IDENTIFICATION OF TECHNOLOGY-RELEVANT ENTITIES BASED ON TREND CURVES

Sukhwan Jung, Rachana Reddy Kandadi, Rituparna Datta,
Ryan Benton and Aviv Segev

Department of Computer Science, University of South Alabama,
Mobile, Alabama, USA

ABSTRACT

Technological developments are not isolated and are influenced not only by similar technologies but also by many entities, which are sometimes unforeseen by the experts in the field. The authors propose a method for identifying technology-relevant entities with trend curve analysis. The method first utilizes the tangential connection between terms in the encyclopedic dataset to extract technology-related entities with varying relation distances. Changes in their term frequencies within 389 million academic articles and 60 billion web pages are then analyzed to identify technology-relevant entities, incorporating the degrees and changes in both academic interests and public recognitions. The analysis is performed to find entities both significant and relevant to the technology of interest, resulting in the discovery of 40 and 39 technology-relevant entities, respectively, for unmanned aerial vehicle and hyperspectral imaging with 0.875 and 0.5385 accuracies. The case study showed the proposed method can capture hidden relationships between semantically distant entities.

KEYWORDS

Technology Forecasting, Trend Curve, Big Data, Academic Articles, Web Pages

1. INTRODUCTION

Identification of relevant terms for a specific technology plays a crucial role in technology funds and government research grants, allowing them to better direct their investments to encourage technological developments beneficial to the target technology. It is also one of the main research fields for stock market prediction as technology development directions affect the stock markets. The current work proposes an approach for identifying any type of entities relevant to the given technology, based on the trend curves of related entities found from recursive encyclopedic connections to the technology. This perspective offers a novel approach of technology trend analysis, granting a possibility of detecting seemingly unrelated entities that cannot be found with conventional means.

The proposed method offers a means of identifying significant entities relevant to a given technology based on term frequency and degree of usage growth. It analyzes technology-relevant entities from Wikipedia in the whole domain of academic articles (academia) and web pages (web) with the help of Google search engine, incorporating both the academic interests and public recognitions of the given entities, each representing the earliest and the latest predictive time windows. Wikipedia, an online encyclopedia with built-in page links between related

David C. Wyld et al. (Eds): ITCSE, NLCA, ICAIT, CAIML, ICDIPV, CRYPIS, WiMo - 2020
pp. 01-13, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.100501

articles, allows the extraction of not only the related technologies but also any related entities, providing generalizability to the proposed method if desired. The authors previously showed technology trends in different datasets contain distinct patterns while sharing an overall shape on different time windows [1]. The comparison of the proposed method on two datasets analyzes the differences in the list of entities deemed relevant to them. In addition, the analysis of entities common in both datasets and their respective trend curves presents an integrated view of the technology-relevant entities over multiple dimensions.

The main contributions of this work are as follows:

- On an algorithmic level, the authors provide an implementation of the proposed method based on the academia and web.
- On a conceptual level, the authors propose a multi-domain approach for identifying any entities relevant to a specific technology based on term frequency and moving gradient, which can be semantically and syntactically unrelated to the technology.
- On a practical level, the authors identify a list of relevant, possibly hidden, entities for the target technology and how different datasets contribute to the result.

Section 2 reviews the related work on technology forecasting with regard to the necessity of normative approaches based on technological curves and their limitations. Section 3 explains the proposed method and experiment in detail. The experiment results in Section 4 show that the proposed method can identify entities related to the given technology with hidden relationships, and Section 5 states the concluding remarks and future work.

2. RELATED WORK

The traditional approach for technology forecasting is a manual approach, including scenario building [2], forecast by analogy [3], and the Delphi method [4]. Scenario building lets analysts generate a series of plausible scenarios with both optimistic and pessimistic developments; these developments aim to be compatible, with substantial effects, with unlikely events that are often disregarded in other methods. Forecast by analogy employs analogical comparison between the known phenomena and the technology trends with the assumption they behave similarly. The Delphi method is a more structured technique, first developed as a systematic and interactive method of forecasting. It relies on a consensus among a panel of experts, which is reached by repeated rounds of questionnaires to the participating experts. The belief is that the variance of the answers will decrease with each iteration and the group will converge towards an answer that can be regarded as correct. The process ends once it either reaches a certain number of rounds or achieves a steady consensus; the answers from the final round determine the result. These manual methods often require a large amount of contribution from numerous field-related experts and hence are expensive to utilize, but still have been used in recent years [5] for its high domain adaptability.

Normative methods such as morphological models [6] and mission flow diagrams [7] are complementary to such processes that attempt to automatically project future behaviors from past data. Based on systems analysis, normative methods view future needs in the field as the scheduled progress of the field and predict future behaviors based on them [7]. Extrapolations on past data are used to analyze changes in the popularity or intensity of a given topic, which can be matched into estimation lines such as linear, polynomial, exponential, and parabolic lines [8]. Extrapolation on a pre-defined technological curve is widely used as well, matching the past data to estimations lines such as Gartner’s hype cycle or other technological growth curves such as S-curve [9]. The future technological stages are then predicted upon the estimation line. The limitations of normative methods suggested in more recent years include incongruencies found from the Gartner dataset and its hype cycle [10] and less generalizability for different technology fields. This indicates that both the manual and extrapolation methods lack the ability to be implemented in related technological fields [11].

Other fields of research tried to address related fields to generate better technology forecasts. The content transition from one topic to another during topic evolution is identified in the form of complementary trend curve patterns, connecting multiple technologies in transitional states [12], [13]. The topics are extracted statistically from a document collection, and the popularity trend curve of each topic is generated by connecting their popularities in discretely divided document subsets per timeslots. The content transition between topics is identified when one topic experiences a significant drop in popularity when the other topic experiences a significant increase, which is translated as the former topic being transferred to the latter one. Technology diffusion can be used for a more specific case of technology transfer where the one is replaced by another, such as LED is replaced by OLED for the TV screen market [14]. The inconsistency problem remains, however; the technology trend curves can vary for different forecast methods and datasets on which they are used. Combining different forecasts of the same technology allows remediation of the disadvantages from individual forecasts at the potential expense of individual advantages [8]. The authors' previous work utilized a combination of forecasts from various datasets to show that the predictive power of different forecasts varies based on the nature of the dataset used. Changes in technology term frequencies in public datasets such as news, books, and web pages are preceded by more academic datasets such as academic articles and patents, resulting in a longer predictive time window for technology growth prediction [1].

The traditional manual approach for technology forecasting requires an extensive amount of time and resources, and cheaper alternatives are highly sought after. Normative methods extrapolate on predefined technological growth curves were successful in forecasting technological development within a given technology field. They showed limited performance in forecasting related technologies. Different technologies do not necessarily follow the same growth curve. However, our work proposes a more generic forecasting model for automatic technology forecasting.

3. METHOD

The proposed method is based on analyzing the frequency trends of technology-relevant entities on academia and web each representing two different dataset orientations – academic and public. Documents in both datasets are timestamped by their publication date and can be sequentially discretized. The method consists of 1) extraction of technology-related entities having recursive encyclopaedic connections to the technology in question, and 2) identification of technology-relevant entities through the entity filtering with their timeline trend curves over two different datasets, incorporating both academic and public interests. The analysis was performed for two selected technologies, Unmanned Aerial Vehicle (UAV) and Hyperspectral Imaging (HSI), identifying relevant entities from Wikipedia articles using the entirety of academia and web. In addition, the entities were evaluated manually to showcase the necessity of multi-datasets and the possible applications of the proposed method.

3.1. Extracting Technology–Related Entities

The technology-related entities were extracted based on the Wikipedia articles. The semi-structured nature of the articles allow multiple extraction approaches; advanced natural language processing such as context recognition can be used to extract terms from the unstructured texts from the articles [15], structured data such as infobox tables or links can be utilized to extract pre-defined terms, and the articles can be read to manually identify the related entities. The see-also section of the Wikipedia article is a list of internal pagelinks manually written by participants and moderators. The see-also section was used in this experiment as its semi-structured nature allows the extracted terms to be not limited to specific contents, domains, or types while providing human-verified semantic, syntactic, or conceptual connections between the original and linked articles.

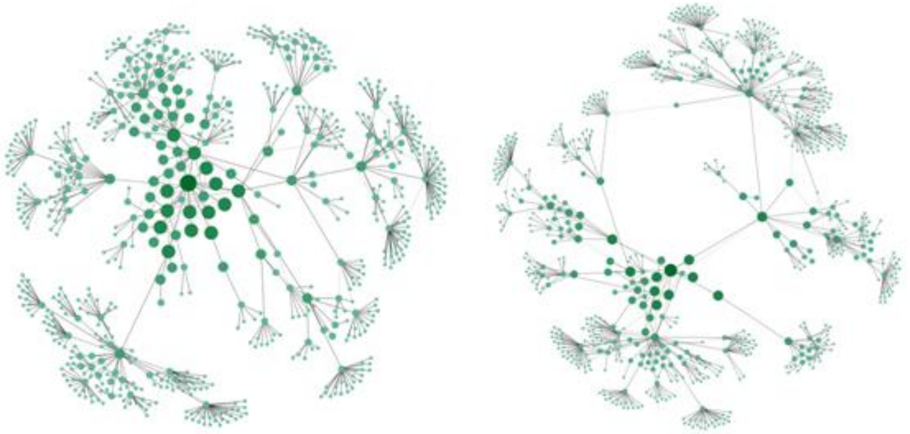
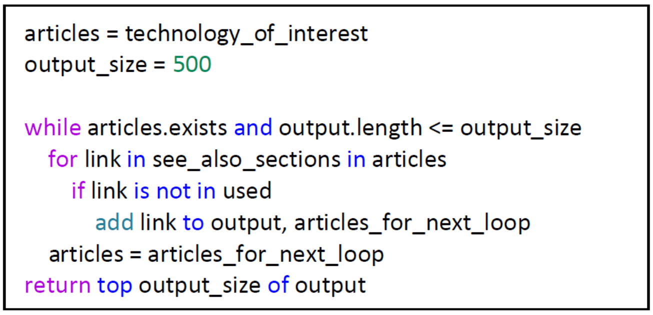


Figure 1(a). Related to *UAV*.

Figure 1(b). Related to *HSI*.

Figure 1. Visualizations of technology-related entities connected by see-also relationships with diminishing node size and color for entities more distant from the given technology.

Given a specified technology, such as UAV, its related entities are extracted by recursively parsing the Wikipedia articles starting from its article. The dedicated library for reading Wikipedia[1] is not ideal when multiple articles are considered, and a different approach is utilized instead. The Wikipedia article is retrieved via the webserver using a specified URL which is then processed with a Python library BeautifulSoup[2] to extract HTML snippet for the see-also section and the pagelinks contained within it. The articles from the collected pagelinks become the first set of technology-related entities with a distance of 1 from the seed article. The algorithm is then run recursively on the extracted articles for breadth-first entity extraction; the see-also sections for articles with distance = n are extracted to get entities with distance = n+1. The recursive search result in exponential growth is the number of entities found.



```
articles = technology_of_interest
output_size = 500

while articles.exists and output.length <= output_size
  for link in see_also_sections in articles
    if link is not in used
      add link to output, articles_for_next_loop
  articles = articles_for_next_loop
return top output_size of output
```

Figure 2. Pseudocode for Extracting Technology-Related Entities.

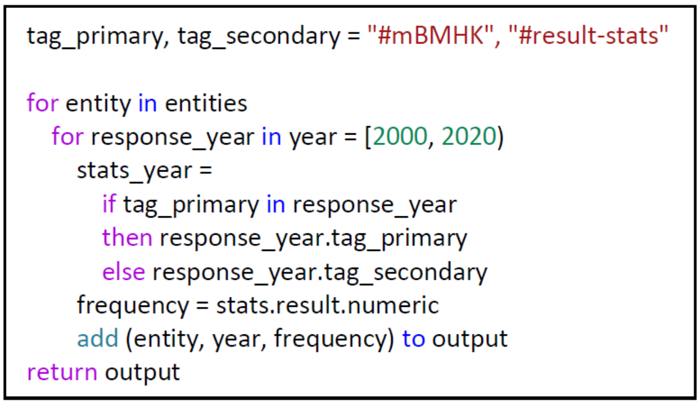
[1] https://pypi.org/project/wikipedia/
[2] https://code.launchpad.net/beautifulsoup/

Examples of entities with distance ≤ 4 are displayed in Figure 1(a) and Figure 1(b) based on two technologies, with articles as nodes and see-also connection as links. The node size and color intensity reflect the distance from the root node, and the graph shows mostly tree structures with only a fraction of the links between branches; such a link indicates that the articles were inversely connected. They show the majority of entities, 71.39% for UAV and 76.49% for HSI, are the furthest from the technology with distance = 4. The exhaustive search can be done for

longer distances for more than a quarter of million entities but is impractical; the authors used the first 500 results as the technology-related entities which can be satisfied with distance ≤ 4. The pseudocode for extracting technology-related entities is shown in Figure 2, where articles are recursively searched until the given number of related articles, 500 in the experiment, are collected. Breadth-first search is done for each of the links in the articles' see-also section. The possibility of cycling is removed by only accessing newly-met articles in the process.

3.2. Extracting Technology–Related Entity Trends

The next step is the extraction of trend curves of the list of the entities found from the previous step. This is achieved by extracting their term frequencies in the large document collections at discrete timeslots, which, in this study, was yearly intervals. The whole domains of academia and web were chosen as the document collections in the experiments. The sheer volume of research publications of the WWW hinders effective searching, and the Google search engine is utilized which searches the documents indexed by Google, respectively exceeding 389 million articles [16] and 60 billion pages [17]. This allowed utilization of Google search engine APIs during the trend curve extraction process, where each data point is the number of search results in a given year. The search result for academia is the number of academic articles containing the term in their titles and abstracts, or full texts when the Google API can access them. For web, the total number of webpages containing the term is used instead. The trend curves are generated by connecting the discrete data points into a series of line graphs. The trend curves are not normalized as in the previous research [18] since the process searches not only for curves with a specific growth pattern but also curves with overall elevated values. All entities are deemed related to the technology in question and are treated equally regardless of their distances from it, or the number of see-also sections between them.



```
tag_primary, tag_secondary = "#mBMHK", "#result-stats"

for entity in entities
  for response_year in year = [2000, 2020)
    stats_year =
      if tag_primary in response_year
      then response_year.tag_primary
      else response_year.tag_secondary
    frequency = stats.result.numeric
    add (entity, year, frequency) to output
return output
```

Figure 3. Pseudocode for Extracting Technology-Related Entity Trends.

Figure 3 shows the pseudocode for the entity trends extraction process. For each entity obtained in the previous section, the Google search result in HTML format is retrieved for every year from 2000 to 2019. The statistical metadata of the response is stored within a HTML div tag identifiable by two possible ids, result-stats and mBMHK, which is extracted as a snippet. The search result count within the snippet is then extracted and stored as the frequency for the year. The only difference between academia and web is the structure of the URL the Google search engine requires; therefore the same implementation is used for both. The number of calls to the Google search engine is limited to 100 per day, and queries were required to be made every 100 seconds.

Results of entity trends extraction are shown in Figure 4 with four graphs. Entities related to both the UAV and HSI share similar patterns, diminishing towards the year 2019 after plateauing at around 2010 in academia in Figure 4(a) and Figure 4(c) while showing exponential growth in web in Figure 4(b) and Figure 4(d). This suggests the entities related to both technologies are experiencing initial hype with the public while the researchers have already passed this stage and show diminished interests in the same entities. Such differences are validated by the authors' previous research on the different time windows for technology growth curves in different datasets, where the technology's development starts with the academic domain and the public inherits the changes afterward [1], [18].

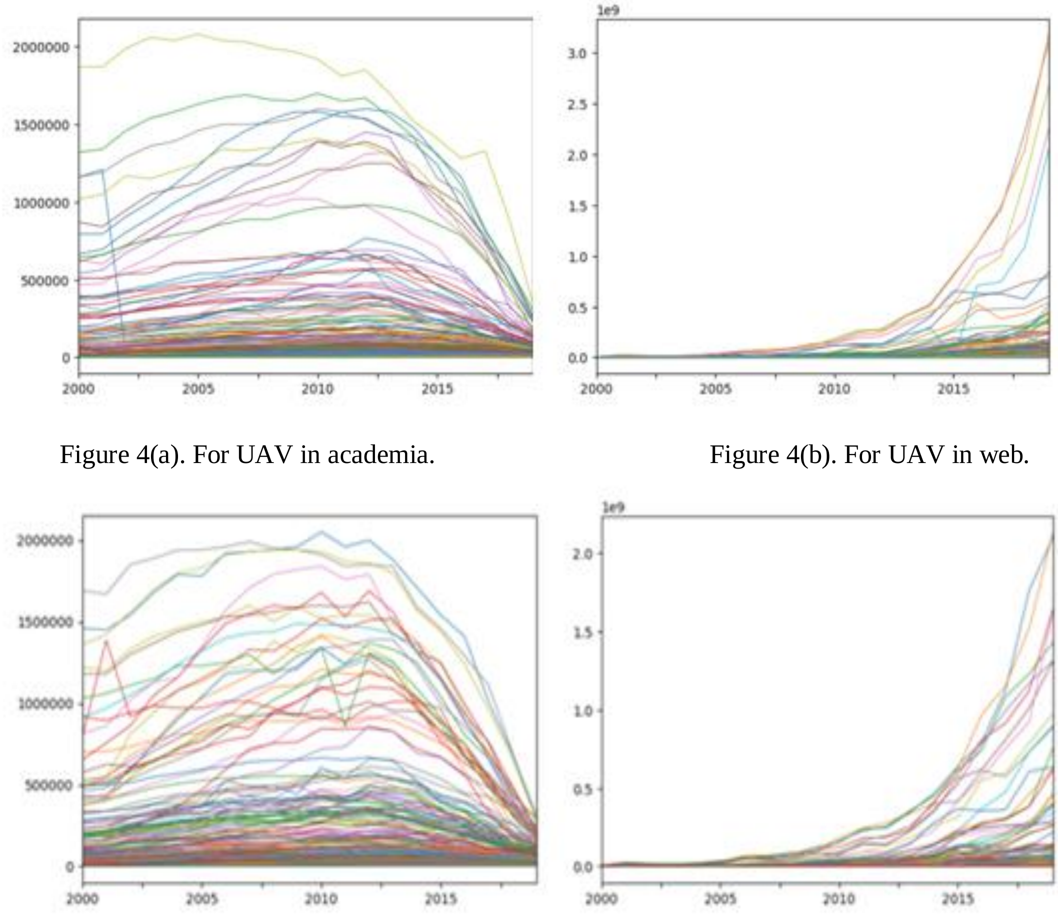


Figure 4(a). For UAV in academia.

Figure 4(b). For UAV in web.

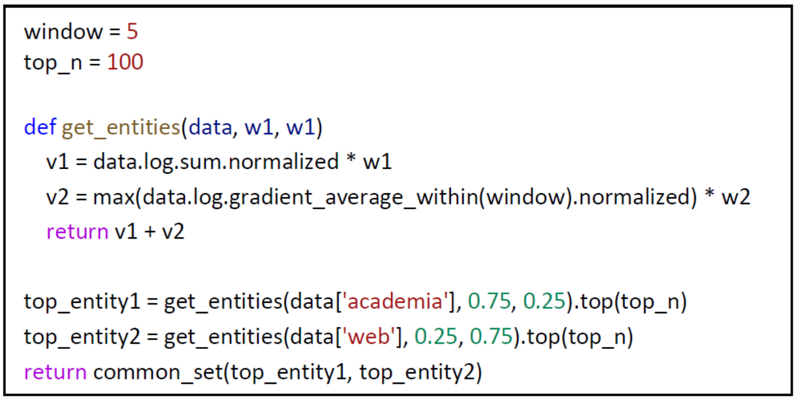
Figure 4(c). For *HSI* in *academia*.

Figure 4(d). For HSI in web

Figure 4. Trend curves for technology-related entities for two technologies of interest in two different document collections.

3.3. Identifying Technology–Relevant Entities

The final process is the identification of technology-relevant entities using the trend curves extracted from the previous section. The candidates are filtered by the combined value of two features: *total_sum* representing the highest trend curves for *academia* and *moving_gradient* representing the highest growth rate at a given interval for *web*. *Total_sum* is calculated as the normalized sum of the frequencies used in the trend curve. *Moving_gradient* is calculated as the maximum normalized average gradient, where the average gradient is calculated for each timeslot using the set time window, which is set to five in this experiment; time windows over 2000 ~ 2019 are reduced to the limit of the extracted data to deal with over/underflow problems. The frequency values vary greatly from 0 to more than 2.0e^9; therefore logscale values are used to reduce the differences between them. The algorithm uses the weighted sum of both *total_sum* and *moving_gradient* to identify top *n* = 100 entities from both *academia* and *web*. Datasets show distinctive differences in their patterns as shown in Figure 4; *academia* shows plateaued curves while *web* shows exponentially growing curves. More weight is given to the feature for the dataset it’s more relevant to; *total_sum* = 0.75 and *moving_gradient* = 0.25 for *academia* and the reverse for *web* as shown in Figure 5.



```
window = 5
top_n = 100

def get_entities(data, w1, w1)
   v1 = data.log.sum.normalized * w1
   v2 = max(data.log.gradient_average_within(window).normalized) * w2
   return v1 + v2

top_entity1 = get_entities(data['academia'], 0.75, 0.25).top(top_n)
top_entity2 = get_entities(data['web'], 0.25, 0.75).top(top_n)
return common_set(top_entity1, top_entity2)
```

Figure 5. Pseudocode for identifying technology-relevant entities.

Technology-relevant entities are identified from the common denominator between the two resulting sets to allow remediation of disadvantages from individual forecasts at the expense of individual advantages [8]. Incorporating both *academia* and *web*, each representing the earliest and the latest predictive time windows, results in a set of relevant entities related to the technology of interest of both the academic interests and public recognitions. Only the entities in the final ranked list from two datasets are selected as the technology-relevant entities, allowing a different number of entities to be found for each technology.

40 for *UAV* and 39 for HIS appeared in both the datasets and were deemed as the technology-relevant entities as shown in Table 1 and Table 2. The score used in this stage is not used for evaluation, hence the entities are not ranked and listed in alphabetical order. They include a range of entities, from high-level domain entities such as *physics* and *data mining* to technology-specific topics such as *ultrasound* and *privacy*, to even seemingly unrelated terms such as *History of the Internet* and *CITES*. The found entities are manually inspected to discern the false positives to calculate the precision of the proposed method at identifying the relevant technologies.

Table 1. List of 40 Technology-Relevant Entities for *UAV*

3D modeling	Core concern	Open architecture	Ranging
Acoustic location	Data mining	Open source	Real time location system
Actuator	Digital identity	Open source hardware	Structured Analysis
Architecture description language	Environment minister	Paper plane	Surveillance
CITES	Integration platform	Privacy	System design
Configuration design	Library (computing)	Privacy by design	System in package
Conservation law	Model aircraft	Privacy laws of the United States	System of record
Continuous integration	Model engine	Privacy policy	System on a chip
Control line	Model ship	Process philosophy	Targeted advertising
Conversation *(disambiguation)*	Modular design	Radio navigation	Vocational education

Table 2. List of 39 Technology-Relevant Entities for *HSI*

Acoustics	Digital divide	Page table
Base address	Digital electronics	Physical symbol system
Black box	Digital recording	Physics
Candidate key	Digital video	Remote sensing
Channel (communications)	Grid computing	Search data structure
Comparison of network diagram software	History of the Internet	Shift register
Computer architecture	Information Age	Simulator
CPU design	Information system	Software diversity
Data (computing)	Internet forum	State machine
Data hierarchy	Machine vision	Ultrasound
Data mining	Memory address register	Value (computer science)
Data processing	Memory model (programming)	Web service
Digital control	Memory protection	Wireless sensor network

4. RESULTS AND DISCUSSIONS

The manual analysis showed that 35 out of the 40 entities found for UAV are relevant, resulting in a precision value of 0.875. Most of the entities fall under six major categories: 1) eight physical components such as actuator, 2) eight vehicle designs methods such as 3D modeling, 3) four navigational features such as radio navigation, 4) four surveying functions such as CITES, 5) six privacy concerns such as digital identity, 6) three other unmanned vehicles, and three uncategorized entities.

The uncategorized technology-relevant entities show hidden connections. Targeted advertising is a marketing strategy optimizing ads to the specific audiences and is mostly employed in the cyberspace, while UAV provides a physical advertising medium in the air capable of following the movement of target audiences over a long period; UAV also allows an easier generation of target-specific contents as well as cheaper aerial accessibility. Data mining is a combination of computer science and statistics seemingly unrelated to UAV, but the increasing number of large-scale datasets such as GIS generated by drones leads to an increased need for data mining to process the raw data.

Figure 6 visualizes Wikipedia articles in a directed graph, where entities are linked by their see-also relationships with diminishing node size with longer distances from the seed. The non-relevant entities acting as a pathway are not colored to distinguish them, while the technology of interest, is colored red to signify the root node in the graph. The tree graph is divided by branches from quadcopter for UAV design and modeling, from human bycatch for navigation and privacy, and micro air vehicle for model and surveillance. The branches do not represent the human categorization; privacy and surveillance branches are far from each other even though the former is the result of the capability of the latter. This suggests that the entities are not necessarily grouped by their graphical structure, nor by their conceptual similarities. CITES, which stands for the Convention on International Trade in Endangered Species of Wild Fauna and Flora, is not related to the surveillance branch, supporting this claim. The graph also explains the existence of seemingly unrelated entities, Conversation (disambiguation) and conservation law; both are connected to the conservation node suggesting that while the former is included as a precaution for mistaking it for conversation, while the latter represents its use in the physics domain.

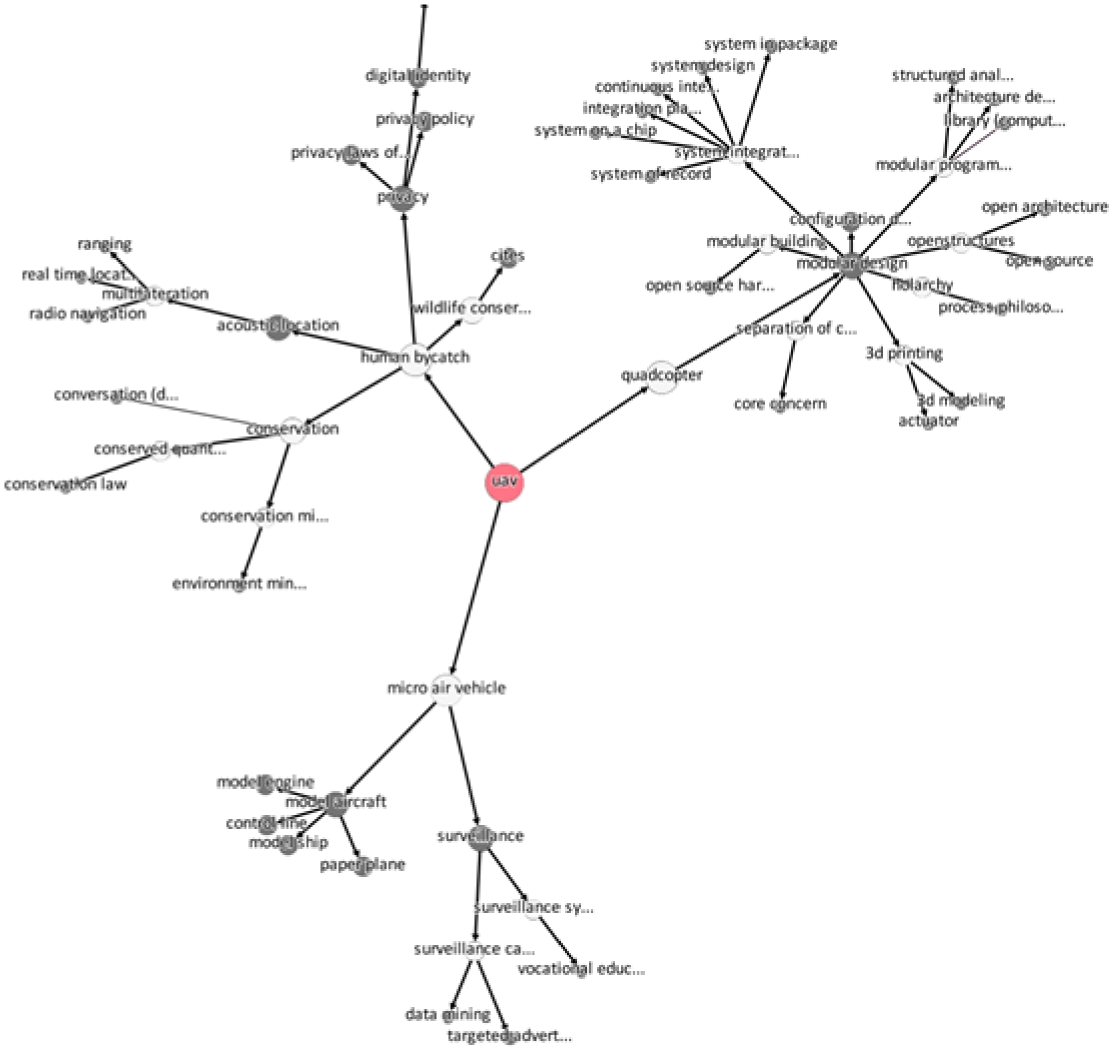


Figure 6. Visualization of paths to the technology-relative entities for UAV.

The manual analysis for HSI resulted in a much lower accuracy of 0.5385, showing only 21 out of the 39 entities as relevant. The majority of the entities are about the actual process of HSI, with eleven related to the data acquisition and preprocessing such as digital video and simulators and eight related to the technical and computational methods used during the process such as digital divide and shift register. Two of the remaining entities are acoustics and ultrasound related to the sound.

The differences in the accuracy can be explained by the skewness of their propagation patterns. Figure 6 for UAV shows a more balanced entity propagation – design perspective, conservation/privacy perspective, and use of micro-size vehicles. On the other hand, the tree graph in Figure 7 for HSI is more skewed towards data (computing) which has a high connection with other entities having less relationship with HSI; 17 out of the 18 unrelated entities were identified from its branch. This shows the danger of utilizing entities with too broad spectrums, where the innate connection to the technology of interest is lost, leading to highly unrelated entities such as history of the internet or physics.

Data mining is more closely related to HSI not only in the graph but also in context, as it is the data analysis technique. More layers are used compared to the related multispectral imaging, increasing the necessity of data mining techniques. Two sound-related entities connected to the root node through sonoluminescence seem unrelated, but acoustics and ultrasound are connected to HSI as they are the non-invasive remote sensing approaches sharing the same goal of collecting information without making physical contact. None of the entities directly connected to the root node in Figure 6 and Figure 7 were identified as technology-relevant entities, which, although not necessarily by design, nonetheless validates the ability of the method to identify remotely-connected entities.

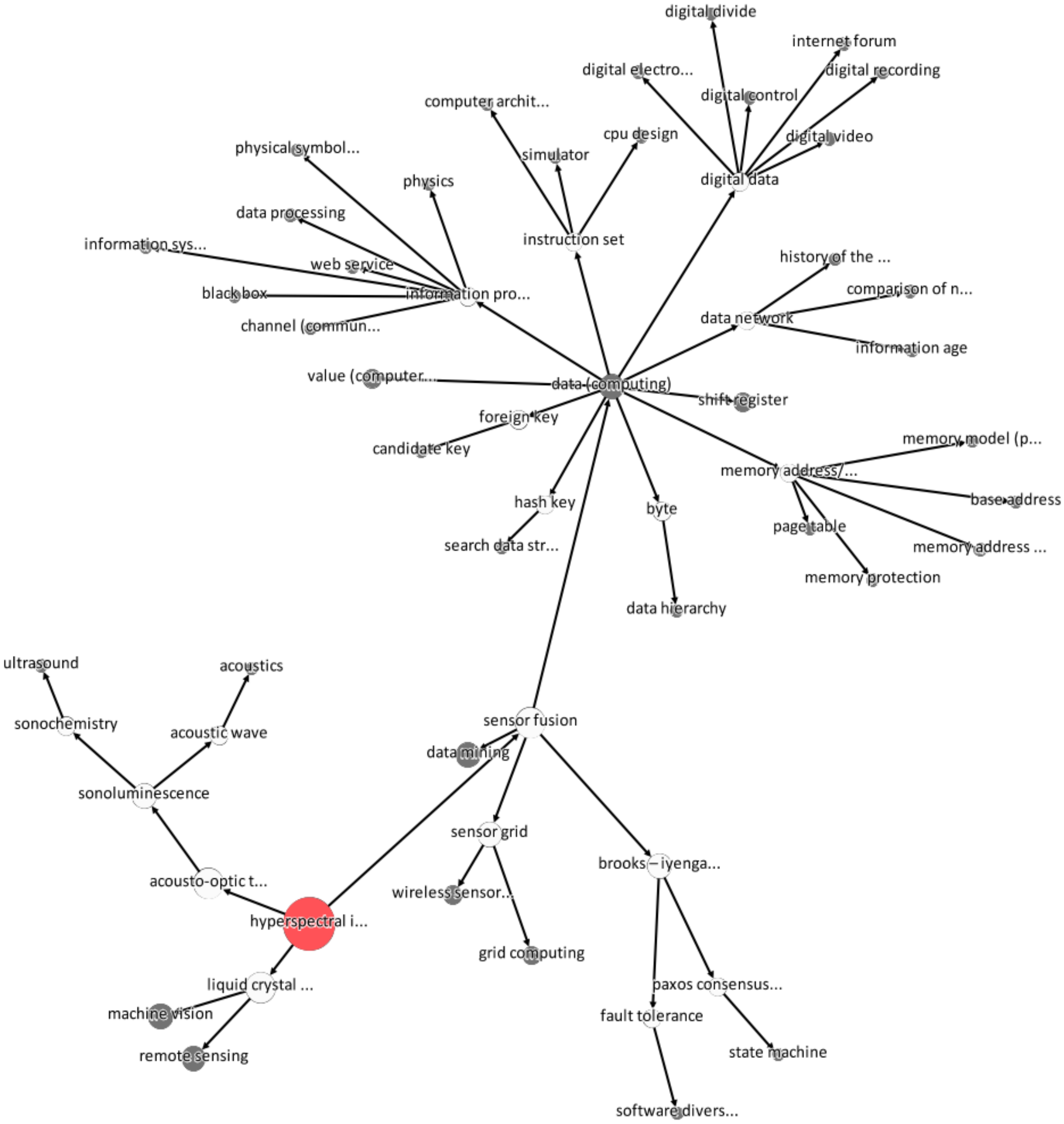


Figure 7. Visualization of paths to the technology-relative entities for HSI.

5. CONCLUSION

The authors propose a method of identifying any type of entities related to a given technology based on their trend curves. The results showed that the entities with recursive relationships in Wikipedia have connections to the target technology not directly observed by either of their encyclopedic descriptions. Case studies revealed that the proposed method can identify entities related to the given technology with hidden relationships. This discovery suggests that the tacit relationships between semantically and syntactically distant technologies can be captured automatically from existing dataset. This opens a path of technology forecasting utilizing the growth of other relevant technologies.

One of the issues for the proposed method is the computational delay when generating the trend curves. The computational complexity is low for trend extraction with *O(nt)* where *n* is the number of trends and *t* is the number of years analyzed. The computation time suffers mostly from the Google search engine API restrictions; the number of query requests is limited to one per 100 seconds. With 20 years to analyze in two different datasets, trend curves for technology-relevant entities can be extracted in over 55.5 hours on a standard computer. Another issue that has an influence on the result is that the relatedness between technology and entities is defined as a binary, treating all related entities equally. Graphical and semantic similarities between them are omitted in the proposed method, rendering it hard to distinguish how related an entity is to the target technology, thus resulting in poor precision for *HSI* due to the entities related to *data (computing)* polluting the entity pool. Future work includes the combination of trend curves with graphical and semantic similarities. Incorporating graphical similarities would allow the method to selectively filter for specific degree of similarities, while implementation of entity filtering with textual similarities would reduce the number of times necessary to generate trend curves, speeding up the method while providing semantic similarity measures between the terms. The future works would also include experimenting on a known case of technology impacted by seemingly unrelated entities to evaluate whether the proposed method can detect such entities beforehand.

ACKNOWLEDGMENTS

This work was supported by the Deep Neural Networks for Forensic and Industry Classifications funded by the Industry Advisory Board (IAB) of an NSF Industry–University Cooperative Research Center (IUCRC) under Grant No. DFII-1903-USA.

REFERENCES

[1] A. Segev, S. Jung, and S. Choi, “Analysis of Technology Trends Based on Diverse Data Sources,” IEEE Trans. Serv. Comput., vol. 2015 Vol.8, no. 06, pp. 903–915, Dec. 2015, doi: 10.1109/TSC.2014.2338855.

[2] P. Durance and M. Godet, “Scenario building: Uses and abuses,” Technol. Forecast. Soc. Change, vol. 77, no. 9, pp. 1488–1492, Nov. 2010, doi: 10.1016/j.techfore.2010.06.007.

[3] K. C. Green and J. S. Armstrong, “Structured analogies for forecasting,” Int. J. Forecast., vol. 23, no. 3, pp. 365–376, Jul. 2007, doi: 10.1016/j.ijforecast.2007.05.005.

[4] N. Dalkey and O. Helmer, “An Experimental Application of the DELPHI Method to the Use of Experts,” Manag. Sci., vol. 9, no. 3, pp. 458–467, Apr. 1963, doi: 10.1287/mnsc.9.3.458.

[5] S. Li, E. Garces, and T. Daim, “Technology forecasting by analogy-based on social network analysis: The case of autonomous vehicles,” Technol. Forecast. Soc. Change, vol. 148, p. 119731, Nov. 2019, doi: 10.1016/j.techfore.2019.119731.

[6] F. Zwicky, “The Morphological Approach to Discovery, Invention, Research and Construction,” in New Methods of Thought and Procedure, Berlin, Heidelberg, 1967, pp. 273–297, doi: 10.1007/978-3-642-87617-2_14.

[7] J. P. Martino, Technological forecasting for decision making, 3rd ed. New York: McGraw-Hill, 1993.

[8] J. S. Armstrong, “Forecasting by Extrapolation: Conclusions from 25 Years of Research,” Inf. J. Appl. Anal., vol. 14, no. 6, pp. 52–66, Dec. 1984, doi: 10.1287/inte.14.6.52.

[9] D. Henton and K. Held, “The dynamics of Silicon Valley: Creative destruction and the evolution of the innovation habitat,” Soc. Sci. Inf., vol. 52, no. 4, pp. 539–557, Dec. 2013, doi: 10.1177/0539018413497542.

[10] O. Dedehayir and M. Steinert, “The hype cycle model: A review and future directions,” Technol. Forecast. Soc. Change, vol. 108, pp. 28–41, Jul. 2016, doi: 10.1016/j.techfore.2016.04.005.

[11] J. S. Armstrong and F. Collopy, “Error measures for generalizing about forecasting methods: Empirical comparisons,” Int. J. Forecast., vol. 8, no. 1, pp. 69–80, Jun. 1992, doi: 10.1016/0169-2070(92)90008-W.

[12] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in Proceedings of the 23rd International Conference on Machine Learning, New York, NY, USA, 2006, pp. 113–120, doi: 10.1145/1143844.1143859.

[13] N. Takeda, Y. Seki, M. Morishita, and Y. Inagaki, “Evolution of Information Needs based on Life Event Experiences with Topic Transition,” in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, Aug. 2017, pp. 1009–1012, doi: 10.1145/3077136.3080703.
[14] Y. Cho and T. Daim, “OLED TV technology forecasting using technology mining and the Fisher-Pry diffusion model,” Foresight, vol. 18, no. 2, pp. 117–137, Jan. 2016, doi: 10.1108/FS-08-2015-0043.
[15] A. Segev, “Circular context-based semantic matching to identify web service composition,” in Proceedings of the 2008 international workshop on Context enabled source and service selection, integration and adaptation organized with the 17th International World Wide Web Conference (WWW 2008) - CSSSIA ’08, Beijing, China, 2008, pp. 1–5, doi: 10.1145/1361482.1361489.
[16] M. Gusenbauer, “Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases,” Scientometrics, vol. 118, no. 1, pp. 177–214, Jan. 2019, doi: 10.1007/s11192-018-2958-5.
[17] M. de Kunder, “The size of the World Wide Web (The Internet),” Mar. 07, 2020. https://www.worldwidewebsize.com/.
[18] A. Segev, C. Jung, and S. Jung, “Analysis of Technology Trends Based on Big Data,” in 2013 IEEE International Congress on Big Data (BigData Congress), Jun. 2013, pp. 419–420, doi: 10.1109/BigData.Congress.2013.65.

AUTHORS

Sukhwan Jung is a Postdoc in the Department of Computer Science, University of South Alabama. He earned his Ph.D. in Knowledge Service Engineering at KAIST in 2020

Rachana Reddy Kandadi is a Master’s degree student in the Department of Computer Science, University of South Alabama. She earned her undergraduate degree from JNTUH in Computer Science and Engineering.

Rituparna Datta is a Computer Research Associate-II in the Department of Computer Science, University of South Alabama. Prior to that, he was an Operations Research Scientist in Boeing Research & Technology (BR&T), BOEING, Bangalore

Ryan Benton is an Assistant Professor in the Department of Computer Science at the University of South Alabama. His research interests lies in the fields of data mining and machine learning, with a current focus upon advanced pattern mining methods, novel graph mining algorithms, and applied applications

Segev Aviv is an Associate Professor in the Department of Computer Science at the University of South Alabama. His research interest is looking for the DNA of knowledge, an underlying structure common to all knowledge, through analysis of knowledge models in natural sciences, knowledge processing in natural and artificial neural networks, and knowledge mapping between different knowledge domains.