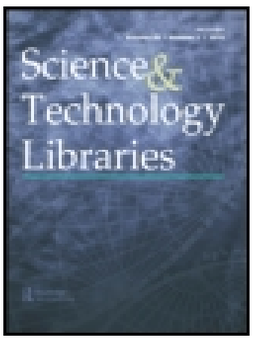


Science & Technology Libraries

ISSN: 0194-262X (Print) 1541-1109 (Online) Journal homepage: https://www.tandfonline.com/loi/wstl20

What Do Websites Say about Internet of Things Challenges? A Text Mining Approach

Mehdi Dadkhah, Mohammad Lagzian, Fariborz Rahimnia & Khalil Kimiafar

To cite this article: Mehdi Dadkhah, Mohammad Lagzian, Fariborz Rahimnia & Khalil Kimiafar (2020): What Do Websites Say about Internet of Things Challenges? A Text Mining Approach, Science & Technology Libraries, DOI: 10.1080/0194262X.2020.1715320

To link to this article: https://doi.org/10.1080/0194262X.2020.1715320

Published online: 24 Feb 2020.

SCIENCE & TECHNOLOGY LIBRARIES
https://doi.org/10.1080/0194262X.2020.1715320

Routledge
Taylor & Francis Group

What Do Websites Say about Internet of Things Challenges? A Text Mining Approach

Mehdi Dadkhah[a], Mohammad Lagzian[a], Fariborz Rahimnia[a], and Khalil Kimiafar[b]

[a]Department of Management, Faculty of Economics and Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran; [b]Department of Medical Records and Health Information Technology, School of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran

ABSTRACT

Internet of Things (IoT) is a popular technological paradigm offering several promised advantages. IoT has been facing different challenges discussed by various researchers. The available literature on IoT challenges is mostly based on experimental studies or reviews. The analysis of webpages as a valuable database can be a good idea to provide an insight into IoT challenges. Thus, this paper was aimed at providing an overall view of webpages discussing IoT challenges. In this regard, the text mining approach was implemented to analyze the related webpages. The obtained results revealed that challenges relate to the security of IoT devices, connectivity, data processing, and storage have been the most prevalent topics in the selected webpages.

KEYWORDS

Internet of Things challenges; IoT adoption; smart devices; Internet of Things; latent Dirichlet Allocation

Introduction

Nowadays, IoT is one of the noteworthy technology paradigms. It has gained popularity in research communities and industries in different sectors (Atzori, Iera, and Morabito 2010; Huang et al. 2017). The early version of the internet contained HTML webpages linked together that were at the top of the internet architecture. After a while, such HTML webpages evolved in the form of web 2.0 that allowed for communication between the end users. Web 2.0 comprises of technologies such as wiki, social networks, blogs, etc. Researchers have put more effort into developing web 2.0 and providing a semantic web. This is where the web pages are understandable by machines. This version of the web is referred to as the semantic web or web 3.0. During the evolution of web technologies, RFID, near field communication, and sensor technologies have evolved as well. The evolutionary convergence of both technologies has led to a new vision facilitating machines communicating with each other via the internet. Such a vision provides a new paradigm entitled "Internet of Things"(Whitmore, Agarwal, and Li 2015). IoT can be defined as the global network of interconnected heterogeneous things (such as sensing and actuating devices) which are

CONTACT Mohammad Lagzian m-lagzian@um.ac.ir Department of Management, Faculty of Economics and Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

connected to the internet and aims to enhance different applications in various sectors (such as logistics, healthcare, manufacturing, etc.) (Chang, Srirama, and Buyya 2019; Gubbi et al. 2013). IoT will have a significant impact on both the internet and the economy. It is anticipated that there would be about 100 billion connected IoT devices and a total economic impact of between $3.9 trillion and $11.1 trillion per year by 2025 (Manyika et al. 2015; Rose, Eldridge, and Chapin 2015). IoT will decrease costs and improve efficiency (Lee and Lee 2015). Furthermore, it is a key enabler for future industrial revolutions (Trappey et al. 2017). There are other papers discussing the advantages and impact of IoT that are outside of the scope of this research.

Although IoT provides some promising benefits, there are challenges limiting the IoT advantages that are addressed by researchers in order to be resolved. There are different studies in the literature investigating IoT challenges such as data management, privacy, security, interoperability, scalability, etc. (Breivold and Kristian 2015; Lee and Lee 2015; Peter and Richard Watson 2017; Tsai, Lai, and Vasilakos 2014). Some researchers have focused merely on a certain type of IoT challenges such as security (Sfar et al. 2018), privacy (Ziegeldorf, Morchon, and Wehrle 2014), ethical challenges (Popescul and Georgescu 2014), etc. Mohammadzadeh et al. (2018) used the Fuzzy Analytic Network Process (FANP) to prioritize the different challenges by considering Iran as the case study. They grouped the different IoT challenges into five categories, including technological, privacy and security, business, legal and regulatory and cultural challenges. Their research indicated that technological challenges and the challenges related to privacy and security have priority over others. In the literature, there are other researches into IoT challenges that reflect the authors' concerns.

Although there are valuable published papers addressing IoT challenges, websites available on the internet can be an appreciated source for analyzing IoT challenges as well. Websites are prized sources of knowledge, since there are web mining techniques to inspect web and extract valuable knowledge from them (Srivastava, Desikan, and Kumar 2005). This attaches significance to web content as a source of knowledge. Bloggers, freelancers, experts, journalists, and many other people share valuable information via websites. In other words, in websites, non-academic people also share information. Although there are some papers reviewing available studies on IoT challenges (Mohammadzadeh et al. 2018; Peter and Richard Watson 2017) and highlighting dominant challenges, no analysis has been performed of websites discussing such IoT challenges, which necessitates an analysis of related webpages to provide an insight into them. This paper aims to answer the below research questions:

Which categories of IoT challenges have been widely discussed in webpages? Which category(s) is more prevalent? Which emotion is behind webpages content in this regard?

Methodology of the study

This research aims to provide an overall view of the webpages that discuss IoT challenges by identifying the categories of IoT challenges that have been discussed in the webpages, their prevalence, and the emotions behind the content. Text mining is appropriate in this regard because it provides the possibility of automatically analyzing large sets of texts and there are also algorithms that fit our research questions. The objective of text mining is to identify relevant patterns out of the data to help decision-making (Aggarwal and Zhai 2012). There are different text mining algorithms and, for the purpose of this research, the topic modeling algorithm seems practical to identify the different categories of IoT challenges. In this model, the key assumption is that any document may have the probability of belonging to multiple topics. In addition, each topic is associated with the probability of involving different words (Aggarwal and Zhai 2012; Bach et al. 2019). This research uses Latent Dirichlet Allocation (LDA) to find topics in the webpages. The dominant categories of the IoT challenges will appear in these topics. LDA is a generative probabilistic model for discrete data such as textual data that aims to provide a short description drawn from the data to enable the processing of a large amount of data while holding up any essential statistical relationships (Blei, Ng, and Jordan 2003). LDA has been widely used in the literature for topic extraction from textual data (Garcia-Rudolph et al. 2019; Moro, Cortez, and Rita 2015; Tian, Revelle, and Poshyvanyk 2009; Tirunillai and Tellis 2014; Wang et al. 2016). The graphic representation of LDA is shown in the following Figure 1.

In this figure, matrix φ indicates the distribution of the T topics over W words from a Dirichlet parameterized by β. In addition, matrix θ shows the distribution of D documents over T topics from a Dirichlet parameterized by α. In fact, a topic z is drawn from the related document's topic distribution θd when word w is generated in document d. However, in such conditions, word w is drawn from the selected topic's word distribution φz. To extract a topic by taking advantage of LDA, we need to estimate φ and θ (Momtazi and Naumann 2013). Various

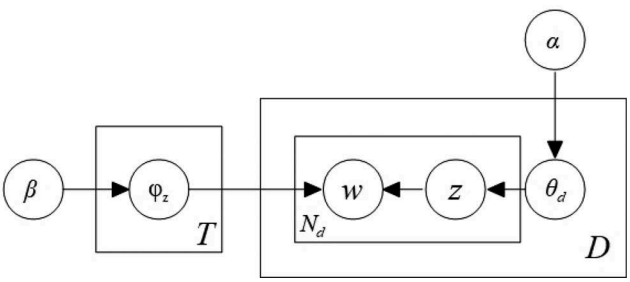


Figure 1. Graphical representation of LDA (from (Blei, Ng, and Jordan 2003; Momtazi and Naumann 2013) with edition).

algorithms have been presented in this regard including Gibbs, the sampling algorithm that is an efficient one for this purpose (Griffiths and Steyvers 2004; Momtazi and Naumann 2013). The input parameters that should be inserted by the user are the number of topics (T) and the values of hyperparameters (α and β). Where α and β are topic smoothing and word/term smoothing parameters, respectively. The larger the values of these parameters over one, the more even the distributions while the more smaller than one the results, the more concentrated the distributions over topics or words (Bastani, Namavari, and Shaffer 2019).

Regarding the text mining process, the related data should be firstly selected, cleaned, and represented, followed by the implementation of the LDA algorithm to it (Rajman and Vesely 2004). The term "Internet of Things challenges" was searched for in the Google search engine on 30th June 2019 and the first 220 results were assessed to select the relevant data. An overview on each webpage has been done to identify it is relevant or not. Academic papers, PDF files, and websites not focusing on IoT challenges were excluded. Finally, 86 webpages were identified to pertain to IoT challenges and they were stored as text files. However, these text files were not ready for data analysis since they contained irrelevant data such as links and words related to different parts of the sites, namely advertisements, menus, categories, etc. These text files were removed as much as possible to include only the text related to IoT challenges (all of the textual content of the related post).

The R tool and related packages were assigned to data analysis in this research (Benoit and Obeng 2019; Csardi and Nepusz 2006; Dowle and Srinivasan 2019; Feinerer and Hornik 2018; Feinerer, Hornik, and Meyer 2008; Fellows 2018b; Grün and Hornik 2011; Hornik, Meyer, and Buchta 2019; Jockers 2015; Neuwirth n.d.; Nikita 2017; R Core Team 2018; Sievert and Shirley 2014, 2015; Silge and Robinson 2016; Wickham 2016, 2017; Wickham et al. 2019; Wickham and Henry 2019). Some R tutorials and their codes along with some editions were used to conduct the related analysis ("An Overview of Text Mining Visualisations Possibilities with R on the CETA Trade Agreement" n.d.; "Read Text Files with Readtext()" n.d.; "Text Mining Example Codes (Tweets)" n.d.; Csardi n.d.; Feinerer 2018; Fellows 2018a; Fetzer 2017; Mair 2018; Neuwirth n.d.; Ognyanova 2015; Oliver 2018; Randy 2018; Sievert and Shirley 2014; Xiao and Miaozhu 2019; Zhao 2012; n.d.). All 86 text files were imported into R and cleaned subsequently. The cleaning process includes converting all letters to lower case and removing punctuation, stop words, number, extra whitespace, and stemming words, preparing the data for future analysis. The analysis identified the most frequent words in the data and their co-occurrence network. Furthermore, it recognized relevant topics, their prevalence, and sentiment analysis. To detect the emotion behind texts in webpages, a sentiment analysis of the data was carried out as well. When human readers study a text, they can understand positive or negative concept behind the text and identify related emotions, which can also be

conducted automatically by machine through text mining that is the subject of sentiment analysis (Liu 2012; Silge and Robinson 2017). There are some semantic lexicons in text mining to realize words and meanings behind them (Silge and Robinson 2017).

Results and discussion

Based on the analysis of the data, Figure 2 shows the word clouds of the related webpages where the size of each word is related to its frequency in the overall data. Figure 3 illustrates the co-occurrence network of the top 11 words. According to these figures, *IoT, device, secure, challenge, internet, things*, etc., were found to be the most frequent words remembering the IoT challenges correctly. This ensured the validity of data collection and the cleaning process since no irrelevant words were observed among the most frequent words. In other words, the word cloud of the data implied its relevance to the IoT challenges. As the stemming process has been used to incorporate the root of the words the words may be incomplete in the figures (some figures support infinite zoom).

LDA was applied to detect the available topics in the documents. By using the library "ldatuning" in R, a total of 11 topics were deemed to be suitable for the concerned data. By implementing LDA to the data, the topic-related keywords were identified. Figure 4 depicts the word clouds related to each topic. The size of each keyword is proportional to its probability in the corresponding topic.

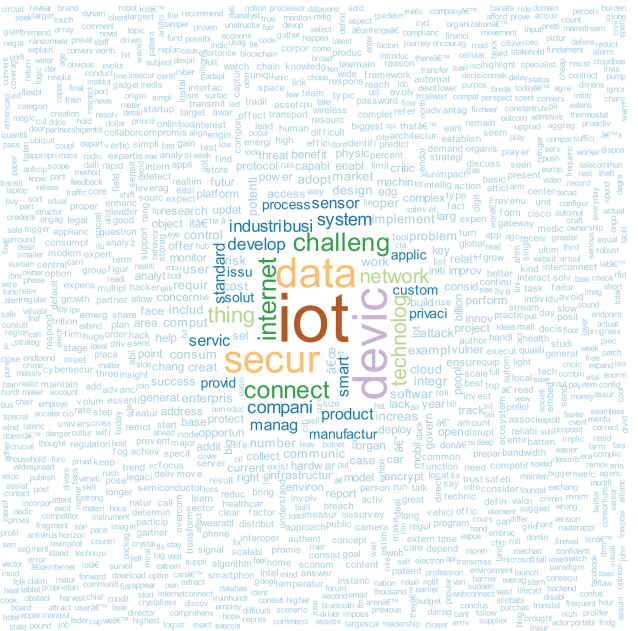


Figure 2. Word cloud of webpages.

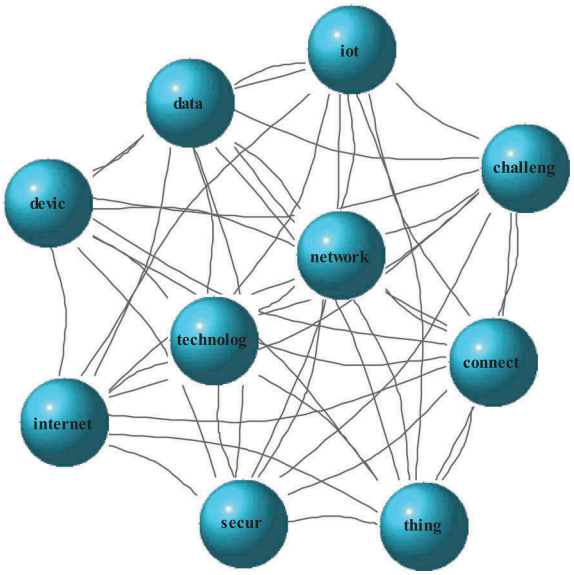


Figure 3. Co-occurrence network of top 11 words.

The LDA algorithm provided the terms associated with each topic and, in turn, the user could label them manually (Bastani, Namavari, and Shaffer 2019; Chang et al. 2009). To label each topic, pertinent terms and webpages were involved wherein a certain topic is more dominant (see Figure 4). Hence, the topic label may be inaccurate in Figure 4. However, the keyword identified would help to gain an overall view of the websites discussing IoT challenges. Topic 1 included IoT security challenges such as device security, data security, authentication, malware, etc. Webpages with this topic investigated IoT security challenges and possible solution (e.g., Katz 2019; Nwazor 2018). Topic 2 focused on the connectivity issues for IoT in large-scale deployment, namely network speed, the interconnection of heterogenous networks, firewalls, etc. (Hanson 2014; Roussey 2016). Processing and storing a large amount of data generated by IoT is another challenge engaged in topic 3 (D'mello 2019; Evans 2018). Topic 4 arose due to the discussion on the webpages of the effectiveness of IoT implementation and related benefits (Delaney and Levy 2017). A lack of a unique standard has led different companies to use various solutions. This entails IoT adoption with challenges. This is as different technologies may be incompatible, wherein developing end-to-end security solutions is burdensome (topic 5) (Bauer, Burkacky, and Knochenhauer 2017). Topic 6 discussed the concerns, challenges, and problems related to IoT adoption (Buntz 2016). Topic 7 encompassed the developmental and implementation challenges related to IoT devices (Rowe n.d.). Topics 8 and 9 appeared due to general terms in the IoT literature. They pertained to smart objects and other popular keywords in the IoT domain. One or two webpages were cited for each topic as an example in the above-mentioned discussion.

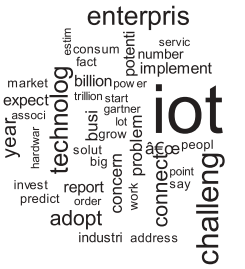
Topic 1: Security challenge of IoT devices

Topic 2: Connectivity challenge of IoT

Topic 3: Challenge of Data processing and storage

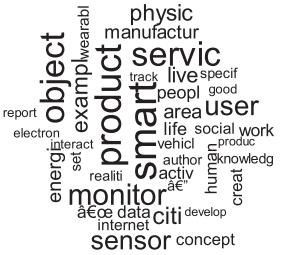
Topic 4: Success of IoT implementation and related benefits

Topic 5: Security and standardaztion challenge of IoT

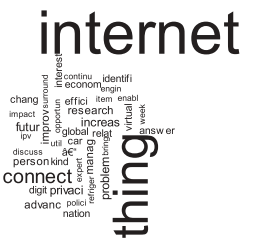


Topic 6: Challenges related to adoption of IoT

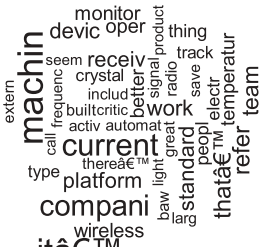
Topic 7: Development/Implementation challenge of IoT



Topic 8: Smart objects in IoT



Topic 9: Internet of Things



Topic 10: Authors did not label this group of words.

Figure 4. Topic-centered Word cloud.

In addition, LDAvis (Sievert and Shirley 2014) was implemented to visualize the result of the LDA analysis. Figure 5 illustrates the topics along with their distances. The prevalence of the corresponding topic was designated by

Figure 5. Topics visualized using LDAvis.

the area of each circle. The inter-distance of topics (similarity between topics) was represented by two axes. As can be observed in Figure 4, there were discussions about IoT challenges, including security, connectivity, data storage and processing, standardization, adoption, and implementation/development challenges, in the selected webpages. Topic 1, i.e., the security challenge of IoT devices, was the most common topic in the webpages. After this topic, topic 2 (connectivity challenge of IoT) and topic 3 (the challenge of Data processing and storage) ranked the most prevalent. The reason for the popularity of these topics might be attributed to their significance as barriers to IoT. The proposed search process for identifying the related webpages was performed in the English language medium without any limitation as to the context. However, the obtained result agrees with that obtained by Mohammadzadeh et al. (Mohammadzadeh et al. 2018). Their study showed that technology, privacy, and security challenges have a high priority over others. Their category for technological challenges consisted of architecture and design, device heterogeneity, addressing, ubiquitous data management, hardware construction, and fault tolerance. Accordingly, our findings reveal that security and technological challenges were the most commonplace topics in the webpages, which might be due to the higher significance of these challenges. There have been other research studies referring to security challenges as one of the major IoT issues (Aldowah,

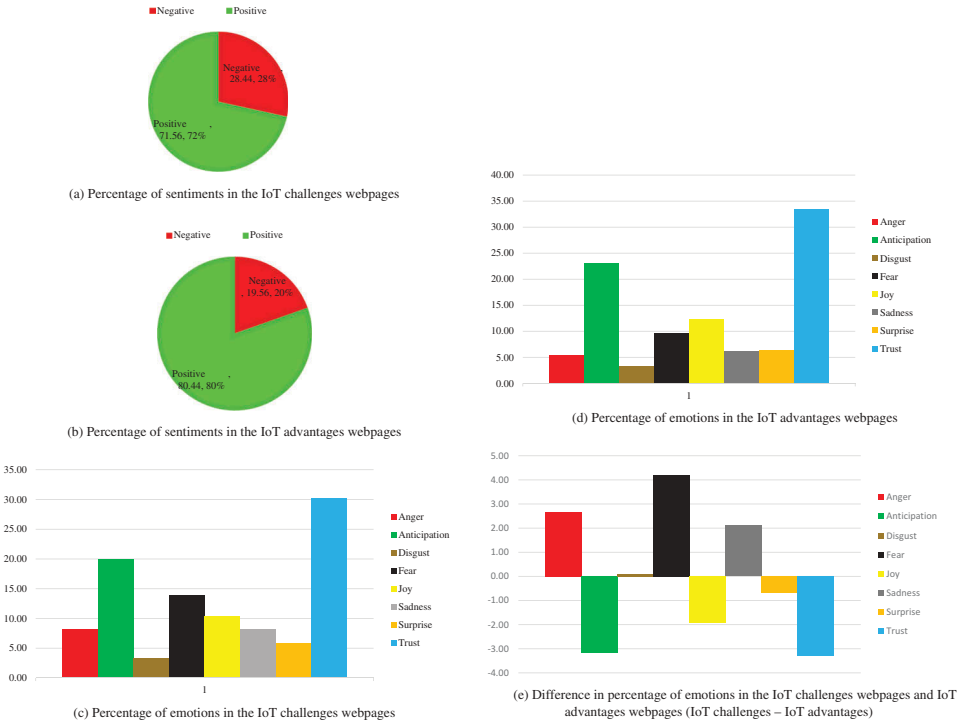
Rehman, and Umar 2018; Sicari et al. 2015). Zubiaga et al. analyzed tweets which discuss IoT, they concluded that security of IoT is a main concern (Zubiaga et al. 2018).

Moreover, the categories of IoT challenges specified by analyzing webpages have been examined in literature. For example, security (Aldowah, Rehman, and Umar 2018; Sicari et al. 2015), connectivity (Mumtaz et al. 2017; Samuel 2016), data processing, and storage (Lee and Lee 2015; Ma, Wang, and Chu 2013), standardization (Rose, Eldridge, and Chapin 2015), etc., have been broadly investigated in the literature. Accordingly, webpages do not typically face new kinds of challenges and they share some similar and already-available categories of challenges that have been presented in literature. It is worth noting that this study aimed at generally discussing IoT challenges, and not doing so in detail.

Some challenges have been highlighted in the analysis, such as security, connectivity, and data processing, storage challenges, etc., more than others. Such challenges are dominant concerns in the realm of IoT challenges on webpages. Future research should pay special attention to these challenges to provide solutions for them. In practice, an individual intending to use IoT needs to have an awareness of these challenges. These are the key barriers to viable IoT implementation. Even though this paper has offered an overview of the webpages discussing IoT challenges, it presents some new questions; 'What are the most prevalent challenges in the literature?' 'Are the results similar to the findings of this paper?' and 'Can the importance of IoT challenges be varied between different contexts and regions?' Future research can answer these questions.

To perform sentiment analysis, available tutorials (Shen 2017; Silge and Robinson 2017) and the NRC lexicon (S. M. Mohammad and Turney 2010, 2012) were used. The NRC lexicon provides two sentiments, positive or negative, and eight emotions, including anger, fear, anticipation, trust, surprise, sadness, joy, and disgust (Mohammad and Turney 2010, 2012). Figure 6(a,c) show the results of the sentiment analysis of the 86 selected webpages about IoT challenges.

Based on the literature (Mohammad 2011a; Garcia-Rudolph et al. 2019; Mohammad and Yang 2011) and after consulting with experts, it was concluded that we could gain an insight into the results of the sentiment analysis when comparer two dataset. Therefore, it was preferred to perform sentiment analysis on the IoT-related data disregarding the IoT challenges, followed by making a comparison between the results. The reason for this is that the number of positive or negative words is not sufficient enough criterion to assess people's opinion. By comparing two different datasets in the same domain, more insight can be provided. In this regard, it was decided to create a dataset from the websites addressing the IoT benefits and advantages. The below query was searched in the Google search engine (search date 29th July 2019):



(a) Percentage of sentiments in the IoT challenges webpages

(b) Percentage of sentiments in the IoT advantages webpages

(c) Percentage of emotions in the IoT challenges webpages

(d) Percentage of emotions in the IoT advantages webpages

(e) Difference in percentage of emotions in the IoT challenges webpages and IoT advantages webpages (IoT challenges – IoT advantages)

Figure 6. Sentiment analysis of webpages about IoT challenges and benefits (Percentage of an emotion has been calculated by considering total occurrence of an emotion in a dataset divided by total occurrence of all emotions on that dataset).

"IoT advantages" OR "IoT potentials" OR "IoT Benefits" OR "IoT profit" OR "IoT promise" OR "Internet of things advantage" OR "Internet of things benefit"

Then, 86 webpages discussing IoT advantages and benefits were selected. Websites which discussed IoT advantages in addition to its challenges were excluded when possible. The sentiment analysis of this dataset was also performed (Figure 6(b,d)), and the obtained results were then compared with the results of the sentiment analysis on IoT challenges dataset. Accordingly, it was possible to compare the results of the sentiment analysis of IoT challenge data with the result of the sentiment analysis of dataset about IoT advantages. The results revealed that positive words in both data were the most dominant, which could be attributed to the discussions about IoT-promised advantages, drivers, IoT success, etc. Negative words ranked second in sentiments, while data pertaining to IoT challenges included more negative words, which could be due to challenges embedded in webpages. In addition, as can be observed in Figure 6, negative emotions such as "fear" were of higher ranks compared to the results of sentiment analysis of IoT advantages dataset. Individuals might fear that such challenges could lead to their IoT project failure. Other emotions could

be compared by considering Figure 6(e). The pattern of colors in the figure is based on the study conducted by Mohammad (2011b).

The sentiment analysis revealed the emotions behind texts in webpages and positive or negative sentiment in this regard. This can help to develop new research in this area, especially for researchers studying technology acceptance. For example, IoT challenges may lead to fear in individuals to use IoT due to the possibility of project failure, health risks, immaturity of IoT devices, etc. Identifying such factors can be vital to IoT technology acceptance. For future information about IoT acceptance, one can refer to cited references (Balaji et al. 2018; Canhoto and Arp 2017; Coughlan et al. 2012). There is a potential for research to understand emotions of people regard to their interaction with IoT or related devices. For example, it is possible to conduct an interview with people to understand reasons which they accept to use or not use IoT devices. Then by doing sentiment analysis, it is possible to understand emotions behind of presented reasons. There is valuable research in the literature which discuss reasons of using or not using IoT devices (Canhoto and Arp 2017; Garg and Kim 2018), but they did not applied sentiment analysis on the interviews. When we are aware with emotions which usually lead to not using of IoT devices, it is possible to plan for addressing reasons of not using with more insights.

Conclusion

This research provided an overall view of the webpages discussing IoT challenges and it has identified the most prevalent challenges present. Based on the results, the challenges related to the security of IoT devices, connectivity and data processing, and storage were the most prevalent. The reason for this may lie in their significance as a barrier to the progression of IoT technology since different webpages mentioned the aforementioned as challenges of IoT. A key contribution of this research was the inspection of websites as the supplementary source of knowledge instead of solely analyzing journal papers. Websites are a valuable knowledge source and they can provide insights for the researchers. Most of the available literature prior to this research focused on academic papers or they briefly referred to the web content concerning IoT challenges. This research has shed a light on the status of the discussion of IoT challenges through websites.

The results of the current research were based on a semi- or fully computer-based analysis of the textual data (natural language) and there might be some errors in comparison to if there had been a human-based analysis of text. In addition, almost all of the most prevalent categories of IoT challenges were identified in research (not all of them and not in detail), which can be assumed to be a limitation of this study.

There are some studies which provide an overview on IoT literature by using statistical methods. (MacDonald and Dressler 2018; Mishra et al. 2016). Using these methods to study IoT challenges can be topic for future research.

Acknowledgments

Dr. Massimo Aria, Associate Professor in Statistics for Social Sciences, PhD in Computational Statistics, Laboratory and Research Group STAD Statistics, Technology, Data Analysis, Department of Economics and Statistics, University of Naples Federico II, Monte S. Angelo, via Cinthia I-80126 Napoli, Italy

Funding

There is no funding resource.

References

Aggarwal, C. C., and C. Zhai. 2012. An introduction to text mining. In *Mining text data*, ed. Charu C. Aggarwal, ChengXiang Zhai, 1–10. Boston, MA: Springer. doi:10.1007/978-1-4614-3223-4.

Aldowah, H., S. U. Rehman, and I. Umar. 2018. Security in Internet of Things: Issues, challenges and solutions. Proceedings of the 3rd International Conference of Reliable Information and Communication Technology 2018 (IRICT 2018), Kuala Lumpur, Malaysia, July 23–24..

An overview of text mining visualisations possibilities with R on the CETA trade agreement. n.d. Accessed July 3, 2019. https://www.bnosac.be/index.php/blog/56-an-overview-of-text-mining-visualisations-possibilities-with-r-on-the-ceta-trade-agreement.

Atzori, L., A. Iera, and G. Morabito. 2010. The Internet of Things: A survey. *Computer Networks* 54 (15):2787–805. doi:10.1016/j.comnet.2010.05.010.

Bach, P., Ž. K. Mirjana, S. Seljan, and L. Turulja. 2019. Text mining for big data analysis in financial sector: A literature review. *Sustainability* 11 (5):1277. doi:10.3390/su11051277.

Balaji, M. S., S. K. Roy, A. Sengupta, and A. Chong. 2018. User acceptance of IoT applications in retail industry. In *The Internet of Things in the Modern Business Environment,*ed. In Lee, 1331–52. USA: IGI Global. doi:10.4018/978-1-5225-2104-4.ch002.

Bastani, K., H. Namavari, and J. Shaffer. 2019. Latent Dirichlet Allocation (LDA) for topic modeling of the CFPB Consumer Complaints. *Expert Systems with Applications* 127 (August):256–71. doi:10.1016/j.eswa.2019.03.001.

Bauer, H., O. Burkacky, and C. Knochenhauer. 2017. Security in the Internet of Things. 2017. https://www.mckinsey.com/industries/semiconductors/our-insights/security-in-the-internet-of-things.

Benoit, K., and A. Obeng. 2019. Readtext: Import and handling for plain and formatted text files. R package version 0.75. https://CRAN.R-project.org/package=readtext.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3 (Jan):993–1022.

Breivold, H. P., and S. Kristian. 2015. Internet of Things for industrial automation–challenges and technical solutions. In *2015 IEEE International Conference on Data Science and Data*

Intensive Systems, 532–39. IEEE, Sydney, NSW, Australia, December 11–13. doi:10.1109/DSDIS.2015.11

Buntz, B. 2016. Top 10 reasons people Aren't Embracing the IoT. 2016. https://www.iotworldtoday.com/2016/04/20/top-10-reasons-people-aren-t-embracing-iot/.

Canhoto, A. I., and S. Arp. 2017. Exploring the factors that support adoption and sustained use of health and fitness wearables. *Journal of Marketing Management* 33 (1–2):32–60. doi:10.1080/0267257X.2016.1234505.

Chang, C., S. N. Srirama, and R. Buyya. 2019. Internet of Things (IoT) and new computing paradigms. In *Fog and Edge Computing: Principles and Paradigms* ed. Rajkumar Buyya, Satish Narayana Srirama, 1–23, Hoboken, NJ: John Wiley & Sons, Inc. doi:10.1109/DSDIS.2015.11

Chang, J., S. Gerrish, C. Wang, B.-G. Jordan L, and D. M. Blei. 2009. Reading tea leaves: How humans interpret topic models. Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 7–10.

Coughlan, T., M. Brown, R. Mortier, R. J. Houghton, M. Goulden, and G. Lawson. 2012. Exploring acceptance and consequences of the Internet of Things in the home. In *Green Computing and Communications (GreenCom), 2012 IEEE International Conference On*, 148–55. Besancon, France: IEEE. doi:10.1109/GreenCom.2012.32

Csardi, G. n.d. R Igraph manual pages. Accessed July 3, 2019. https://igraph.org/r/doc/.

Csardi, G., and T. Nepusz. 2006. The Igraph software package for complex network research. *InterJournal Complex Systems* 1695.

D'mello, A. 2019. The Top 3 challenges of preparing IoT data. 2019. https://www.iot-now.com/2019/02/08/92826-top-3-challenges-preparing-iot-data/.

Delaney, K., and E. Levy. 2017. Internet of Things: Challenges, breakthroughs and best practices. 2017. https://connectedfutures.cisco.com/report/internet-of-things-challenges-breakthroughs-and-best-practices/.

Dowle, M., and A. Srinivasan. 2019. Data table: Extension of `data.Frame`. R package version 1.12.2. https://CRAN.R-project.org/package=data.table

Evans, C. 2018. Internet of Things challenges in storage and data. 2018. https://www.computerweekly.com/news/252450705/Internet-of-things-challenges-in-storage-and-data.

Feinerer, I. 2018. Introduction to thetmpackagetext mining in R. https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf.

Feinerer, I., and K. Hornik. 2018. Tm: Text mining package. R package version 0.7-6. https://CRAN.R-project.org/package=tm.

Feinerer, I., K. Hornik, and D. Meyer. 2008. Text mining infrastructure in R. *Journal of Statistical Software* 25 (5):1–54. doi:10.18637/jss.v025.i05.

Fellows, I. 2018a. Package 'Wordcloud'. https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf.

Fellows, I. 2018b. Wordcloud: Word clouds. R package version 2.6. https://CRAN.R-project.org/package=wordcloud.

Fetzer, T. 2017. Topic models. http://www.trfetzer.com/wp-content/uploads/EC994_Topic-Modeling1.pdf.

Garcia-Rudolph, A., S. Laxe, J. Saurí, and M. B. Guitart. 2019. Stroke survivors on twitter: Sentiment and topic analysis from a gender perspective. *Journal of Medical Internet Research* 21 (8):e14077. doi:10.2196/14077.

Garg, R., and J. Kim. 2018. An exploratory study for understanding reasons of (not-) using Internet of Things. *CHI '18: CHI Conference on Human Factors in Computing Systems Montreal QC Canada April, 2018*. New York, NY: Association for Computing Machinery. ISBN:978-1-4503-5621-3

Griffiths, T. L., and M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences* 101 (suppl 1):5228–35. doi:10.1073/pnas.0307752101.

Grün, B., and K. Hornik. 2011. Topicmodels: An R package for fitting topic models. *Journal of Statistical Software* 40 (13):1–30. doi:10.18637/jss.v040.i13.

Gubbi, J., R. Buyya, S. Marusic, and M. Palaniswami. 2013. Internet of Things (IoT): A Vision, architectural elements, and future directions. *Future Generation Computer Systems* 29 (7):1645–60. doi:10.1016/j.future.2013.01.010.

Hanson, J. 2014. 5 Challenges of Internet of Things Connectivity. 2014. https://community.arm.com/iot/b/blog/posts/5-challenges-of-internet-of-things-connectivity.

Hornik, K., D. Meyer, and C. Buchta. 2019. Slam: Sparse lightweight arrays and matrices. R package version 0.1-45. https://CRAN.R-project.org/package=slam.

Huang, J., Q. Duan, C.-C. Xing, and H. Wang. 2017. Topology control for building scalable energy-efficient Internet of Things. *IEEE Wireless Communications* 24:1. doi:10.1109/MWC.2017.1600193WC.

Jockers, M. L. 2015. Syuzhet: Extract sentiment and plot arcs from text. https://github.com/mjockers/syuzhet.

Katz, H. 2019. IoT cybersecurity challenges and solutions. 2019. https://www.allot.com/blog/iot_cybersecurity_challenges_and_solutions/.

Lee, I., and K. Lee. 2015. The Internet of Things (IoT): Applications, Investments, and Challenges for Enterprises. *Business Horizons* 58 (4):431–40. doi:10.1016/j.bushor.2015.03.008.

Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*. 5. Morgan & Claypool. doi:10.2200/S00416ED1V01Y201204HLT016

Ma, M., P. Wang, and C.-H. Chu. 2013. Data management for Internet of Things: Challenges, approaches and opportunities. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 1144–51. Beijing, China: IEEE.

MacDonald, K. I., and V. Dressler. 2018. Using citation analysis to identify research fronts: A case study with the Internet of Things. *Science & Technology Libraries* 37 (2):171–86. doi:10.1080/0194262X.2017.1415183.

Mair, P. 2018. *Modern psychometrics with R*. Springer International Publishing. Use R!.

Manyika, J., M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon. 2015. Unlocking the potential of the Internet of Things. McKinsey Global Institute. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world.

Mishra, D., A. Gunasekaran, S. J. Childe, T. Papadopoulos, R. Dubey, and S. Wamba. 2016. Vision, applications and future challenges of Internet of Things: A bibliometric study of the recent literature. *Industrial Management & Data Systems* 116 (7):1331–55. doi:10.1108/IMDS-11-2015-0478.

Mohammad, S. 2011a. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. Proceedings of the ACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), 2011, Portland, OR.

Mohammad, S. M. 2011b. Even the abstract have colour: Consensus in Word-Colour associations. n *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*, 368–73. Portland, Oregon: Association for Computational Linguistics.

Mohammad, S. M., and P. D. Turney. 2010. Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34. Association for Computational Linguistics. https://dl.acm.org/doi/10.5555/1860631.1860635

Mohammad, S. M., and P. D. Turney. 2012. Crowdsourcing a word–emotion association Lexicon. *Computational Intelligence* 29 (3):436–65. doi:10.1111/j.1467-8640.2012.00460.x.

Mohammad, S. M., and T. W. Yang. 2011. Tracking sentiment in Mail: How genders differ on emotional axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 70–79. Portland, OR: Association for Computational Linguistics.

Mohammadzadeh, A. K., S. Ghafoori, A. Mohammadian, R. Mohammadkazemi, B. Mahbanooei, and R. Ghasemi. 2018. A Fuzzy analytic network process (FANP) approach for prioritizing Internet of Things challenges in Iran. *Technology in Society* 53:124–34. doi:10.1016/j.techsoc.2018.01.007.

Momtazi, S., and F. Naumann. 2013. Topic modeling for expert finding using latent dirichlet allocation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3 (5):346–53. doi:10.1002/widm.1102.

Moro, S., P. Cortez, and P. Rita. 2015. Business Intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation. *Expert Systems with Applications* 42 (3):1314–24. doi:10.1016/j.eswa.2014.09.024.

Mumtaz, S., A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, and J. Rodriguez. 2017. Massive Internet of Things for industrial applications: Addressing wireless IIoT Connectivity challenges and ecosystem fragmentation. *IEEE Industrial Electronics Magazine* 11 (1):28–33. doi:10.1109/MIE.2016.2618724.

Neuwirth, E. n.d. Package 'RColorBrewer. https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf.

Nikita, M. 2017. Ldatuning: Tuning of the Latent Dirichlet allocation models parameters. R package version 0.2.2. https://github.com/nikita-moor/ldatuning.

Nwazor, T. 2018. IoT security challenges and 5 effective ways to handle them. 2018. https://dzone.com/articles/iot-security-challenges-and-5-effective-ways-to-ha-1.

Ognyanova, K. 2015. Network visualization with R. POLNET 2015 Workshop, Portland OR. http://www.kateto.net/wp-content/uploads/2015/06/Polnet%202015%20Network%20Viz%20Tutorial%20-%20Ognyanova.pdf.

Oliver, J. 2018. Heatmaps in R. 2018. https://jcoliver.github.io/learn-r/006-heatmaps.html.

Peter, R. J., and B. Richard Watson. 2017. Research challenges for the Internet of Things: What role can or play? *Systems* 5:1. doi:10.3390/systems5010024.

Popescul, D., and M. Georgescu. 2014. Internet of Things–some ethical issues. *The USV Annals of Economics and Public Administration* 13 (2 (18)):208–14.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rajman, M., and M. Vesely. 2004. From text to knowledge: document processing and visualization: A text mining approach. In *Text mining and its applications*, ed. Spiros Sirmakessis, 7–24. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-45219-5

Randy, Q. 2018. Systematic literature review with quantitative analysis of scientific production: Scientometrics using R. 2018. http://rstudio-pubs-static.s3.amazonaws.com/410444_09a66192d93a467cb8e67c427b8b5277.html.

Read Text Files with Readtext. n.d. Accessed July 3, 2019. https://cran.r-project.org/web/packages/readtext/vignettes/readtext_vignette.html.

Rose, K., S. Eldridge, and L. Chapin. 2015. *The Internet of Things: An overview–understanding the issues and challenges of a more connected world*. United States: The Internet Society (ISOC).

Roussey, B. 2016. 5 challenges facing the Internet of Things. 2016. http://techgenix.com/internet-of-things-challenges/.

Rowe, M. n.d. DevOps challenges in the Internet of Things. Accessed July 21, 2019. https://techbeacon.com/enterprise-it/devops-challenges-internet-things.

Samuel, S. S. I. 2016. A review of connectivity challenges in IoT-smart home. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, 1–4. doi:10.1109/ICBDSC.2016.7460395.

Sfar, R., E. N. Arbia, Y. Challal, and Z. Chtourou. 2018. A roadmap for security challenges in the Internet of Things. *Digital Communications and Networks* 4 (2):118–37. doi:10.1016/j.dcan.2017.04.003.

Shen, W. 2017. Text and sentiment analysis on the world's most read book. 2017. https://rpubs.com/wes-shen/sentiment-analysis-most-read-book.

Sicari, S., A. Rizzardi, L. A. Grieco, and A. Coen-Porisini. 2015. Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks* 76 (January):146–64. doi:10.1016/j.comnet.2014.11.008.

Sievert, C., and K. Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. Baltimore, MD.

Sievert, C., and K. Shirley. 2015. *LDAvis: Interactive visualization of topic models*. R package version 0.3.2. https://CRAN.R-project.org/package=LDAvis.

Silge, J., and D. Robinson. 2016. Tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software* 1 (3):37. doi:10.21105/joss.00037.

Silge, J., and D. Robinson. 2017. *Text mining with R: A tidy approach*. USA: O'Reilly Media.

Srivastava, T., P. Desikan, and V. Kumar. 2005. Web mining–concepts, applications and research directions. In *Foundations and advances in data mining*, ed. Wesley Chu, Tsau Young Lin, 275–307. Berlin, Heidelberg: Springer.

Text Mining Example Codes (Tweets). n.d. Accessed July 3, 2019. https://rstudio-pubs-static.s3.amazonaws.com/66739_c4422a1761bd4ee0b0bb8821d7780e12.html.

Tian, K., M. Revelle, and D. Poshyvanyk. 2009. Using latent dirichlet allocation for automatic categorization of software. In *2009 6th IEEE International Working Conference on Mining Software Repositories*, 163–66. doi:10.1109/MSR.2009.5069496.

Tirunillai, S., and G. J. Tellis. 2014. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research* 51 (4):463–79. doi:10.1509/jmr.12.0106.

Trappey, A. J. C., C. V. Trappey, U. H. Govindarajan, A. C. Chuang, and J. J. Sun. 2017. A review of essential standards and patent landscapes for the Internet of Things: A key enabler for industry 4.0. *Advanced Engineering Informatics* 33 (August):208–29. doi:10.1016/j.aei.2016.11.007.

Tsai, C.-W., C.-F. Lai, and A. V. Vasilakos. 2014. Future Internet of Things: open issues and challenges. *Wireless Networks* 20 (8):2201–17. doi:10.1007/s11276-014-0731-0.

Wang, S.-H., Y. Ding, W. Zhao, Y.-H. Huang, R. Perkins, W. Zou, and J. J. Chen. 2016. Text mining for identifying topics in the literatures about adolescent substance use and depression. *BMC Public Health* 16 (1):279. doi:10.1186/s12889-016-2932-1.

Whitmore, A., A. Agarwal, and D. X. Li. 2015. The Internet of Things—A survey of topics and trends. *Information Systems Frontiers* 17 (2):261–74. doi:10.1007/s10796-014-9489-2.

Wickham, H. 2016. *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, H. 2017. Tidyverse: Easily install and load the "Tidyverse." R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse.

Wickham, H., and L. Henry 2019. Tidyr: Easily tidy data with "spread()" and "Gather()" functions. R package version 0.8.3. https://CRAN.R-project.org/package=tidyr.

Wickham, H., R. François, L. Henry, and M. Kirill 2019. Dplyr: A grammar of data manipulation. R package version 0.8.0.1. https://CRAN.R-project.org/package=dplyr.

Xiao, N., and L. Miaozhu 2019. Scientific journal and sci-fi themed color palettes for ggplot2. 2019. https://nanx.me/ggsci/articles/ggsci.html.

Zhao, Y. 2012. *R and data mining: Examples and case studies*. USA: Academic Press.

Zhao, Y. n.d. Social network analysis Accessed July 3, 2019. http://www.rdatamining.com/examples/social-network-analysis.

Ziegeldorf, J. H., O. G. Morchon, and K. Wehrle. 2014. Privacy in the Internet of Things: Threats and challenges. *Security and Communication Networks* 7 (12):2728–42. doi:10.1002/sec.795.

Zubiaga, A., R. Procter, and C. Maple. 2018. A longitudinal analysis of the public perception of the opportunities and challenges of the internet of things. *PLoS ONE* 13 (12):e0209472. doi:10.1371/journal.pone.0209472.