

The Field of Social Robotics as Means of Technology Selection to Address Country Specific Social Issues

Cristian Mejia, Yuya Kajikawa

Tokyo Institute of Technology, Graduate School of Innovation Management, Japan

Abstract—This paper proposes a methodology to identify plausible robotic technologies to address country-specific social issues. Two stages are described. Firstly, we identify and rank country specific social issues by mining semantic relations in a newspaper database. A collection of news about Japan are analyzed through topic models and the social issues are extracted from the topics obtained. In the second stage those social issues are linked to robotic technologies by exploring the academic landscape of the social robotics, a field that has been serving to bridge the gap between robotics and society. A set of robotic technologies for the input country is shown as result. This methodology can be used by policy makers to identify technologies that best fit the necessities of a selected country.

I. INTRODUCTION

In the quest for bringing solutions to social needs there is the struggle on how to rightly identify both, the social concerns that affects most of the people, and the type of solutions that could be provided. It is needed to look for methods to identify social issues that are most relevant at present and bring solution to them timely. Also, some countries that share similar social issues may try to solve them by using different approaches, others may lack the capabilities to conduct research domestically and have to transfer solutions from overseas. In any case, there is the necessity of exploring the solution space to understand what have been done, what can be improved, or what option fits better to the problems.

As concluded by Gassmann and Zeschky [1] when studying product innovation through analytical thinking, some solutions might not be new for some, but can bring innovation to others, and decision makers should be open minded and promote the exhaustive search for external solutions and exploiting them.

The solution space is embedded in a system composed by several dimensions like markets, policy, technology, culture, science and industry [2], [3] ever interacting with each other. Among them, special attention has been given to technology and its relation to society [4], [5] as both affect each other in different parts of the innovation process [6] and the nature of this interactions determines the direction and speed of technological change [7]. In this paper we focus on technology, particularly in technology research and their connection to society. We propose a method to explore the possible solutions technology may bring to different social issues by using data mining techniques.

One of the outputs of technology research comes in the form of articles in academic journals, from where solutions can be extracted, but the amount of publishing is increasing

every day that we may found ourselves lost in the information sea [8]. The constant evolution and diversification of technology research makes the comprehensive understanding and finding of all possible solutions to a problem a difficult task. Even experts may overlook solutions that appear in distant disciplines. Data mining enables researchers and decision makers to overcome such concerns by applying techniques that reveal the intricate connections among several fields of knowledge otherwise hidden in the large databases of information.

Certainly, academic papers are not the sole output of technology research, but they have characteristics that facilitate large-scale analysis to reach conclusions faster. One of those characteristics are the cited references they contain, and the knowledge network that can be created of. These citation networks have been broadly studied [9], [10] and the use of them dates back to 1965 with the work of de Sola Price [11]. They have been used to map the landscape of science [12], create taxonomy of knowledge [13], find research and technology fronts [14] among other uses [15], [16]. Innovation research has also been subject of citation network analysis by Kajikawa et al [17]. In their work, research on “innovation” was studied through a network of papers, that was divided into tightly connected clusters that represented specific knowledge of the discipline. This analysis helped to understand the development of innovation studies and drawn recommendations for future directions. Citation networks have proven to be a sound methodology in the science and technology studies, and innovation.

In the mentioned works the unit of measure to create the citation networks is the academic article. A large corpus of papers enable researchers to map the knowledge of one or multiple disciplines from where solutions might be drawn. To extent the benefits of datamining to the analysis of society problems, there is the necessity to find the unit of measure to map across different social issues in different countries with the same versatility of papers in science and technology.

Newspapers are a good source of information about society. They cover topics that deserve debate, action and legal regulation [18], influence the formation of public opinion [19] and they are always up to date. Newspaper data has been used in innovation research to measure public expectations towards technology and to understand legitimacy in technology adoption [20]. In the work of Melton et al [21] international newspaper articles were used to measure the hype and disappointment cycles in relation to seven alternative fuel vehicle technologies to give recommendations towards the displacement of fossil fuels. Nevertheless, the use of this type of data has been scarce in

the literature, but its value and crescent interest of application in technology studies has been identified [22].

The challenge when analyzing newspaper articles is their unstructured nature, they lack of bibliographic references, and vary in size and vocabulary usage. To deal with that, we can apply topic models, a methodology that has been used to analyze this type of data in the social science and humanities [23]. Topic models are algorithms that extract the themes, or topics, that compose a collection of documents. Their automated nature facilitates the exploration of large corpus of text, an advantage over previous research that has been relied on the actual reading and manual coding or classification of news articles to reach conclusions. When applying a topic model in a collection of articles about a country or region, the model outputs the different topics that have received attention in the press, some of them are expected to contain social issues.

Topic models and citation networks are well developed methodologies, and we want contribute in this stream of research by bringing them together for the use of knowledge discovery by exploiting the text similarity of topics and networks. Concretely, this paper explores the possibility of finding solutions to social issues by exploiting the semantic relations between topic models from newspapers articles and citation networks of academic publications. The former representing society, and the later technology. The proposed methodology may help to scope over the large, and ever increasing, amount of technology research and extract from them plausible solutions to concerns in society.

We show our methodology by linking the topic model of newspaper articles about Japan to the network of academic papers on social robotics. Social robots are interactive partners that perform a social role [24]. They are also categorized as service robots, having as main characteristics a high level of complexity and autonomy [25]. The idea of linking robotics to social issues is not new, Andrade et al [26] reviewed robotic technologies for health care, and Ittipanuvat et al [27] for the elderly society. Our research contributes on

these explorations by assessing multiple social issues at once according to the necessities of the selected country.

The rest of the paper is as follows. We describe the process of creating the topic model and citation network, and how to establish semantic connections between them. The methodology is then applied to newspaper articles on Japan, and academic articles of robotics and their connections are analyzed to figure out the robotics technologies that may bring a solution to a social concern. We conclude the paper with a brief summary of the results and suggestions for future work.

II. METHODOLOGY

The overview of this research is presented in Fig. 1. It consists of 2 stages of data processing where social issues are extracted from newspaper articles, and robotic technologies are obtained from journal articles respectively. Finally, the connections between journal social issues and robotics are established by text similarity. Detailed explanation is provided in the following sections.

Stage 1: Topic Model

The first step is the data collection (1.1). The dataset of news articles related to a target country is extracted from a newspaper database. Next, a cleaning process is conducted on this dataset by removing stopwords and infrequent terms that appeared 5 times or less in the text corpus.

Once the news are obtained and cleaned, topic model is conducted to uncover the underlying themes in the corpus. Topic models are algorithms that analyze and classify the words in a corpus of unstructured text, like news articles, to find the underlying themes present in the set of documents [28]. This themes, or topics, can be described as sets of words that co-occur repetitively across the corpus. More precisely, a topic is a multinomial distribution over the total vocabulary found in all the documents [29].

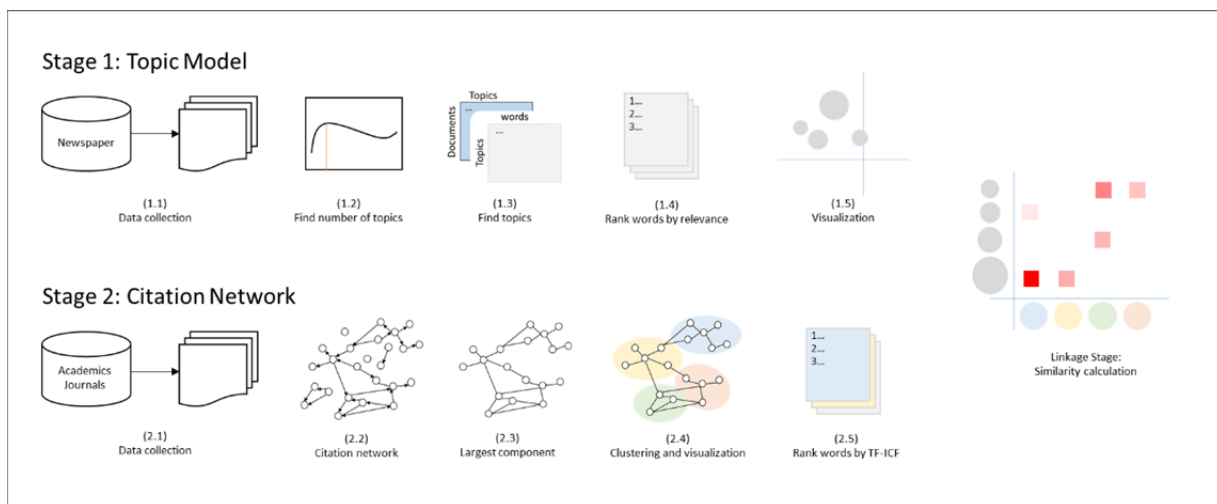


Fig. 1 Overview of this research. Topics from a topic model and clusters from a citation network are obtained and compared.

Similarly, documents are represented as multinomial distribution over topics. Simply put, documents are made of topics, and topics are made of words. In practice, the output of those algorithms are the document-topic matrix and the topic-word matrix filled with their respective conditional probabilities. Having that information, we are able to calculate the probability of a word w given a document d as follows:

$$p(w|d) = \sum_{t=1}^T p(w|t)p(t|d). \quad (1)$$

Where t stands for each of all possible T topics.

In fact, the news data we collected describes the left side of the equation, and we have to get the topics on the right side. There are many algorithms to achieve this, in this paper we use Gibbs sampling which is a Markov chain Monte Carlo algorithm that iterates over all words in the corpus to classify them into the desired T number of topics [30]. The Gibbs sampling algorithm is broadly used in topic modeling and has been applied on patents, journal articles, and newspaper corpus [19], [31], [32].

To analyze news articles using Gibbs sampling we need to set the number of topics T beforehand (step 1.2). This is not a trivial task, because too few or too many topics compromise their human interpretability. So that, the number of topics is selected by running the topic model using several alternatives of T and choosing the one with the highest log-likelihood, which at the same time serves as evaluation method for the model [33]. The Maximum likelihood in topic models can be described as the number of topics T that most likely reproduce the observed dataset, the news, when using (1).

Once the number of topics is chosen, the document-topic matrix and the topic-word matrix are extracted (step 1.3). The topic model can be considered finished by simply ranking the words by their probability. However, this has been proven to be a suboptimal solution and a better method of ranking words in topics called relevance (step 1.4) was proposed by Sievert and Shirley [34]. Relevance is a sorting function driven by a parameter that can take a value between 0 and 1, it represents a tradeoff between how exclusive the word is to the topic (relevance = 0) and how frequent a word appears in that topic (relevance = 1). The authors found that an optimal solution lies in about 0.6, thus we used that value of relevance in this paper.

The final step (1.5) is the visualization of the model. Distance between each pair of topics is calculated using the Jensen-Shannon divergence method, resulting in a T by T distance matrix. This is then reduced to T by 2 matrix using multidimensional reduction of the Principal Components. Once reduced, topics can be plot in 2-dimensional space. We used the LDAvis package in R to perform the visualization.

Stage 2: Citation Network

Journal articles are different from news articles in several aspects, one key difference is the use of bibliographic

references. Each academic paper is required to report previous works from where its results were built upon, creating a network of connections to previous papers. This citation networks, also called academic landscapes, have been used to study knowledge flows [35] and research fronts[36].

To create a citation network first we have to extract the bibliographic information of papers from a database of academic journals (step 2.1). Papers are treated as nodes, and each node gets connected to the papers they mention in their references (step 2.2). This is represented as arrows pointing backwards in time. In the following step, the directions of the arrows are removed and the largest connected component is extracted (step 2.3). We use the largest component because it captures the most information possible of the field we are analyzing.

In the next step (2.4), the network is clustered using a modularity maximization algorithm, which at the same time determines the best number of partitions for the network. As result, the citation network is divided in C number of clusters.

Finally, (step 2.5), the abstract of the papers in each cluster is analyzed and words ranked according to the term frequency – inverse cluster frequency TF-ICF weighting.

Linkage Stage

The previous two stages outputted T topics and C clusters, for news and papers respectively. The words in topics are ranked by their relevance weighting, whereas those in clusters are ranked by TF-ICF weighting. In this part of the process we establish semantic relations between topics and clusters by using cosine similarity.

Cosine similarity measures the similarity of two vectors by calculating the cosine of the angle between them. It is used in text mining, where documents are represented as vectors of size V , being V the size of vocabulary in the corpus (i.e. the set of unique words). Any type of weighting can be applied to the words as long as they are transformed into normalized vectors, and all documents share the same vector space. Therefore, to compute the cosine similarity we have to go through two preparation steps: Obtain the total vocabulary, and normalize the vectors.

The total vocabulary is simply the union of the vocabulary found in the topics and the one found in the clusters: $V_{total} = V_{topics} \cup V_{clusters}$. Then, topics and clusters are expressed as vectors of size V_{total} which are sparse, filled with their respective weights. Following, the vectors are normalized.

Finally, cosine similarity, which is the dot product of normalized vectors, is computed over all possible pairs of topic and cluster vectors. The result is a T by C similarity matrix that can be visualized as a heatmap, from where the most similar topic and cluster can be observed.

III. DATA

The methodology presented above is text driven. Given that the topics in newspaper articles are linked to academic papers in English, we chose an international newspaper to

source the information of Japan. The New York Times was selected as database because is worldwide known, regarded as a quality provider of news, and several academic research has been done on its corpus [21], [29].

News articles about Japan were extracted from The New York Times by using the Application Programming Interface (API) they provide. The New York Times API covers a comprehensive archive of international news dating back to September 1851 [37]. Furthermore, they keep a controlled vocabulary, also called Times Tags or Times Topic Pages, to refer to specific people, events or places. As query, instead of searching the news that contains the term “Japan” we search for the Times Topic “Japan” in order to be sure we get the news articles *about* Japan. Results were filtered to get only those of blog, business, education, food, front-page, health, open, opinion, public editor, science, technology, times topic, topics, and world sections (i.e. those sections that are unlikely to represent a social issue as obituary, dining & wine were not taken into account). Because we are interested in recent social issues we retrieved articles from January 2010 to November 2015, resulting a total of 1528 news articles.

For the papers, the Web of Science core collection by Thomson Reuters was used to retrieve bibliographic data about robotic technologies. The Web of Science is known as a reliable source of journal articles, indexing more than 12000 journals in basically all disciplines since 1970. The query was as follow “(social* or service* or personal or consumer) NEAR/2 robot*” spanning trough all years available. 3908 records were extracted.

IV. RESULTS

1528 news articles about Japan were extracted from The New York Times, and 3908 journal articles about social robotics were retrieved from the Web of Science. In this section we present the characteristics of each of those datasets and how they are related.

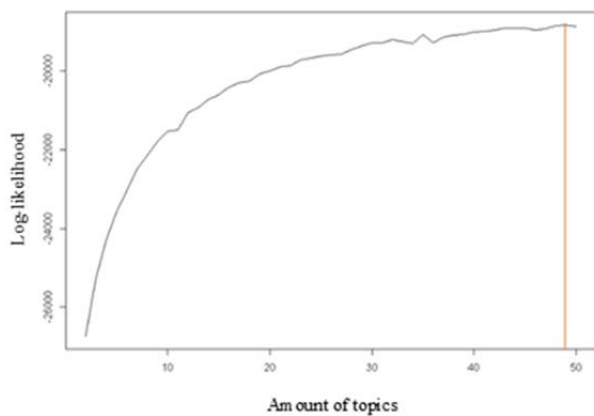


Fig. 2 Estimation of the number of topics based on maximum log-likelihood. The highest value is reached at 49 topics.

The news articles were analyzed through topic models. We fitted the model using different amount of topics in a range from 2 to 50, and determined that 49 topics was the best solution for the news data based on the maximum log-likelihood as is shown in Fig. 2.

The 49 topics about Japan are shown in Fig 3. Axes are the two principal components after multidimensional reduction representing relative distance among topics, those appearing closer each other are likely to be semantically related. Topics are ordered based on the size, which is an indication of how many words the topic collected in relation of all words in the corpus. Topics were named by human judgment based on their contents as shown in table 1. Not all topics are about social issues. Because we are mining news about a country it is expected to get all kinds of topics related to that country. So that, topics about Japanese films, politics and so on can be spotted in the model. The largest topic 1, seems to be meaningless. It collects general words that cannot be allotted in other topics. These type of topics is called junk topics, and do not contribute to the analysis. They are also expected and are subject of study for its automatic detection and elimination [38].

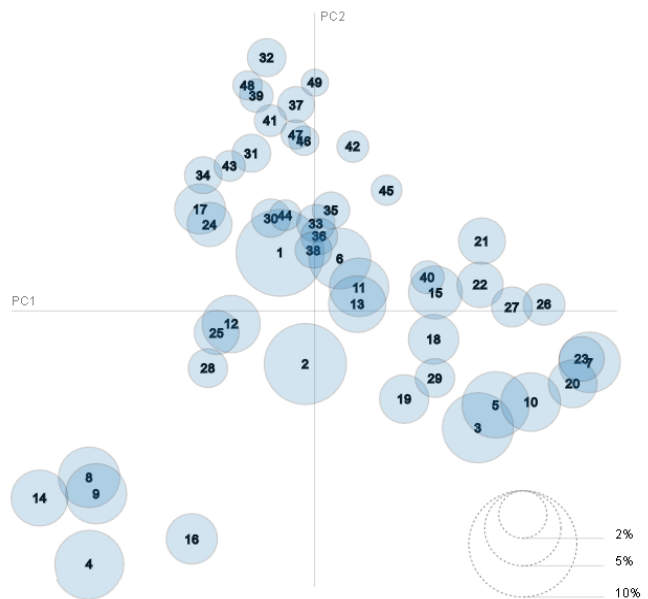


Fig. 3 49 topics about Japan obtained from news articles in The New York Times by applying topic models. Axes are the two principal component after multidimensional reduction. They represent relative distance.

The bottom-left corner of Fig.3 groups 5 topics all referring to nuclear energy, particularly about the issues after the Fukushima incident. It is clear, according to the news in our dataset, that nuclear energy and its implications have received great attention in Japan, and by grouping them together is the main issue observable. Other topics that can be regarded more strictly as social issue are 17.- Local migrations and 32.- Birthrate and elderly society. Nevertheless, many others also contain a *social* component.

TABLE 1
NAMES OF TOPICS AFTER MANUALLY INSPECTING THEIR CONTENTS.

List of topics		
1. General (Junk Topic)	17. Local Migration	33. Journalism
2. Government	18. Airspace	34. Undersea Exploration
3. US Diplomatic Relations	19. Political Scandals	35. Whaling
4. Fukushima Daiichi	20. Military	36. Investment
5. Senkaku Islands	21. World War II	37. Overseas Education
6. Japanese and Foreigners	22. Historical Relations with China	38. Middle East
7. China	23. War Apologies	39. Culture and Traditions
8. Concerns on Radioactivity	24. Damages on Infrastructure	40. North Korea
9. Nuclear Reactor Cooling	25. Earthquake and Tsunami	41. Internet
10. Prime Minister	26. South Korea	42. Law and Religion
11. Economics	27. United Nations	43. Climate Change
12. Accident and Disasters	28. Nuclear Energy Policy	44. Academic Research
13. Politics	29. U.S. Military Bases	45. Film
14. Radiation Leaks	30. Electronics Companies	46. Social Demonstrations
15. International	31. International Business	47. Sports
16. Nuclear Plant Reactivation	32. Birthrate and Elderly Society	48. Research on Health
		49. Literature and Arts

Table 2 shows the top words and news of selected topics. The junk topic collects words that do not convey meaning as a whole, and it is verified when looking at the news which seems to be unrelated. The other topics are coherent. Neglecting topic 1, topic model was able to map themes about Japan in a semantically meaningful way. Some of the topics may point to social issues, others are just an account of people, things or events. The next task is to link the topics to robotic technologies.

TABLE 2

Topic 1: General (Junk Topic)	
Top words	Question, people, don't, time, make, things, sense, point, hard, ways...
Top News	<ul style="list-style-type: none"> Living with Mistakes Is Nuclear Power Simply Too 'Brittle'? Ghosts in a Secular Age
Topic 9: Nuclear Reactor Cooling	
Top Words	Fuel, reactor, rods, cooling, containment, water, vessel, spent, hydrogen, pools...
Top News	<ul style="list-style-type: none"> A Look at the Mechanics of a Partial Meltdown In Fuel-Cooling Pools, a Danger for the Longer Term Chemistry 201: Why Is Fukushima So Gassy?
Topic 32: Birthrate and Elderly Society	
Top Words	School, children, women, parents, son, mother, care, female, marriage, elderly...
Top News	<ul style="list-style-type: none"> On The Liberal Marriage Hypothesis A New Ratio for the Japanese Cram School Desperate Hunt for Day Care in Japan

An academic landscape was created by exploiting the citation relations of the articles on social robotics. The largest component of such citation network is shown in Fig. 4. It consists of 1562 nodes, and 3079 edges, divided in 33 clusters. The figure also shows the top 5 clusters by number of nodes. In the network each cluster captures the research areas on social robotics. Table 3 shows the contents of selected clusters.

TABLE 3

COMPOSITION OF SELECTED CLUSTERS. TOP 10 WORDS AND TOP 3 PAPERS ARE SHOWN

Cluster 1: Robots for therapy and education	
Top words	Child, autism, asd, social, disorder, emotion, therapy, expression, imitation, autistic
Top papers	<ul style="list-style-type: none"> A survey of socially interactive robots [46] Socially intelligent robots: dimensions of human-robot interaction [47] Anthropomorphism and the social robot [48]
Cluster 2: Human Robot Interaction	
Top words	Older, social, child, acceptance, dementia, attitude, participant, assistive, paro, adult
Top papers	<ul style="list-style-type: none"> Interactive robots as social partners and peer tutors for children: A field trial [49] Acceptance of Healthcare Robots for the Older Population: Review and Future Directions [50] Social Robots for Long-Term Interaction: A Survey [51]
Cluster 16: Robots for Hazardous Environments	
Top words	Unmanned, climbing, inspection, façade, vehicle, tele, wall, industry, adhesion, sea
Top papers	<ul style="list-style-type: none"> Intelligent legged climbing service robot for remote maintenance applications in hazardous environments [52] Tele-operated climbing and mobile service robots for remote inspection and maintenance in nuclear industry [42] Comparing speed to complete progressively more difficult mobile robot paths between human tele-operators and humans with sensor-systems [53]

Semantic connections were established among papers and news articles by calculating the similarity of terms that conforms each topic and cluster. Only the 100 most relevant words (forming a total vocabulary of 5559) were used to compute the similarity matrix that is presented in the form of a heatmap in Fig. 5. The intensity of red imply more similarity. A higher similarity is an indication that they are semantically related. However, because the nature of language, it is possible that some pairs share words that are meaningless in the opposite context. The similarity matrix is just a tool to rank connected pairs, but they still have to be manually evaluated to assess the quality of their relation.

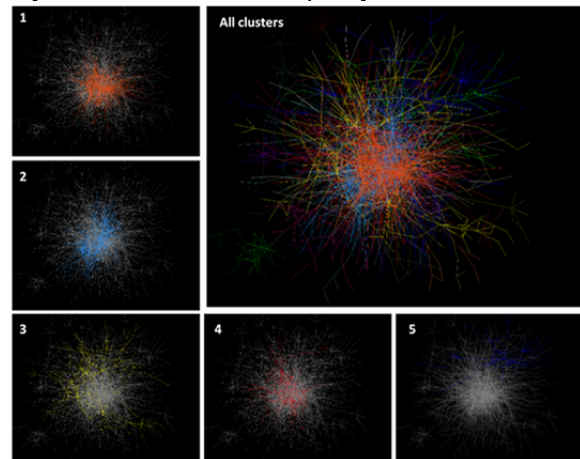


Fig. 4 Citation network of social robotics. The top 5 clusters are: 1. Robots for therapy and education, 2. Human-Robot interaction, 3. Mobility in closed space, 4. Robot interaction in the work place, 5. Robots for domestic tasks.

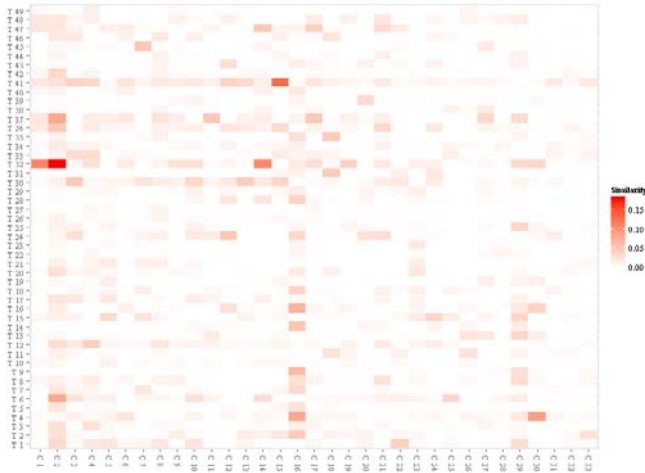


Fig. 5 Similarity between topics of Japan (y-axis) and clusters of social robotics (x-axis)

The 10 most similar topic and cluster pairs are shown in Fig 6. Three topics are associated to the clusters of human-robot interaction and robotics for hazardous environments. On the other direction 3 clusters point to the topic birthrate and elderly society. At first glance the linkage seems to have a good semantic relation, and a further analysis was conducted to verify their relatedness.

Birthrate and elderly society is a topic that collects several social issues. Some news were shown above in table 1, and some other headlines like “The Fertility Implosion”, “How in the World Will We Care for All the Elderly?”, “Japan’s Need for Women Workers”, appear in this topic along with some associated to children education. Each of them refers to known social concerns of Japan like fertility, elderly society, women empowerment, and others. This topic is linked to the clusters of human-robot interaction, robotics for therapy and education, and emotional attachment to robots, being a plausible connection between social issues and possible solutions through robotics. On the many possibilities of robotics for elderly care has already been reported by Ittipanuvat [27], we can highlight the cases of robot Paro [39] and Keepon [40] robots companions for the elderly and

children, as examples of the many socially interactive robots that belong to the clusters. On table 3 we indicated more papers suitable for this linkages. For the emotional attachment to robots cluster we can mention the work of Bainbridge [41] on the benefit of physical interaction with robots in games.

Another group of linkages is the one connecting nuclear issues to robots for hazardous environments. Examples of headlines in these topics are: “Accident at Fukushima Daiichi Nuclear Plant”, “Japan Restarts Nuclear Reactor”, “Japan Plans Safety Assessments of Nuclear Plants” and those shown in table 1. A clear concern on nuclear energy is observed. For the issue on safety on nuclear energy production robotics may provide some solutions. Cluster 16 refers to tele operated robots able to move in unstructured terrain or climb vertical surfaces, some of them specifically designed for nuclear environments [42]. Linkage to cluster 30 is also present, this contains indication on the design for safer robots [43].

Not all connections discovered have a problem – solution relation. The second highest similar linkage joins the topic of internet and the cluster ubiquitous robotics. While it is true that both share many words, they belong to different semantic space. The topic internet, contains news related to the Internet as a whole having headlines of the type “Hackers Find Way to Outwit Tough Security at Banking Sites” or “Virus Infects Computers in Japan’s Parliament” that are unrelated to the ubiquitous robotics. The cluster refers to technologies connecting the robot to multiple sensors in the environment [44] or to tele-operate robots through the Internet [45].

Lastly, the topics of Japanese and foreigners, and overseas education where linked to the cluster of human-robot interaction. Both topics contains stories of foreigners living in japan, or Japanese living abroad. This kind of news use several words conveying emotional meaning (e.g. friends, feel) that happen to be present in the human-robot interaction cluster as well. Because the topics do not relate to specific issues or concerns, no solution can be associated from the cluster.

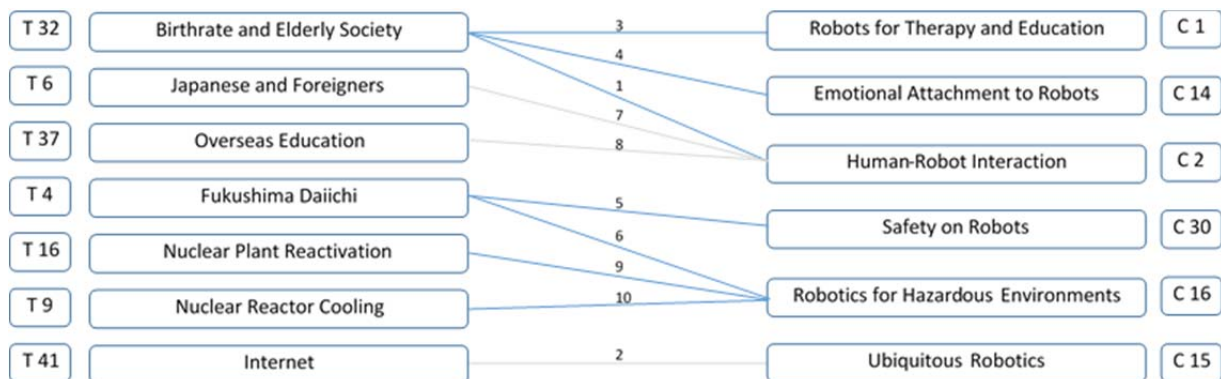


Fig. 6 Pairs of topics about Japan and clusters of social robotics having the highest similarity. The number in the connection is the ranking based on similarity, blue lines indicate a problem-solution relationship.

As shown above, linkage between topics and clusters by measuring the similarity of their text only, cannot be considered as evidence of a problem – solution relation for all connections. However, it is useful to recognize possible options and after inspection decide whether exist technologies that answers the problems observed.

V. DISCUSSION

In this paper we established semantic connections between topics about Japan found in an international newspaper and social robotic technologies published in academic articles. Semantic relations were helpful to connect local but broad issues of the selected country, to specific technologies in the robotics field. We used the robotics technologies as example, but other technology research fields may be used to connect the same topic model. Comparing the connections of multiple technologies to the topics of the same country might be useful for policy makers to spot research opportunities. By studying the topic model, important social issues (i.e larger topics) that don not show strong connection to technologies may serve as indicator of those opportunities. So that, the connection map either represented as heatmap or as the linkages in Fig. 6, might be beneficial for analyzing both, stronger and weaker relationships. Such comparison deserve deeper analysis that will be conducted in future research.

Limitations were found in this research. Newspaper articles about Japan were extracted from an international newspaper. Even though this source is regarded as reliable, the use of domestic newspapers may be expected. Some Japanese newspapers also offer information in English but the availability of those databases was limited and could not be used. Nevertheless, the topic model described here reveals the themes on Japan that are relevant from an international point of view, which is at the same time a hint of relevancy, because topics that get great attention locally are prone to be displayed internationally as well. More research is needed to understand the usage of international press as proxy to evaluate local issues. On the other hand, the advantage of international outlets of news is that one source may be used to analyze several countries, or regions. These opportunities are yet to be explored.

Extracting text similarity served as indicator of relatedness, but not all the observed connections shown a solution to a problem. The tasks of separating the social issues from general topics, and identifying solutions from the connections still depends on human reasoning, and further work will be required to understand how completely automate them. However, the methodology is helpful to reduce the effort of selecting the best alternatives.

Further work is needed to evaluate the performance of topic models compared to other text mining methodologies in the domain mapping and research fronts identification tasks

when analyzing academic articles. We will consider such questions in the future.

VI. CONCLUSIONS

In this paper we found robotic technologies that might be considered as solution to some social issues of Japan. Newspaper articles served as source of information to map topics related to Japan in the past 5 years, and from the topics the issues were extracted. On the other hand, different domains were obtained by clustering the citation network of academic articles of robotics. Linkage from the topics to robotic technologies were obtained by measuring the similarity in the text contents.

Japan's concerns on birthrate and elderly society were reflected in the topic model and linked to specific areas in robotics as emotional attachment to robots and human robot interaction, from where examples of companion robots were identified. Three topics related to nuclear energy issues were also identified and connected to robots for hazardous environments. While the problems were concentrated on Japan, the technology solutions were found internationally.

We do not claim all those connections can be considered new findings for the Japanese policy maker or researcher, on the contrary, part of that knowledge is expected. but we highlight the ease of discovering those connections by mining newspapers trough topic models and establishing a linkage by measuring content similarity. Topic models allow to study several topics related to a country at once, which can be used to assess the opportunity to respond to social problems by using technology already available, or to spot the necessity to develop new.

This paper explored the applicability of topic models for knowledge discovery. Newspapers report social issues that are relevant at present, thus the combination of topic models in news and citation networks in papers may help to bring solutions in the right timing. The results can be used by policy makers to identify technologies that contributes to solve social problems in their countries, and by researchers and managers to find new targets for their technologies.

ACKNOWLEDGMENTS

This research is supported by Grants-in-Aid for Scientific Research (A) (Grants ID:26245044), Japan Society for the Promotion of Science (JSPS), and by a MEXT scholarship.

REFERENCES

- [1] O. Gassmann and M. Zeschky, "Opening up the Solution Space: The Role of Analogical Thinking for Breakthrough Product Innovation," *Creat. Innov. Manag.*, vol. 17, no. 2, pp. 97–106, 2008.
- [2] F. W. Geels, "From sectoral systems of innovation to socio-technical systems," *Res. Policy*, vol. 33, no. 6–7, pp. 897–920, 2004.
- [3] B. Elzen, F. W. Geels, and K. Green, *System Innovation and the Transition to Sustainability. Theory, Evidence and Policy*. Edward

- Elgar, 2004.
- [4] R. Williams and D. Edge, "The social shaping of technology," *Res. Policy*, vol. 25, no. 6, pp. 865–899, 1996.
- [5] M. S. Jørgensen, U. Jørgensen, and C. Clausen, "The social shaping approach to technology foresight," *Futures*, vol. 41, no. 2, pp. 80–86, 2009.
- [6] J. Fleck and P. Edinburgh, "Innofusion or Diffusion? The Nature of Technological Development in Robotics," University of Edinburgh, Programme on Information and Communication Technologies, 1988.
- [7] B. Carlsson and R. Stankiewicz, "On the nature, function and composition of technological systems," *J. Evol. Econ.*, vol. 1, no. 2, pp. 93–118, 1991.
- [8] G. S. Davidson, B. Hendrickson, D. K. Johnson, C. E. Meyers, and B. N. Wylie, "Knowledge mining with VxInsight: Discovery through interaction," *J. Intell. Inf. Syst.*, vol. 11, no. 3, pp. 259–285, 1998.
- [9] F. Radicchi, S. Fortunato, and A. Vespignani, "Citation networks," *Underst. Complex Syst.*, pp. 233–257, 2012.
- [10] K. Borner and A. Scharnhorst, "Visual conceptualizations and models of science," *Journal of Informetrics*, vol. 3, no. 3, pp. 161–172, 2009.
- [11] D. J. De Solla Price, "Networks of scientific papers," *Science (80-)*, vol. 149, no. 3683, p. 510, 1965.
- [12] Y. Kajikawa, F. Tacao, and K. Yamaguchi, "Sustainability science: the changing landscape of sustainability research," *Sustain. Sci.*, pp. 431–438, 2014.
- [13] R. Klavans and K. W. Boyack, "Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?," p. 26, 2015.
- [14] N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima, "Detecting emerging research fronts based on topological measures in citation networks of scientific publications," *Technovation*, vol. 28, no. 11, pp. 758–775, 2008.
- [15] O. Persson, "Identifying research themes with weighted direct citation links," *J. Informetr.*, vol. 4, no. 3, pp. 415–422, 2010.
- [16] L. Leydesdorff and I. Rafols, "Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations," *J. Informetr.*, vol. 5, no. 1, pp. 87–100, 2011.
- [17] Y. Kajikawa, M. Hashimoto, I. Sakata, Y. Takeda, and K. Matsushima, "Academic Landscape of Innovation Research and National Innovation System Policy Reformation in Japan and the United States.," *Int. J. Innov. Technol. Manag.*, vol. 9, no. 6, p. -1, 2012.
- [18] S. Janssen, G. Kuipers, and M. Verboord, "Cultural Globalization and Arts Journalism: The International Orientation of Arts and Culture Coverage in Dutch, French, German, and U.S. Newspapers, 1955 to 2005," *Am. Sociol. Rev.*, vol. 73, no. 5, pp. 719–740, 2008.
- [19] P. DiMaggio, M. Nag, and D. Blei, "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding," *Poetics*, vol. 41, no. 6, pp. 570–606, 2013.
- [20] F. W. Geels and B. Verhees, "Cultural legitimacy and framing struggles in innovation journeys: A cultural-performative perspective and a case study of Dutch nuclear energy (1945-1986)," *Technol. Forecast. Soc. Change*, vol. 78, no. 6, pp. 910–930, 2011.
- [21] N. Melton, J. Axsen, and D. Sperling, "Moving beyond alternative fuel hype to decarbonized transportation," *Nat. Energy*, vol. 1, no. February, pp. 1–29, 2016.
- [22] D. Rotolo, D. Hicks, and B. Martin, "What is an emerging technology?," *SPRU Work. Pap. Ser.*, vol. 2015–06, no. 10, pp. 1–40, 2015.
- [23] J. W. Mohr and P. Bogdanov, "Introduction-Topic models: What they are and why they matter," *Poetics*, vol. 41, no. 6, pp. 545–569, 2013.
- [24] F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer, "Understanding Social Robots," *Second Int. Conf. Adv. Comput. Interact.*, no. Section II, pp. 169–174, 2009.
- [25] T. Haidegger, M. Barreto, P. Gonçalves, M. Habib, S. K. V. Ragavan, H. Li, A. Vaccarella, R. Perrone, and E. Prestes, "Applied ontologies and standards for service robots," *Rob. Auton. Syst.*, vol. 61, no. 11, pp. 1215–1223, 2013.
- [26] A. O. Andrade, A. a. Pereira, S. Walter, R. Almeida, R. Loureiro, D. Compagna, and P. J. Kyberd, "Bridging the gap between robotic technology and health care," *Biomed. Signal Process. Control*, vol. 10, pp. 65–78, 2014.
- [27] V. Ittipanuvat, K. Fujita, I. Sakata, and Y. Kajikawa, "Finding linkage between technology and social issue: A Literature Based Discovery approach," *J. Eng. Technol. Manag. - JET-M*, vol. 32, pp. 160–184, 2014.
- [28] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models," *IEEE Signal Process. Mag.*, vol. 27, pp. 55–65, 2010.
- [29] D. Newman, C. Chemudugunta, P. Smyth, and M. Steyvers, "Analyzing entities and topics in news articles using statistical topic models," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3975 LNCS, pp. 93–104, 2006.
- [30] T. Griffiths and M. Steyvers, "Finding scientific topics.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101 Suppl, pp. 5228–5235, 2004.
- [31] Z. Hu, S. Fang, and T. Liang, "Empirical study of constructing a knowledge organization system of patent documents using topic modeling," *Scientometrics*, vol. 100, no. 19, pp. 787–799, 2014.
- [32] M. Rosen-Zvi, M. Rosen-Zvi, T. Griffiths, M. Steyvers, P. Smyth, and P. Smyth, "The author-topic model for authors and documents," *Proc. 20th Conf. Uncertain. Artif. Intell.*, pp. 487–494, 2004.
- [33] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, "Evaluation methods for topic models," *Proc. 26th Annu. Int. Conf. Mach. Learn. - ICML '09*, no. 4, pp. 1–8, 2009.
- [34] C. Sievert and K. Shirley, "LDavis: A method for visualizing and interpreting topics," *Proc. Work. Interact. Lang. Learn. Vis. Interfaces*, pp. 63–70, 2014.
- [35] I. Sakata, H. Sasaki, M. Akiyama, Y. Sawatani, N. Shibata, and Y. Kajikawa, "Bibliometric analysis of service innovation research: Identifying knowledge domain and global network of knowledge," *Technol. Forecast. Soc. Change*, vol. 80, no. 6, pp. 1085–1093, 2013.
- [36] N. Shibata, Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, "Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications," *Technol. Forecast. Soc. Change*, vol. 78, no. 2, pp. 274–282, 2011.
- [37] The New York Times, "Times Developer Network - Article Search API v2." [Online]. Available: http://developer.nytimes.com/docs/read/article_search_api_v2. [Accessed: 02-Feb-2016].
- [38] L. AlSumait, D. Barabá, J. Gentle, and C. Domeniconi, "Topic significance ranking of LDA generative models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5781 LNAI, no. PART 1, pp. 67–82.
- [39] W. L. Chang, S. Šabanović, and L. Huber, "Use of seal-like robot PARO in sensory group therapy for older adults with dementia," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2013, pp. 101–102.
- [40] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon: A playful robot for research, therapy, and entertainment," *Int. J. Soc. Robot.*, vol. 1, pp. 3–18, 2009.
- [41] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *Int. J. Soc. Robot.*, vol. 3, no. 1, pp. 41–52, 2011.
- [42] B. L. Luk, K. P. Liu, A. A. Collie, D. S. Cooke, and S. Chen, "Tele-operated climbing and mobile service robots for remote inspection and maintenance in nuclear industry," *Ind. Rob.*, vol. 33, no. 3, pp. 194–204, 2006.
- [43] R. Woodman, A. F. T. Winfield, C. Harper, and M. Fraser, "Building safer robots: Safety driven control," *Int. J. Rob. Res.*, vol. 31, no. 13, pp. 1603–1626, 2012.
- [44] R. B. Rusu, B. Gerkey, and M. Beetz, "Robots in the kitchen: Exploiting ubiquitous sensing and actuation," *Rob. Auton. Syst.*, vol. 56, no. 10, pp. 844–856, 2008.
- [45] D. Schulz, W. Burgard, D. Fox, S. Thrun, and A. B. Cremers, "Web interfaces for mobile robots in public places," *Robot. Autom. Mag. IEEE*, vol. 7, no. 1, pp. 48–56, 2000.
- [46] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Rob. Auton. Syst.*, vol. 42, pp. 143–166, 2003.
- [47] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 362, no. April, pp. 679–704, 2007.

- [48] B. R. Duffy, "Anthropomorphism and the social robot," *Rob. Auton. Syst.*, vol. 42, pp. 177–190, 2003.
- [49] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial," *Human-Computer Interact.*, vol. 19, pp. 61–84, 2004.
- [50] E. Broadbent, R. Stafford, and B. MacDonald, "Acceptance of healthcare robots for the older population: Review and future directions," *Int. J. Soc. Robot.*, vol. 1, pp. 319–330, 2009.
- [51] I. Leite, C. Martinho, and A. Paiva, "Social Robots for Long-Term Interaction: A Survey," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 291–308, 2013.
- [52] B. L. Luk, D. S. Cooke, S. Galt, A. A. Collie, and S. Chen, "Intelligent legged climbing service robot for remote maintenance applications in hazardous environments," *Rob. Auton. Syst.*, vol. 53, no. 2, pp. 142–152, 2005.
- [53] D. Sanders, "Comparing speed to complete progressively more difficult mobile robot paths between human tele-operators and humans with sensor-systems to assist," *Assem. Autom.*, vol. 29, no. 3, pp. 230–248, Jul. 2009.