Exploratory Analysis of Internet of Things (IoT) in Healthcare: A Topic Modeling Approach

Ramakrishna Dantu  Indika Dissanayake  Sridhar Nerur
California State University, University of North Carolina, University of Texas at Arlington
Sacramento Greensboro Arlington
dantu@csus.edu j_dissan@uncg.edu snerur@uta.edu

Abstract

The rapid integration of the physical and cyber worlds through the Internet of Things, or IoTs, is transforming our lives in ways that we could not have imagined even five years ago. Although they are still in their infancy, IoTs have already made a significant impact, particularly in the healthcare domain. The purpose of this study is to unravel key themes latent in the sparse but growing academic literature on the application of IoTs in healthcare. Specifically, we performed topic modeling and identified five dominant clusters of research, namely, privacy and security, wireless network technologies, applications, data, and smart health and cloud. Our results show that research in healthcare IoT has mainly focused on the technical aspects with little attention to social concerns. In addition to categorizing and discussing the topics identified, the paper provides directions for future research.

1. Introduction

Current technological trends are rapidly transforming organizations into digital businesses that will engage with the world in ways that we cannot yet fathom. AI, NLP, Blockchain, Cloud-readiness, Internet of Things (IoTs), and augmented & virtual reality are some of the technologies at the heart of this revolution. These have serious implications for how organizations evolve their enterprise architecture, plan their IT infrastructure, manage their data from its source to the point at which it is consumed (i.e., data governance), and manage their security/privacy. This study is an exploratory analysis of the current state of academic research on the application of IoTs in the healthcare domain.

The Internet of Things, or IoTs, refers to a network of interconnected objects and devices with embedded software and sensors that facilitate the interchange of data and communication between the physical and virtual (i.e., Internet/Cyber) worlds [1]. The last few years have witnessed an explosion of data in terms of volume (amount of data), variety (data formats), and velocity (speed at which data is generated). It is estimated that Internet users around the world generate over 2.5 quintillion bytes of data every day [2]. This growth was fueled in no small measure by the proliferation of Internet-based applications and the ubiquity of smartphones, tablets, wearables, smartwatches, and other handheld devices. IoTs rely on the portability and interconnectivity of such devices to provide enormous amounts of real-time data that organizations can harness to derive actionable insights.

IoTs are redefining how people connect, interact, and exchange information, transforming practically every industry. Examples of Internet-enabled devices include kitchen appliances such as refrigerators, smart TVs, vending machines, transportation vehicles such as cars, and even devices used for monitoring the heart (e.g., pacemakers) and blood pressure. This list of “things” continues to grow with more devices being added every day. By 2020, over 20 billion connected things will be in use [3] and is projected to grow to over 75 billion by 2025 [4]. These connected devices are expected to generate more than 20 trillion gigabytes of data by 2025 and will have huge implications for healthcare, particularly in the areas of remote and real-time monitoring of patients, telemedicine, chronic illness management, physician-patient interaction, patient engagement, elderly care, and influencing patient behavior, to name a few. By 2020, 40% of IoT-related technology will be health-related, constituting a $117 billion market that will be more than for any other domain [5]. It is conceivable that our lives will change dramatically as IoTs realize their fullest potential.

The application of IoTs in the healthcare domain is still in its nascent stage. However, over the last few years, there has been a marked increase in the number of articles, both academic and non-academic, that have looked at IoTs in the context of healthcare. Researchers and industry experts have discussed a variety of issues and recurring themes that are latent in this corpus. Unraveling these themes and assessing the extant literature in terms of what has been done so far should be of immense value to academics and practitioners alike. Such an effort would not only help us address current opportunities and challenges in the use of IoTs in healthcare but also anticipate what lies ahead. This exploratory study aims to satisfy this need by using
text analytics, particularly topic modeling, to explicate the dominant themes underlying the existing corpus on the use of IoTs in the healthcare domain.

The remainder of the paper has been structured as follows. In the next section, we provide a brief overview of healthcare IoT. Next, we provide a description of the research methodology used in this study. This is followed by a discussion of the results and an elaboration of the topics identified. The concluding section summarizes our study and provides directions for future research.

2. Healthcare IoT

IoT solutions have been either applied or proposed in several areas of patient care, including but not limited to monitoring blood glucose levels [6], blood pressure [7], heart rate [8], blood oxygen level, and medication. For example, Istepaninan et al.[6], examined continuous, non-invasive monitoring of glucose levels using mobile Internet of Things (m-IoT). Yang et al.[8] designed and implemented a wearable using cloud-based IoT architecture for monitoring electrocardiogram (ECG). In a research study related to blood oxygen saturation measurement, authors Fu and Liu [9] proposed an IoT-based near-infrared portable tissue oximeter to measure blood oxygen saturation and heart rate parameters. These are but a few examples of the growing number of healthcare IoT applications.

Healthcare is subject to stringent HIPAA (Health Insurance Portability and Accountability Act) regulations [10], which provides data privacy and security provisions for safeguarding medical information. IoT adoption in healthcare can be fostered with better IoT privacy and security features. From a security standpoint, studies have proposed several IoT frameworks consisting of rules, guidelines, protocols, and standards that can enable implementation of IoT applications in various domains (See [11], [12]). Given the need for protecting privacy and confidentiality of patient health information and securing electronic health records, privacy and security aspects of IoTs are of considerable interest to both researchers and practitioners.

The overall objective of using IoT solutions in healthcare is to provide real-time, high quality patient care, particularly to those who don’t have the means to access medical facilities or see a physician. IoT solutions typically involve wearable devices or sensors ([13], [14]) that are connected over a network. These wireless technologies enable the collection and of data related to patients, providers, and hospitals, among others, in the cloud. Rapid strides in big data analytics now make it possible to analyze data in real-time using the cloud infrastructure. Researchers have proposed frameworks for collecting and processing IoT-related big data in a cloud environment ([15], [16]), and have discussed the attendant challenges and opportunities.

A recent study by Ng et al. [17] conducted a semantic similarity analysis of IoTs in general and identified 10 critical factors, namely, frameworks and challenges, current situation, interactions, security issues, application domains, data management, smart cities, and recommender systems [17]. We take one step further by specifically focusing on IoTs in the healthcare domain. Furthermore, we employ topic modeling to identify key research themes latent in the growing body of literature on the application of IoTs in healthcare. We further distinguish our work by articulating the social, technical, and organizational perspectives apparent in the extant literature.

3. Methodology

3.1 Data Collection

![Figure 1. Number of IoT-healthcare articles per year](image-url)

The Web of Science database was used to retrieve abstracts of papers on the application of IoTs in the healthcare domain. The search term\(^1\) yielded 819 articles. Of these, only 661 abstracts were available for analysis. As can be seen in Figure 1, the number of IoT-related articles published in the healthcare domain

---
\(^1\) Results: 819 (from Web of Science Core Collection)
You searched for: TOPIC: ((IoT OR "Internet of Things") AND "healthcare")
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Retrieved at 2:50 pm on March 28, 2018
grew at a very slow rate until 2013 and increased dramatically thereafter.

Out of the 661 abstracts extracted, 578 had keywords. Figure 2 represents the frequency distribution of keywords that were present in at least five articles. It is evident from the figure that technical issues dominate the current discourse on IoTs. Our exploratory analysis suggests that social and organizational aspects are underrepresented in the extant literature.

![Figure 2. Frequency Distribution of Keywords](image)

### 3.2. Topic Modeling

Latent Dirichlet Allocation (LDA), a popular algorithm for topic modeling (see [18]), was used to identify latent themes in the abstracts. Prior to performing LDA, the abstracts were preprocessed to reduce the inherent noise in text data. The first step was to normalize the text by converting it to lower case. Next, we removed stopwords, punctuation and digits. Stopwords refer to frequently occurring words (e.g., “the”, “a”, “is”, “of”) that are likely to diminish interpretability of the results. Likewise, punctuations and digits are eliminated because they add no value to the analysis. The final preprocessing step involved lemmatization. Researchers often use stemming (i.e., reducing words to a root form) or lemmatization (i.e., reducing to a base dictionary form or a lemma). In this study, we use lemmatization, as the form it reduces to is more interpretable and meaningful (see [19]). The lemmatized text was then subjected to LDA.

LDA requires us to specify the number of topics to be extracted, which is often not easy to determine. In our study, we used the LDA algorithm available in Python’s Scikit-Learn module to retrieve 50 topics. An examination of the relevance of the topics suggested that only about 25 topics were dominant in the corpus. Therefore, we proceeded to extract 25 topics using LDA implemented in MALLET\(^2\) (Machine Learning for Language Toolkit) from the University of Massachusetts at Amherst [20]. Subsequently, we identified five emerging categories across these 25 topics based on similarities and overlap among them. Finally, using the probability distribution of abstracts across the topics, we extracted five articles that contribute the most to each of the 25 topics, giving us a total of 125 articles. Two of the authors independently read these 125 articles and ratified the identified clusters.

### 4. Results and Discussion

This study focuses on five overarching and dominant clusters from the total of 25 identified. These reflect topics on privacy and security, wireless network technologies, applications (e.g., design/development of device/applications in healthcare, such as patient monitoring systems), data, and smart health and cloud.

\(^2\) [http://mallet.cs.umass.edu](http://mallet.cs.umass.edu)
Furthermore, our findings suggest that a number of studies are at the intersection of these clusters. For example, some studies focused on Radio Frequency Identification (RFID) authentication mechanisms involving aspects of wireless network technologies and security. To further validate our clusters, we used a software tool called VOSViewer (see www.vosviewer.com) that has been widely used for visualizing and understanding scientific domains [21]. Figure 3 shows the term map generated by VOSViewer based on the co-occurrence of words in our corpus. As can be seen, the cluster compositions are quite consistent with the topics that we abstracted from the results of our topic modeling analysis. In order to gain deeper insight into these clusters, we analyzed them with a view to eliciting social, technical, and organizational perspectives reflected in the current literature. The resulting framework is shown in Figure 4.

![Figure 3. Topic Clusters](image)

As we can see from the figures 3 and 4, most of the articles focused primarily on the technical challenges of IoTs. Our analysis shows that social challenges of IoT have received very little attention. For instance, technical aspects of security and privacy (e.g., advanced encryption mechanisms and authentication schemes) have been widely investigated but social concerns with regard to security & privacy of all stakeholders (e.g., patients’ trust), appear to be understudied. Moreover, ecological impacts (e.g., rare earth metals and toxic chemicals in devices) [22], awareness, and training are some other social challenges that received scant attention. Likewise, the other clusters (e.g., data) favor the technical dimension to social and organizational issues.

Furthermore, within the technical domain of healthcare IoTs, scholars seem to have lavished more attention on wireless network technologies. As can be seen in Figure 2, the most frequent keywords (e.g., wireless sensor networks, rfid, 6lowpan) are related to wireless network technologies.
To understand research trends more clearly, we graphed the frequency distribution of a few prominent keywords across time. This distribution, shown in Figure 5, also reveals that early research on IoT healthcare has largely failed to adopt a sociotechnical systems (STS) [23] that takes into account how people, processes, organizational structure, and technology will be affected by the proliferation of IoTs in our daily lives.

It is clear that security and privacy have become critical factors for the development and application of IoTs in any domain [18]. The need to address these concerns is even more important in healthcare, as it deals with highly sensitive medical data. Based on our analysis, the core articles in this cluster focus on several topics related to privacy and security of medical data and devices, including RFID authentication mechanisms, medical data security, and privacy. Security, privacy concerns, trust, safety, and mechanism are some of the keywords commonly associated with this cluster.

RFID is one of the key technologies used in IoT deployment in the healthcare domain. For example, RFID devices are used in wearable devices such as patient monitoring systems to identify and access patients’ medical records in real-time. Some research questions explored in this area include proposing and developing secure mutual RFID authentication schemes and protocols (e.g., elliptic curve cryptography based authentication protocols) that are suitable for healthcare applications [24] [25]. This area also partially overlaps with the wireless network technologies cluster.

4.1. Cluster 1: Security & Privacy

Figure 4. Challenges of IoT in healthcare

Figure 5. Keywords studied across the years
Protecting the privacy and confidentiality of patient data is at once critical and challenging. Privacy violations could hurt not only the patients but also healthcare firms [26]. These violations could lead to legal battles, erode patients’ trust, and threaten the long-term viability of healthcare firms. Some examples for research topics explored in this area include proposing blind cloud storage to protect privacy through maintaining anonymity of patients’ medical records [27], and advanced encryption methods such as attribute-based encryption [28]. We have noticed that privacy and confidentiality also partially overlaps with the data cluster.

4.2. Cluster 2: Data

The application of IoTs in healthcare generates massive amounts of medical data every day. This opens up a window of research opportunities in the areas of collecting, storing, handling, and analyzing medical data, improving data quality, and safeguarding data privacy and security. Some common keywords that appeared in this cluster include data, big data, method, approach, and algorithm.

Health IoT applications commonly use patient wearable devices, mobile, and social media applications to collect data. Research related to data collection process include topics such as improving validity of data [29], accurately and precisely taking measurements in real-time [30], and collecting and aggregating data using different sources such as mobile, wearables, and social media applications [31]. Not surprisingly, researchers have also focused their attention on effectively handling big data, which naturally overlaps with topics related to the cloud & smart health. Dominant topics in the big data realm included evolving an appropriate architecture to handle large-scale data (e.g., Hadoop-based medical emergency management system) [32] as well as facilitating the sharing of large and diverse real-time data among all the devices within a healthcare system [33].

Analyzing healthcare data is particularly challenging. Several studies have focused on clustering and classification methods as well as on algorithms suitable for medical data. Examples include classification schemes to group human behaviors and daily activities into different categories [34], and mechanisms to detect patients’ activities and events (e.g., sitting, falling) [35].

4.3. Cluster 3: Applications

This cluster includes the broad category of healthcare-related IoT applications, which includes the design and development of equipment and/or their components. Some common key words that appeared in this cluster include monitoring, diagnosis, treatment, medication, and caregiver. Patient monitoring devices and applications (mobile, wearable devices, and web based applications) are some of the main topics investigated in this cluster. Researchers have primarily focused on design and development of devices used for proactively detecting symptoms and monitoring physical and cognitive functions of patients. Typical examples include sensing systems for facilitating long-term mobility of the elderly [36], detecting early symptoms of dementia in elderly people living alone [37], detecting aging related neurological diseases [38], detecting voice disorders [39], monitoring foot pressure wirelessly [40], detecting vital signals [41], and devices for pregnant women to check their and fetus vitals [42]. In addition to patient monitoring systems, IoT applications are also used in preventive management systems, and detecting diseases such as zika, diabetes, and chikungunya [43] [44].

IoT adoption is another area in this cluster that has received scant attention. Researchers on IoT adoption have investigated facilitators of and barriers to patients’ adoption of healthcare IoTs [45], and have also explored ways to facilitate IoT adoption by urban poor communities [46].

4.4. Cluster 4: Wireless Network Technologies

IoT applications and wearable devices use wireless sensors to remotely collect and transmit data. This cluster discusses topics including but not limited to wireless sensor network standards and protocols for healthcare applications [47], network mobility management (NEMO) (to provide location independent continued connectivity for patients) [48], and power efficiencies (e.g., energy autonomous wearable healthcare devices) [49]. Some common key words that appeared in this cluster include node, network, protocol, energy efficiency, and ipv6.

4.5. Cluster 5: Cloud & Smart Health

This cluster includes research related to cloud & fog computing, and smart health applications such as IoT applications used in monitoring patients living in smart homes. Quite understandably, topics in this cluster overlap with research themes in the IoT application cluster that we discussed earlier. For
example, smart health applications are also part of IoT applications. Some common key words that appeared in this cluster are cloud, fog, computing, smart, and home. Tyagi et al. [50] proposed a cloud-based conceptual framework for IoT applications in the healthcare domain. Bibani et al. [51] proposed a prototype to demonstrate how an IoT healthcare application can be designed and developed in hybrid cloud/fog environment.

5. Conclusions and Future Research

We are living in an increasingly connected world where physical appliances, autonomous vehicles, drones, wearables, and smart “things” (e.g., homes, phones, watches, televisions, energy meters) equipped with software and sensors can seamlessly integrate the physical and virtual worlds. The Internet is what facilitates the interchange of data among these “things”, and hence this network is referred to as the “Internet of Things” (i.e., IoTs). Research on IoTs has proliferated in recent years and there is a need for us to understand what has been explored thus far in order to anticipate what challenges and opportunities lie ahead.

Since IoTs are rapidly transforming the healthcare industry, our paper focused on the research that has been done in this area. Specifically, this paper explored academic research related to IoTs in the healthcare domain. Our preliminary analysis showed that researchers are mainly focused on technical aspects and many important social aspects are under investigated. Using topic modeling, we identified five main research areas, namely, security & privacy, data, wireless network technologies, cloud & smart health, and applications. Furthermore, we identify the social, technical, and organizational themes apparent in the corpus we studied. Thus, our paper elucidates key research themes and provides a platform for evolving a roadmap for future research on the transformative potential of IoTs in the healthcare domain.

Our analysis suggests that most of the studies on IoTs in healthcare adopt a rather mechanistic, instrumental approach grounded in hard systems thinking [23]. The advocates of sociotechnical systems (STS) thinking have long argued that an integrated and holistic approach that blends the technical (i.e., hard systems, mechanistic metaphor) with the social dimension (i.e., soft systems thinking, human/stakeholder concerns) is paramount to build work systems that guarantee organization effectiveness while enhancing the quality of life of humans. An STS perspective that is sensitive to the psychological needs of stakeholders (e.g., patients) and to the empowerment (e.g., training) and active engagement of humans is conspicuously absent from the current discourse. Such an approach would also embrace broader ramifications of IoTs, such as ethical and social dilemmas that inevitably ensue when new technologies are introduced. It is apparent that opportunities to research the interplay among IoTs, people, processes, and organizational structures abound.

As an extension to this study, the authors intend to investigate practitioners’ perspective of IoT applications in healthcare by exploring website data. This would enable us to determine how compatible academic research is with state-of-the-art in industry. We believe this is a small but important step towards understanding what research has to be undertaken to transform the delivery and quality of healthcare through IoTs.

6. References


