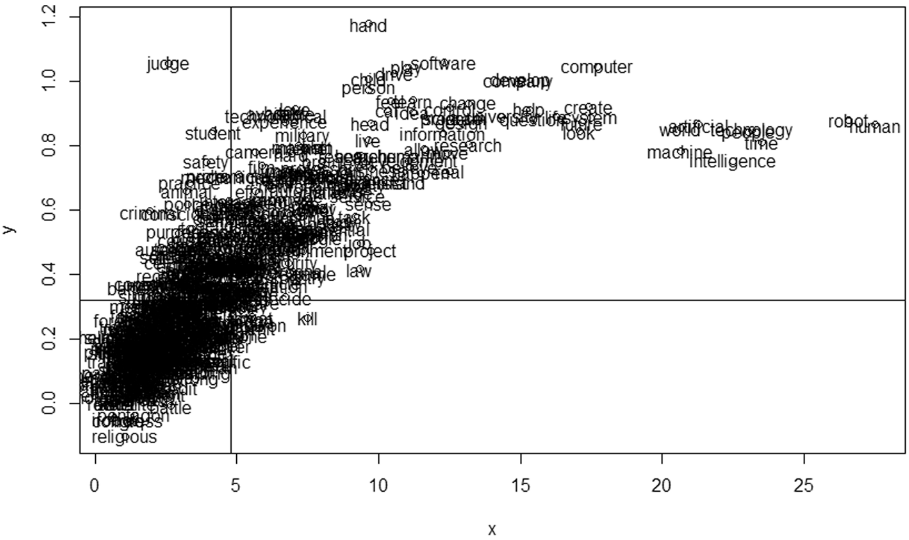

Fig. 6 Keyword emergence map regarding ethical issue in AI

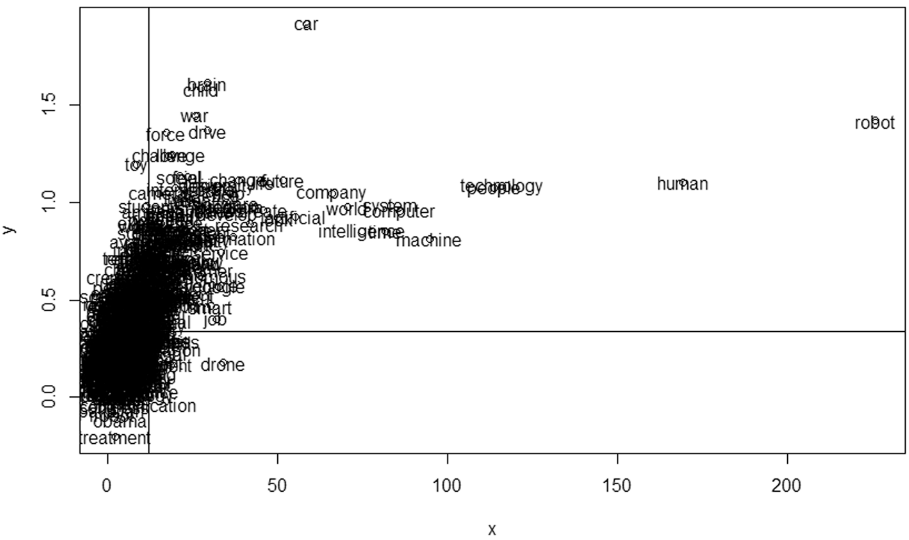

Fig. 7 Keyword issue map regarding ethical issue in AI

interpretation of future signals derived by a group of keywords, we decided to focus in particular on a set of keywords that commonly locate in the same quadrant in the KIM and KEM, respectively. First, we constructed a matrix of all possible keyword co-occurrence pairs using WORDij software (Danowski 2010), the program reads through the original text and identifies when specific words occur together, and pay attention to a group of keywords, (i.e., network clusters,) which occur commonly in the same quadrant. Secondly, we examined the keywords of each candidate cluster and confirmed whether the cluster should be determined as an issue topic through reviewing original texts. This method has

Table 1 Strong signals regarding emerging ethical issue in AI

Topic	Keywords	Definition
Privacy	Camera, security, automated, surveillance, drone, government, public, police	Privacy infringement and personal information leakage issues caused by artificial intelligence products and automated surveillance technology
Safety	Car, self-driving, unmanned, vehicle, safety, accident	Controversy related to accident risk and ethical algorithm of artificial intelligence specifically related to self-driving vehicle
Human abuse	Animal, abuse, crime, damage	Human abuse issue related to malicious learning, use in crime, violence against robots, etc.

Table 2 Weak signals regarding emerging ethical issue in AI

Topic	Keywords	Definition
Legal liability	Legal, liability, responsibility, conflict, car, drive, accident	The issues of locus of responsibility, scope of responsibility, legal issues including judicial process for accident caused by artificial intelligence
Killer robots	War, military, use, drone, robot, attack, software	Issues related to the development of combat robots with autonomous decision making ability applied with artificial intelligence
Singularity	Humanity, destroy, hawking, singularity, fear, overcome	A variety of disputes over the dystopia that comes when artificial intelligence transcends human ability

been used in research paper network (Doerfel and Barnett 1999; Stek and van Geenhuizen 2016), community structure (Girvan and Newman 2002; Jung and Valero 2016), and structure based on social relationship (Choi and Park 2015). This task was repeated for all candidate clusters, and the final 11 issue topics were selected and the name of each issue topic was defined.

3.4 Identification of future signal

For the final step, we classified each issue topic as strong signal, weak signal, not so strong but well-known signal, and latent signal. This classification was based on where most keywords of each issue topic locates at the quadrants of KEM and KIM. The contents of each topic were examined by reviewing the original articles that include keywords and additional literature review. The validity of the definition and classification of each topic was further verified through review of the several experts in information systems and peer researcher.[9] (Ponomareva and Sokolova 2015).

Strong signal Issues regarding 'privacy', 'safety', and 'human abuse' have received a lot of attention recently. A number of people have understood the problems and recognized its seriousness, and stakeholders are taking action accordingly. That is, the 'interpretation' of signals proceeded considerably. For example, several lawmakers in the U.S. have attempted to regulate the use of drones in compromising positions (Table 1).

Weak signal These issues relate to potential future problems, and that of AI in the future, when applications are realized in our daily lives. Many people have not yet felt any

[9] Peer reviewers include participants of WATEF International Conference 2016: DISC (data, innovation, social, and conversion).

Table 3 Not strong but well-known signals regarding emerging ethical issue in AI

Topic	Keywords	Definition
Imaginary fear	Race, scientific, fiction, scenario	Vague fears about the artificial intelligence and robots appeared in SF movies and comics, and
Identity	Creature, similar, synthetic, humanity, memory, humanoid	Identity confusion issue by the appearance of a humanoid robot

Table 4 Latent signals regarding emerging ethical issue in AI

Topic	Keywords	Definition
Final safeguard	Big red button, violence, evil, victim, damage, driving	Defense system against runaway of artificial intelligence and robotic systems that can be caused by imperfect technology
Emotional robot	Emotion, mimic, human, gesture, expression, heart, communication, child	The issue of the proliferation of human-emotional robots and their side effects
Decision by robot	Surgery, doctor, expert, replicate, brain, cognitive, software	Responsibility, socio-economic issues due to the emergence of robots supporting decision making

effects, and the impacts are being studied or discussed by a small but significant number of experts and institutions (Table 2).

Not so strong but well-known signal These are about vague fears and human identity crisis based on imaginations of the distant future. Issues are well known for many people as it has been being used as a material for many Science-Fiction films for decades. These issues have no specific direction whether to become strong signals or not (Table 3).

Latent signal Issues regarding ‘Final safeguard’, ‘Emotional robot’, and ‘Decision by robot’ have potential signals that cannot be determined as to whether they become weak signals. These issues might just vanish or just be repeatedly exposed to the public without any impact, that is, as a ‘not so strong but well-known signal’ (Table 4).

4 Conclusion

Attempts have been made to find patterns and predict the future from information that explodes in the Big Data era. However, the methodological foundation is still challenging. If one applies a methodology that lacks credibility when processing and analyzing large amounts of information, the prediction may be less accurate than relying on the intuition of some experts. Research on theoretical basis and analytical techniques for predicting the future based on quantitative evidence from big data is still at an early stage.

In this paper, a methodological contribution for detecting future signals was demonstrated; in that it provides an improved methodology for verifying the reliability and validity of analytical results by adding qualitative analysis to the existing quantitative analysis method. In the case study on the ethical issues of AI, we discovered that feasibility of the revised method is useful to evaluate the status of issue spreading from potential to emerging issues. Through the combination of quantitative and qualitative analysis, the revised methodology is, in nature, ambidextrous approach in which the analysts are able to balance between rigor and flexibility, which is widely accepted in various field of research and management (Lee et al. 2017). In practice, the method is expected to contribute to government and business stakeholders by providing a comprehensive view to the world we see now, encompassing from latent signals to widely known signals and their meanings as well.

Despite the contributions, there are limitations and rooms for future research. Starting from the analysis techniques presented in this study, more sophisticated methodology and theory development is necessary. Firstly, in the third stage of the revised framework in assessing the degree of 'interpretation' we only employed keyword clustering and topic selection method based on keyword co-occurrence. For in-depth understanding of how receivers interpret potential signals or weak signals, it needs to consider adoption of semantic network analysis (Cho et al. 2012; Sudhahar et al. 2015; Wang and Rada 1998; Wasserman and Faust 1994) with a time series approach. Semantic network analysis focuses on the structural relationship among words to capture meaning and is also useful for future study to reveal hidden phenomena, social perceptions, and foresight as time (Carley and Kaufer 1993). Integration of semantic network analysis in the framework and the validation is required in the follow-up study.

Another limitation, although it is not a scope of the current study, is regarding questions on how to understand the future potentials of the weak signals. Weak signal has a potential to be converted into a strong signal, not all of them have a same degree of potentials. It is an important research topic to be dealt with in depth as to in which condition a weak signal is likely to develop into a strong signal. One potential research topic of interest would be finding semantic network patterns in clustering keywords from weak or potential signal toward strong signal and patterns that present decrease in strength for the strong signal (Thorleuchter et al. 2014; Thorleuchter and Van den Poel 2015; Pratama 2017). Such efforts would immensely improve the accuracy for predicting the trajectory of future signs.

Appendix

See Table 5.

Table 5 Top 30 keyword with high frequency

Keywords	Total term frequency	Average term frequency	Total document frequency	Average document frequency	Degree of visibility (DOV)	Degree of diffusion (DOD)
Software	633	35.167	222	12.333	0.998	1.058
Computer	1550	86.111	319	17.722	0.952	1.046
Play	435	24.167	198	11.000	1.011	1.042
Drive	533	29.611	190	10.556	1.365	1.025
Child	497	27.611	174	9.667	1.582	1.008
Develop	635	35.278	270	15.000	0.935	1.005

Table 5 continued

Keywords	Total term frequency	Average term frequency	Total document frequency	Average document frequency	Degree of visibility (DOV)	Degree of diffusion (DOD)
Company	1190	66.111	269	14.944	1.041	0.993
Person	374	20.778	174	9.667	0.953	0.974
Learn	543	30.167	202	11.222	0.956	0.941
Feel	425	23.611	188	10.444	1.132	0.938
Change	695	38.611	237	13.167	1.113	0.931
Create	826	45.889	313	17.389	0.961	0.921
Control	557	30.944	225	12.500	1.068	0.916
Love	341	18.944	127	7.056	1.243	0.914
Help	652	36.222	276	15.333	1.042	0.909
Car	1052	58.444	186	10.333	1.914	0.908
Idea	468	26.000	202	11.222	1.014	0.903
Life	823	45.722	293	16.278	1.102	0.890
System	1498	83.222	315	17.500	0.982	0.887
Science	546	30.333	227	12.611	1.042	0.886
Robot	4060	225.556	478	26.556	1.417	0.880
Question	636	35.333	278	15.444	0.975	0.877
Program	656	36.444	228	12.667	0.961	0.877
Future	932	51.778	309	17.167	1.112	0.872
Experience	287	15.944	121	6.722	0.891	0.871
Head	317	17.611	175	9.722	0.937	0.870
Design	521	28.944	233	12.944	1.081	0.868
Human	3047	169.278	495	27.500	1.103	0.864
Artificial	996	55.333	383	21.278	0.924	0.864
World	1273	70.722	372	20.667	0.974	0.854

References

Ansoff, H.I.: Managing strategic surprise by response to weak signals. Calif. Manag. Rev. **18**(2), 21–33 (1975)

Carley, K., Kaufer, D.: Semantic connectivity: an approach for analyzing symbols in semantic networks. Commun. Theory **3**(3), 183–213 (1993)

Cho, S.E., Choi, M.G., Park, H.W.: Government-civic group conflicts and communication strategies: a text analysis of TV debates on Korea's import of U.S. beef. J. Contemp. East. Asia **11**(1), 1–20 (2012)

Choi, S., Park, H.W.: Networking interest and networked structure: a quantitative analysis of Twitter data. Soc. Sci. Comput. Rev. **33**(2), 145–162 (2015)

Choo, C.W.: The art of scanning the environment. Bull. Am. Soc. Inf. Sci. Technol. **25**(3), 21–24 (1999)

Choo, C.W., Auster, E.: Environmental scanning: acquisition and use of information by managers. Ann. Rev. Inf. Sci. Technol. **28**, 279–314 (1993)

Coffman, B.: Weak signal research. Part I. Introduction, MG Taylor Corporation (1997)

Dator, J. A.: Advancing futures: Futures studies in higher education. Greenwood Publishing Group (2002)

Danowski, J.A.: WORDij 3.0 [Computer Program]. University of Illinois at Chicago, Chicago (2010)

Doerfel, M.L.: What constitutes semantic network analysis? A comparison of research and methodologies. Connections **21**(2), 16–26 (1998)

Doerfel, M. L., Barnett, G. A.: A Semantic Network Analysis of the International Communication Association. Human Communication Research. **25**(4), 589–603 (1999)

Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)

Hellsten, I., Dawson, J., Leydesdorff, L.: Implicit media frames: automated analysis of public debate on artificial sweeteners. Public Underst. Sci. **19**(5), 590–608 (2010)

Hiltunen, E.: Was it a wild card or just our blindness to gradual change. J. Futures Stud. **11**(2), 61–74 (2006)

Hiltunen, E.: The future sign and its three dimensions. Futures **40**(3), 247–260 (2008)

Hong, S.W., Kim, Y.E., Bae, K.J., Park, Y.W., Park, J.K.: Development of analysis model for R&D environment change in search of the weak signal. J. Korea Technol. Innov. Soc. **12**(1), 189–211 (2009)

Hong, Y.J., Shin, D., Kim, J.H.: High/low reputation companies' dialogic communication activities and semantic networks on Facebook: a comparative study. Technol. Forecast. Soc. Change **110**, 78–92 (2016)

Hsu, C.L., Park, S.J., Park, H.W.: Political discourse among key Twitter users: the case of Sejong City in South Korea. J. Contemp. East. Asia **12**(1), 65–79 (2013)

Julien, P.A., Andriambeloson, E., Ramangalahy, C.: Networks, weak signals and technological innovations among SMEs in the land-based transportation equipment sector. Entrep. Reg. Dev. **16**(4), 251–269 (2004)

Jung, K., Valero, J.N.: Assessing the evolutionary structure of homeless network: social media use, keywords, and influential stakeholders. Technol. Forecast. Soc. Change **110**, 51–60 (2016)

Kamppinen, M., Kuusi, O., Soderlund, S.: Tulevaisuudentutkimus Perusteet ja Sovellukset (Futures Studies, Foundations and Directions). Suomalaisen Kirjallisuuden Seura, Helsinki (2002)

Kuosa, T.: Futures signals sense-making framework (FSSF): a start-up tool to analyse and categorise weak signals, wild cards, drivers, trends and other types of information. Futures **42**(1), 42–48 (2010)

Kuusi, O., Hiltunen, E.: The signification process of the future sign. J. Futures Stud. **16**(1), 47–66 (2011)

Lee, Y.-J., Lee, J.-H., Ham, K.S.: Balancing efficiency and flexibility in software project: the role of team collective improvisation, behavioral integration, and member diversity. J. Contemp. East. Asia **16**(1), 22–45 (2017)

Maurits Butter, M.L., Cagnin, C., Carabias, V., Könnölä, T., van Rij, V., Klerx, J., Schape Rinkel, P., Amanatidou, E., Saritas, O., Harper, J.C., Pace, L.: Scanning for early recognition of emerging issues; dealing with the unexpected, an operational framework for the identification and assessment of new future developments. Workshop paper: SESTI methodology, workshop 26 October 2010

Murtagh, F., Pianosi, M., Bull, R.: Semantic mapping of discourse and activity, using Habermas's theory of communicative action to analyze process. Qual. Quant. **50**(4), 1675–1694 (2016)

Ponomareva, J., Sokolova, A.: The identification of weak signals and wild cards in foresight methodology: stages and methods (No. WP BRP 46/STI/2015). National Research University Higher School of Economics (2015)

Pratama, A.B.: Online-based local government image typology: a case study on jakarta provincial government official YouTube videos. J. Contemp. East. Asia **16**(1), 1–21 (2017)

Smith, M.: Catalyzing social media scholarship with open tools and data. J. Contemp. East. Asia **14**(2), 87–96 (2015)

Stek, P.E., van Geenhuizen, M.S.: The influence of international research interaction on national innovation performance: a bibliometric approach. Technol. Forecast. Soc. Change **110**, 61–70 (2016)

Sudhahar, S., Veltri, G., Christianini, N.: Automated analysis of the US presidential elections using Big Data and network analysis. Big Data Soc. **2**(1), 1–28 (2015)

Thorleuchter, D., Van den Poel, D.: Technology classification with latent semantic indexing. Expert Syst. Appl. **40**(5), 1786–1795 (2013a)

Thorleuchter, D., Van den Poel, D.: Protecting research and technology from espionage. Expert Syst. Appl. (2013b). doi:10.1016/j.eswa.2012.12.051

Thorleuchter, D., Van den Poel, D.: Weak signal identification with semantic web mining. Expert Syst. Appl. **40**(12), 4978–4985 (2013c)

Thorleuchter, D., Van den Poel, D.: Idea mining for web-based weak signal detection. Futures **66**, 25–34 (2015)

Thorleuchter, D., Scheja, T., Van den Poel, D.: Semantic weak signal tracing. Expert Syst. Appl. **41**(11), 5009–5016 (2014)

Wang, W., Rada, R.: Structured hypertext with domain semantics. ACM Trans. Inf. Syst. **16**(4), 372–412 (1998)

Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)

Yoo, S.H., Park, H.W., Kim, K.H.: A study on exploring weak signals of technology innovation using informetrics. J. Technol. Innov. **17**(2), 109–130 (2009)

Yoon, J.: Detecting weak signals for long-term business opportunities using text mining of Web news. Expert Syst. Appl. **39**(16), 12543–12550 (2012)