Users Issues in using the Internet of Things Systems

Andrew Truelove, Farah Naz Chowdhury, Omprakash Gnawali, and Mohammad Amin Alipour
Department of Computer Science, University of Houston, United States

***Abstract*—Internet of Things (IoT) systems are bundles of networked sensors and actuators that are deployed in an environment and act upon the sensory data that they receive. These systems, especially consumer electronics, have two main cooperating components: a device and a mobile app. The unique combination of hardware and software in IoT systems presents challenges that are lesser known to mainstream software developers and might require innovative solutions to support the development and integration of such systems.**

In this paper, we analyze the more than 90,000 reviews of ten IoT devices and their corresponding apps and extract the issues that users encountered in using these systems. Our results indicate that issues with *connectivity*, *timing*, and *update* are particularly prevalent in the reviews. Our results call for a new software-hardware development framework to assist the development of reliable IoT systems.

I. Introduction

Internet of Things systems (IoT) are sets of interconnected sensors and actuators that potentially backed and managed by servers on the Internet. These systems are becoming part of "smart" solutions to the everyday life of users. For example, traditional thermostat, a solution for controlling the room temperature, can be replaced by smart thermostat that can extract the users' preferences and can be controlled remotely.

Despite the popularity of IoT solutions, the development of such systems still seems to a form of art, and the potential issues facing users are largely unknown. A systematic identification of problems would enable researchers to devise tools, techniques, and frameworks to support effective development of such systems. In this paper, we use the users review on Amazon and Google marketplaces to elicit the issues in IoT systems. We particularly focus on IoT consumer electronics that are used by home users. Most consumer electronics have two main components: a physical device, and a mobile app. Marketplaces such as Amazon.com and app stores allow users to leave reviews about devices and the mobile apps.

In this paper, we analyze over 90,000 reviews from ten IoT consumer electronic systems to understand what are the common issues that users are facing. We evaluate all reviews from January to mid-October 2018 for ten popular devices from Amazon.com and their corresponding Android apps from Google Play. Our results indicate that issues with *connectivity*, *timing*, and *update* are particularly prevalent in the reviews. The results call for new software-hardware development framework to assist development of reliable IoT systems.

Contributions. This paper makes the following contributions.

- We identify technical issues in ten consumer IoT systems by analyzing users reviews on Amazon and Google Play.
- We make data and analysis code available.

II. Related Work

There is a large body of work in analyzing users' reviews to elicit the issues in software systems. To the best of our knowledge, extracting users' issues in IoT technology, at least in the form of consumer electronics, have not been explored.

Atrozi et al. [4] survey the definitions, architecture, fundamental technologies, and applications of the Internet of Things. They note that IoT has been deployed in the area of mobile apps and that mobile devices will expand the IoT market as they continue to develop. Alur et al. [3] provide a list of challenges in development of IoT systems. Fu et al. [7] report the potential safety and security issues in IoT systems.

Maalej and Nabil [14] used the probabilistic technique to automatically classify app reviews into one of four types: bug reports, feature requests, user experience and ratings. Hoon et al. [11] studied how reviews evolve over time and the characteristics of the reviews. Pagano and Maalej [16] explore how and when users leave their feedback and also analyzed the content of the review. AppEcho allows users to leave feedback in situ that is exactly when the user discovered an issue [17]. AR-Miner extracts informative information from the reviews using topic modeling [5].

Hermanson [10] looked at whether perceived ease of use and perceived usefulness were widely discernable in user reviews submitted to apps on the Google Play Store. The author collected 13,099 reviews from the Google Play Store and found that only 3% of the reviews had information relating to perceived usefulness and that less than 1% of the reviews had any mention of perceived ease of use.

CLAP is a tool to help developers parse through app reviews when rolling out an update [18]. CLAP categorizes user reviews based on the information they contain. It groups related reviews together, and then automatically prioritize which groups should be prioritized for the next app update. Gu and Kim [9] propose SUR-Miner, a pattern –based parsing technique which parses aspect-opinion pairs from review sentences to produce effective user review summarization. Di Sorbo et al. [6] propose a tool called SURF which summarizes thousands of app reviews and generates a detailed interactive agenda on recommended updates and changes to the app. Licorish et al. [13] used content analysis and regression to provide insights into the nature of reviews provided by the users. Mujahid et al. [15] looked at user reviews of wearable apps. The authors manually sampled and categorized 6 android wearable apps. They found that the most frequent complaints involved functional errors, lack of functionality and cost.

TABLE I: IoT Devices and Applications Used in this Study

First Row: Name of App Second Row: Name of Device	Description
Amazon Alexa Amazon Echo Dot (2nd Gen)	A virtual assistant. App connects to a variety of devices with speakers and microphones that allows the user to interface with the service.
ecobee ecobee4 Smart Thermostat	Connects to a thermostat that can be controlled by the app.
Google Home Google WiFi System, 1-Pack	A virtual assistant. App connects to a variety of devices with speakers and microphones that allows the user to interface with the service.
Insteon for Hub Insteon Hub	Connects to a hub device that, in turn, connects to a number of other Insteon devices, including light switches, lamps, and security camera. Through the hub, the user can control all connected devices with the app.
Kevo Kevo Lock (2nd Gen)	Connects to a door lock that can be installed in the user's door. Lock can be controlled with the app.
Nest Nest T3007ES Thermostat	Connects to a thermostat that can be controlled by the app.
Philips Hue Philips Hue Starter Kit	Connects to light bulbs whose intensity and color are controlled by the app.
SmartThings (Samsung Connect) SmartThingsSmart Home Hub	Connects to a variety of Samsung-branded devices. These devices can be controlled through the app.
Tile Tile Mate	Connects to a small, square-shaped device that can be attached to a number of personal belongings. The device connects to the internet, allowing its location to be tracked through the app.
WeMo WeMo Mini Smart Plug	Connects to a number of WeMo-branded devices, including cameras, light bulbs, and electrical plugs. These devices can be controlled through the app.

TABLE II: Characteristics of Reviews Considered in this Study

System		Total	Review Length (char)				
			Min.	25%	50%	75%	Max
Amazon Alexa	App Reviews	5,785	1	18	56	135	2,027
	Device Reviews	54,289	3	44	92	192	7,632
ecobee	App Reviews	917	4	68	133	229	1,572
	Device Reviews	598	14	148.8	336.5	644	12,390
Google Home	App Reviews	7,051	2	26	73	157	1,996
	Device Reviews	1,859	9	102	240	468	9,526
Insteon	App Reviews	70	7	71.5	118.5	264.8	532
	Device Reviews	121	19	113	316	621	2,232
Kevo	App Reviews	461	3	33	93	206	1,724
	Device Reviews	296	15	154.8	337	719.2	5,016
Nest	App Reviews	1,798	3	61	135	242	1,877
	Device Reviews	1,431	9	83.5	210	462	5,139
Philips Hue	App Reviews	1,231	3	64	137	248	1,553
	Device Reviews	667	9	69	146	303.5	4,833
SmartThings	App Reviews	9,973	2	18	58	139	2,662
	Device Reviews	417	7	89	214	487	3,998
Tile	App Reviews	1,480	2	34	90.5	194	1,718
	Device Reviews	2,149	7	62	137	256	3,209
WeMo	App Reviews	3,177	2	40	85	177	1,833
	Device Reviews	2,013	5	100	215	385	7,841

III. Method

In this section, we first describe the data selection and characteristics of the review data used in this study.

A. Characteristics of Data

Table I lists the IoT systems (Devices and their corresponding apps) used in this study. These systems encompass a wide domain including conversational assistants, thermostats, electronic locks, and tracking devices. The price of the devices ranged from about $25 to $200 at the time of writing.Six of these systems were used in a previous study of IoT apps by Kaaz et al. [12], and the remaining four systems are based on a Google search for popular IoT apps. For each system, we found device on Amazon.com and the corresponding app on Google Play. We note that for some of the devices there are multiple versions of the products on Amazon website. In such cases, we chose the ones which had more reviews.

For each system, we extracted the reviews from the Amazon website and the corresponding app reviews from the Google Play Store. We collected reviews that were posted during a 10-month period starting from the beginning of January 2018 to mid-October of the same year.

Table II shows statistics about the number and length of reviews for products and apps. The table provides some noteworthy insights. For instance, with all IoT systems, the maximum review length was always higher in the device reviews than in the app reviews. It is possible that Amazon allows a higher character limit in its reviews than the Google Play Store. Moreover, users have to use a mobile phone to enter the app reviews, but they can use computers for leaving reviews for the devices on Amazon. Typing on a computer can be easier for many users than on phones.

For seven out of ten systems, more reviews were collected from the Google Play Store than Amazon. The three exceptions to this pattern are Amazon Alexa, Insteon, and Tile. With Amazon Alexa, this could be explained by the fact that Amazon is both the creator of the device and the curator of the storefront. As a first-party product, the Echo Dot likely receives some level of favoritism, likely expressed through increased promotion on the Amazon.com web site. This promotion could lead to more purchases and ultimately, more reviews. This favoritism may also explain why the Google Home app received so many more reviews than the Google Home device. The reason Insteon is an exception is probably due to the fact that it received fewer reviews overall. There is only a difference of 51 reviews between the app reviews and the device reviews. If Insteon had received more reviews during the time frame studied, the number of reviews may have more closely matched the pattern of the other systems. With Tile, no explanation for its anomalous behavior is immediately apparent. It is worth noting that Tile, as an IoT system, is fairly unique out of all the systems studied. These facts will be explored in more depth and explained in detail in the later sections of the paper.

B. Topic Modeling

We used Latent Dirichlet Allocation (LDA) to identify the most important topics users feel most strongly about [8]. By creating topics from the text of these reviews, it is possible, that some topics will be comprised of words that speak to a component of the app or device that users are complaining about. For example, if a topic contains the words "bad", "battery", and "drain", then we could infer that complaints about battery life are a significant topic in the user reviews. We used Gensim library [1] with the default configurations to generate a list of topics. For each review, we used LDA to generate three topics and return the ten words for each topic that contributed the most to that topic.

IV. Issues mentioned in IoT System Reviews

This section describes the result of analysis of users reviews for the systems in our study. For each IoT system, we generated three topics made up of ten words. Our results listed these ten words in the order of how much they contributed to that topic. For brevity, we discuss the analysis of two systems

in details here, and we add the results of the topics discussed in the reviews of other systems in Appendix A.

Tables III and IV depict the words for each topic for the Amazon Alexa and SmartThings apps. Tables V and VI display the topics for the corresponding devices. Beside each word is a number from 0 to 1 that reflects the magnitude at which that word contributed to the topic. When it comes to interpreting the LDA results, it was clear some words in a list appeared to be more important than others. Determining the usefulness of a word was based on a combination of its position in the list and the magnitude value the word had been assigned. A higher magnitude means a word contributed to the topic more strongly, meaning it is likely to be more integral in identifying the topic created by the LDA. At the same time, each topic list spans a different range of values between the magnitude of the first word and the magnitude of the tenth word. In some cases, the final few words had magnitudes so low to appear almost negligible, but in other cases, the final words carried magnitudes not all that lower than the value for the first word in that list.

For example, in Table V, the eighth word in Topic 3 is "christmas", which has a magnitude of 0.017. Though its position near the end of the list means this word may be one of the least important words in Topic 3, its impact is not entirely negligible. Compare the magnitude value of "christmas" in Topic 3 to the magnitudes found in Topic 1. The only word in Topic 1 with a magnitude higher than 0.017 is the first word, "speaker", which has a magnitude value of 0.021. Every word following has a lower magnitude value than "christmas". This means that "christmas" had more of an impact on its topic than nine of the ten words listed for Topic 1. This would suggest that the magnitude values of each word relative to the other magnitude values in the same topic carry more importance than the absolute position in any list.

If an IoT system is receiving significantly different rating distributions from the app store page and device store page, perhaps the kinds of topics generated from the app reviews and the device reviews may illustrate why.

A. Apps vs. Device

In a very general sense, the topics for the apps had more instances of words with negative sentiment than the topics for the devices. Though there are plenty of positive words in both the app and device topics, when a negative word like slow, bad, waste, or useless does appear, it seems to be more likely to be in an app review topic. Additionally, words such as control and connect appear more prominently in the app review topics, which may be an indicator of what issues users are running into when suing the app. The word update is particularly common in the app review topics.

Observation 1: Topics for the apps had more instances of words with negative sentiment than the topics for the devices.

As an example, none of the topics for the SmartThings Hub device contain any significantly negative language (Table VI).

TABLE III: Amazon Alexa App LDA Topics

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
good	0.037	love	0.021	connect	0.018
music	0.026	device	0.018	time	0.016
play	0.019	update	0.016	wifi	0.015
great	0.015	slow	0.013	phone	0.014
nice	0.011	home	0.011	keep	0.013
amazing	0.008	list	0.010	update	0.012
control	0.008	awesome	0.007	android	0.009
song	0.007	take	0.007	device	0.009
voice	0.006	please	0.007	tried	0.008
time	0.006	phone	0.007	best	0.008

TABLE IV: SmartThings App LDA Topics

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
great	0.035	phone	0.036	tv	0.044
love	0.024	uninstall	0.030	connect	0.028
smartthings	0.023	permission	0.019	good	0.026
device	0.022	update	0.015	device	0.020
easy	0.016	bloatware	0.015	phone	0.017
home	0.014	disable	0.014	smart	0.015
smart	0.013	apps	0.014	time	0.013
classic	0.013	remove	0.012	bluetooth	0.011
useful	0.011	device	0.011	update	0.011
awesome	0.009	delete	0.011	remote	0.009
Topic 1 Summary: Ease of Use		Topic 2 Summary: Desire to Remove App from Device		Topic 3 Summary: Connecting Phone with App	

The only instance of somewhat negative language comes from a single appearance of the word issue in Topic 3, and even then, the word has a fairly low magnitude value of 0.006. Meanwhile, the topics for the SmartThings app (Table IV) contain significantly more negative language, particularly in Topic 2, where words like uninstall, bloatware, remove, and delete are all found. The presence of the words permission and update in this topic suggest that something about the SmartThings app's permission requirements and updates is being associated with users wanting to remove the app from their device.

Overall, the observations that can be made from these LDA results are fairly general. There are exceptions to the general observations identified above; some negative words do appear in topics for the device reviews, for example. Though the topics provide some guidance as to what kinds of issues users of the apps are facing, it may be possible to refine the results to make these issues more apparent. We decided to see if running an LDA specifically on the app reviews that came with a low star rating might provide more helpful information.

TABLE V: Amazon Echo Dot LDA Topics

Topic 1	Topic 1 Magnitude	Topic 2	Topic 2 Magnitude	Topic 3	Topic 3 Magnitude
speaker	0.021	music	0.045	love	0.139
device	0.016	play	0.025	great	0.084
sound	0.016	fun	0.024	easy	0.032
good	0.015	question	0.019	gift	0.023
home	0.011	ask	0.017	product	0.020
smart	0.010	weather	0.015	bought	0.020
better	0.009	thing	0.015	room	0.017
time	0.008	time	0.013	christmas	0.017
voice	0.008	answer	0.013	house	0.015
quality	0.007	know	0.013	family	0.012
Topic 1 Summary: Good Sound Quality		Topic 2 Summary: Capbility of Features		Topic 3 Summary: Good Gift for Family	

TABLE VI: SmartThings Hub LDA Topics

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
light	0.011	device	0.020	device	0.021
home	0.008	smartthings	0.016	smart	0.014
smart	0.007	home	0.010	home	0.013
time	0.006	product	0.009	new	0.007
smartthings	0.006	light	0.009	easy	0.007
lock	0.005	time	0.008	smartthings	0.007
sensor	0.005	smart	0.008	issue	0.006
product	0.005	support	0.007	automation	0.005
device	0.005	zwave	0.007	control	0.005
back	0.004	great	0.007	phone	0.005

TABLE VII: Prominent Words from LDA Topics of 1-Star App Reviews

System	Words				
	1	2	3	4	5
Amazon Alexa	hate	device	time	update	useless
ecobee	update	thermostat	time	internet	connection
Google Home	music	chromecast	time	device	update
Insteon	device	time	find	waste	version
Kevo	lock	update	door	phone	time
Nest	camera	thermostat	update	home	time
Philips Hue	light	update	bridge	time	connection
SmartThings	phone	permission	uninstall	access	connect
Tile	phone	time	find	battery	key
WeMo	device	time	product	switch	update

B. Issues in low-rated systems

We filtered the app reviews so only reviews that had a minimal 1-star rating were left in the text. The goal behind running the LDA on only the 1-star reviews was to see if it was possible to identify the aspects of the app and devices that were leaving users with a negative impression. As such, we did not focus on words dealing with sentiment or emotion. Instead, we looked at words related to the functionality and features of the apps and devices. Table VII shows some of the noteworthy words that appeared in the topics for each app and Table VIII shows the same for device when LDA looked only at the 1-star reviews. These are words that had high magnitude values or that appeared in multiple topics.

Going over all the topics, a handful of relevant words seemed to appear with a greater frequency than others in the apps. For example, for all apps except for Kevo, at least one topic contained either the word "connect" or "connection". The prevalence of these words suggests that users of these apps have experienced some issue with connecting their phone to another device or network. The frequency in which "connect" and "connection" appears can mean that these connection issues are perhaps a greater source of frustration for users of IoT apps in general. Another noteworthy word was update. This word appeared in topics for all apps except for Insteon for Hub and Tile. It is important to note that the context for this word may not be the same in every appearance in the tables. For example, it is possible that some topics use "update", because an update was the source of a problem. It is also possible that the word appears in the context of users requesting an update to fix a problem with the app. However, the prevalence of the word does indicate that updates are an important part of app development and care should be taken in determining how they are implemented.

Home was another common word that appeared for six apps. With Google Home, this is not all that surprising, since home is part of the app's name. As for the other apps, the frequency of the word might suggest that many of these apps are indeed utilized for personal, home use. Making sure that these apps remain suited to this kind of use is another important thing for developers to keep in mind.

The word that appeared with the greatest frequency, however, was "time". This word appeared for all 10 apps; for most of the apps. With the exceptions of ecobee and Insteon, "time" actually appeared in at least two of the three topics for every app. Similar to "update", "time" does not necessarily have a single meaning in every one of its appearances. For apps like Philips Hue, the word appears to refer to the user's ability to configure through the app the time in which their light bulbs are set to turn on, turn off, change color, and so on. In these cases, the word "time" seems to relate more to scheduling functions of the app. In other cases, such as with Amazon Alexa, "time" appears in conjunction with words like "slow". Here, "time" is used to refer more to the duration of a function. The word appears in at least one of these contexts for every app. The prevalence of the word suggests that issues involving time are also an important element of these low-rated reviews. Resolving issues involved with timing settings as well as working to reduce the duration of app functions appear to both be issues app developers may want to pay attention to.

Observation 2: Issues with *connectivity*, *timing* and *update* are prevalent in the reviews of apps.

In the 1-star rated device reviews, in addition to timing and connectivity issues were also prominent. Anothe topic that seems to frustrate users in half of the devices is "support". Closer investigation of term "support" in the reviews revealed that in Insteon this term is largely referring to the issue of discontiuation of support of certain device, e.g., "INSTEON has stopped supporting their first cameras". In a fast-paced market such as IoT, abandaning of product might happen, but it is far from ideal. It indicates that the design of systems does not afford an efficient maintanence of the systems. Unsupported devices also known as zombie devices pose serious security, privacy and safety threats to the users [7] In ecobee, Google Home, Nest, , and Philips Hue, the term "support" was mostly refering to the customer support.

Observation 3: Issues with *connectivity*, *timing*, and *support* are prevalent in the reviews of the device.

V. Discussion

The intent behind running topic modeling on the app and device reviews was to help identify those functions and features of the IoT system that appeared to be the most important to its users. After seeing the greater distribution of 1-star reviews in the apps compared to the devices, we were interested in discovering whether the LDA results would in particular help

TABLE VIII: Prominent Words from LDA Topics of 1-Star Device Reviews

System	Words				
	1	2	3	4	5
Amazon Alexa	time	device	star	music	sound
ecobee	thermostat	support	product	system	temperature
Google Home	wifi	device	router	product	support
Insteon	support	device	customer	sensor	year
Kevo	lock	door	phone	time	product
Nest	thermostat	support	product	time	heat
Philips Hue	bulb	light	bridge	support	turn
SmartThings	product	device	home	time	new
Tile	phone	battery	key	time	product
WeMo	device	switch	connect	smart	time

identify the characteristics of the apps that were causing users to leave negative reviews. The topics generated by the LDA from the full review texts provided fairly general information. Negative words appeared to be more common in the app review topics than in the device review topics, for example.

Running LDA on the 1-star app reviews only seemed to produce slightly more tangible results. Words like "time", "update", and "connect" were particularly frequent among these topics. Each of these words is related to different aspects of an app's functionality that can be a focus for developers. Though it is likely that the process can be refined further to be more effective, the results suggest that topic modeling approaches such as LDA can be used to help identify issues users may be dealing with when using the app.

The three prominent issues of timing, connectivity, and update shed lights on some facets of IoT systems that are rarely encountered in developing mainstream software systems. Powerful processors, abundant memory, optimizing compilers have largely resolved the problem of timing and efficiency in the development of software. However, in systems that work on limited processing power and memory such as IoT devices and the mobile systems, efficiency has become an issue.

Moreover, fast, reliable networks with negligible latency is a given in the development of traditional software systems. It has been achieved by development of technologies and tools that reduce the latency of network connections; for example, nowadays, almost all cloud service providers automatically move the running instances of applications to data centers closer clients. It seems that we need new technologies to address this problem for IoT systems.

The problem of automatic update and backward compatibility in traditional software systems have been under investigation for many years. Nowadays, thanks to standardization of operating systems and protocols there are frameworks that strive to (almost) seamless update of software. For example, Android, Windows, and MacOS allow developers to update their applications using the corresponding app stores. However, update for IoT systems that a large portion of the hardware and protocols have not been standardized poses new challenges that require new tools and techniques.

Understanding issues and obstacles in operational IoT system allows us to devise techniques and tools to support effective development of these systems. We believe that analysis of user reveiews can contribute in better understanding of these systems by extracting first-hand experiences of users. We released the dataset and the source code of this study at https://github.com/atruelove/AppReviewAnalysis to replicate the study and to facilitate further analysis of the reviews.

VI. Threats to Validity

There are the following main threats to the validity of this study. First, our analysis was small in scope, we only used relatively recent reviews of a small number of IoT systems in our study. We also included the reviews from the Google Play app store not other app stores. Although small in scope, we believe that this study will provide the first glimpse of the users' issues in IoT systems. Second, we used LDA for topic modeling. It is known that LDA suffers from some limitations such as order effect [2]. To address these limitations, for given proposed words as topics, we manually checked the words to understand the intended meaning in the reviews and make sense of them.

VII. Conclusion

In this paper, we analyzed the reviews of ten IoT devices from Amazon and the corresponding apps from the Google Play Store. To the best of our knowledge, it is the first analysis of such systems. Our results suggest that (1) there are more negative topics in the mobile apps than the devices, and (2) efficiency, connectivity, and update seem to be prevalent issues in such systems. Our results call for the development of new tools and techniques to support practitioners to address these issues. We released the dataset and the source code of this study at https://github.com/atruelove/AppReviewAnalysis to facilitate further analysis of the reviews.

References

[1] gensim: topic modelling for humans. https://radimrehurek.com/gensim/, 9 2018.

[2] Agrawal, A., Fu, W., and Menzies, T. What is wrong with topic modeling? and how to fix it using search-based software engineering. *Information & Software Technology 98* (2018), 74–88.

[3] Alur, R., Berger, E., Drobnis, A. W., Fix, L., Fu, K., Hager, G. D., Lopresti, D., Nahrstedt, K., Mynatt, E., Patel, S., et al. Systems computing challenges in the internet of things. *Computing Community Consortium (CCC) Technical Report* (2016).

[4] Atzori, L., Iera, A., and Morabito, G. The internet of things: A survey. *Computer Networks 54*, 15 (2010), 2787 – 2805.

[5] Chen, N., Lin, J., Hoi, S. C. H., Xiao, X., and Zhang, B. AR-miner: mining informative reviews for developers from mobile app marketplace. In *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014* (New York, New York, USA, 2014), ACM Press, pp. 767–778.

[6] Di Sorbo, A., Panichella, S., Alexandru, C. V., Shimagaki, J., Visaggio, C. A., Canfora, G., and Gall, H. C. What would users change in my app? summarizing app reviews for recommending software changes. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering - FSE 2016* (New York, New York, USA, 2016), ACM Press, pp. 499–510.

[7] Fu, K., Kohno, T., Lopresti, D., Mynatt, E., Nahrstedt, K., Patel, S., Richardson, D., and Zorn, B. Safety, security, and privacy threats posed by accelerating trends in the internet of things. *Computing Community Consortium (CCC) Technical Report 29*, 3 (2017).

[8] FUJINO, I. Refining lda results and ranking topics in order of quantity and quality with an application to twitter streaming data. In *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)* (2014), IEEE, pp. 209–216.

[9] GU, X., AND KIM, S. " what parts of your apps are loved by users?"(t). In *Automated Software Engineering (ASE), 2015 30th IEEE/ACM International Conference on* (2015), IEEE, pp. 760–770.

[10] HERMANSON, D. New directions: Exploring Google Play mobile app user feedback in terms of perceived ease of use and perceived usefulness.

[11] HOON, L., VASA, R., SCHNEIDER, J.-G., AND GRUNDY, J. An analysis of the mobile app review landscape: trends and implications. *Technical report, Swinburne University of Technology* (2013), 1–23.

[12] KAAZ, K. J., HOFFER, A., SAEIDI, M., SARMA, A., AND BOBBA, R. B. Understanding user perceptions of privacy, and configuration challenges in home automation. In *Visual Languages and Human-Centric Computing (VL/HCC), 2017 IEEE Symposium on* (2017), IEEE, pp. 297–301.

[13] LICORISH, S. A., SAVARIMUTHU, B. T. R., AND KEERTIPATI, S. Attributes that predict which features to fix: Lessons for app store mining. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering* (2017), ACM, pp. 108–117.

[14] MAALEJ, W., AND NABIL, H. Bug report, feature request, or simply praise? On automatically classifying app reviews. In *2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings* (aug 2015), IEEE, pp. 116–125.

[15] MUJAHID, S., SIERRA, G., ABDALKAREEM, R., SHIHAB, E., AND SHANG, W. Examining User Complaints of Wearable Apps: A Case Study on Android Wear. *Proceedings - 2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems, MOBILESoft 2017*, August (2017).

[16] PAGANO, D., AND MAALEJ, W. User Feedback in the AppStore: An Empirical Study (submitted). *RE '13: Proceedings of the 21st International Requirements Engineering Conference* (2013), 125–134.

[17] SEYFF, N., OLLMANN, G., AND BORTENSCHLAGER, M. AppEcho: A User-Driven, In Situ Feedback Approach for Mobile Platforms and Applications. In *Proceedings of the 1st International Conference on Mobile Software Engineering and Systems - MOBILESoft 2014* (New York, New York, USA, 2014), ACM Press, pp. 99–108.

[18] VILLARROEL, L., BAVOTA, G., RUSSO, B., OLIVETO, R., AND DI PENTA, M. Release planning of mobile apps based on user reviews. *Proceedings of the 38th International Conference on Software Engineering - ICSE '16* (2016), 14–24.

APPENDIX

A. Results of topic modeling for IoT apps and devices

TABLE XIV: Philips Hue App

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”light”	0.041	”version”	0.011	”update”	0.031
”room”	0.016	”light”	0.010	”bridge”	0.028
”great”	0.014	”color”	0.007	”light”	0.023
”update”	0.013	”location”	0.007	”home”	0.018
”scene”	0.012	”scene”	0.007	”connect”	0.014
”time”	0.011	”routine”	0.007	”time”	0.008
”bulb”	0.010	”good”	0.007	”control”	0.008
”new”	0.010	”gen”	0.006	”connection”	0.008
”turn”	0.010	”easy”	0.006	”new”	0.007
”feature”	0.009	”feature”	0.006	”find”	0.006

TABLE XV: Tile App

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”great”	0.021	”love”	0.020	”phone”	0.039
”tile”	0.018	”phone”	0.018	”key”	0.029
”find”	0.016	”found”	0.010	”find”	0.020
”battery”	0.016	”key”	0.009	”time”	0.017
”time”	0.014	”easy”	0.008	”tile”	0.013
”phone”	0.014	”find”	0.008	”keep”	0.010
”key”	0.011	”location”	0.008	”lost”	0.009
”product”	0.010	”great”	0.008	”location”	0.008
”never”	0.009	”best”	0.008	”bluetooth”	0.008
”year”	0.008	”good”	0.007	”ring”	0.008

TABLE XVI: WeMo App

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”switch”	0.025	”great”	0.037	”device”	0.035
”light”	0.024	”easy”	0.033	”time”	0.023
”update”	0.021	”time”	0.027	”wifi”	0.019
”home”	0.021	”device”	0.018	”connect”	0.019
”turn”	0.017	”setup”	0.017	”switch”	0.013
”device”	0.015	”good”	0.015	”setup”	0.012
”smart”	0.012	”love”	0.013	”rule”	0.011
”plug”	0.011	”buggy”	0.012	”network”	0.011
”firmware”	0.011	”slow”	0.009	”plug”	0.011
”product”	0.011	”product”	0.008	”reset”	0.009

TABLE XVII: ecobee4 Smart Thermostat

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”thermostat”	0.017	”thermostat”	0.019	”thermostat”	0.022
”sensor”	0.015	”temperature”	0.011	”easy”	0.012
”room”	0.011	”home”	0.008	”great”	0.011
”device”	0.009	”house”	0.008	”support”	0.010
”great”	0.008	”feature”	0.007	”install”	0.008
”home”	0.008	”smart”	0.007	”product”	0.008
”support”	0.007	”love”	0.007	”sensor”	0.007
”temperature”	0.006	”sensor”	0.006	”system”	0.006
”smart”	0.006	”time”	0.006	”wire”	0.006
”house”	0.006	”room”	0.006	”house”	0.006

TABLE XVIII: Google WiFi System

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”wifi”	0.041	”router”	0.022	”wifi”	0.026
”house”	0.025	”easy”	0.019	”router”	0.018
”system”	0.014	”device”	0.016	”device”	0.017
”easy”	0.014	”wifi”	0.014	”network”	0.016
”router”	0.014	”setup”	0.011	”mesh”	0.007
”great”	0.013	”great”	0.011	”point”	0.007
”home”	0.013	”speed”	0.010	”house”	0.007
”signal”	0.013	”love”	0.008	”setup”	0.007
”speed”	0.012	”house”	0.008	”time”	0.006
”product”	0.009	”network”	0.007	”speed”	0.006

TABLE XIX: Insteon Hub Device

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”switch”	0.015	”device”	0.016	”device”	0.024
”control”	0.008	”year”	0.014	”new”	0.016
”product”	0.007	”product”	0.011	”scene”	0.012
”purchase”	0.007	”support”	0.010	”sensor”	0.011
”buy”	0.006	”new”	0.009	”light”	0.011
”de”	0.006	”account”	0.009	”switch”	0.008
”house”	0.005	”month”	0.008	”old”	0.008
”device”	0.005	”died”	0.007	”year”	0.008
”return”	0.005	”control”	0.007	”home”	0.007
”time”	0.005	”bought”	0.006	”product”	0.007

TABLE XX: Kevo Lock (2nd Gen)

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”lock”	0.029	”lock”	0.032	”lock”	0.046
”door”	0.014	”phone”	0.018	”time”	0.018
”key”	0.013	”door”	0.017	”door”	0.012
”phone”	0.009	”key”	0.011	”key”	0.012
”time”	0.008	”time”	0.009	”battery”	0.011
”product”	0.008	”unlock”	0.008	”phone”	0.009
”open”	0.007	”open”	0.007	”open”	0.007
”touch”	0.005	”plus”	0.007	”unlock”	0.006
”kwikset”	0.005	”great”	0.006	”great”	0.005
”bluetooth”	0.004	”bluetooth”	0.006	”fob”	0.005

TABLE XXI: Nest T3007ES Thermostat

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”easy”	0.027	”thermostat”	0.023	”thermostat”	0.020
”thermostat”	0.026	”wire”	0.018	”time”	0.009
”great”	0.019	”unit”	0.011	”house”	0.009
”love”	0.018	”system”	0.010	”home”	0.009
”install”	0.017	”installed”	0.009	”day”	0.008
”home”	0.015	”old”	0.007	”support”	0.008
”temperature”	0.013	”hvac”	0.007	”product”	0.007
”control”	0.009	”support”	0.006	”heat”	0.007
”product”	0.008	”new”	0.006	”ac”	0.007
”house”	0.008	”install”	0.005	”degree”	0.006

TABLE XXII: Philips Hue Starter Kit

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”product”	0.017	”light”	0.040	”light”	0.044
”bulb”	0.016	”home”	0.020	”bulb”	0.034
”great”	0.016	”bulb”	0.018	”great”	0.018
”light”	0.011	”control”	0.012	”love”	0.017
”easy”	0.009	”switch”	0.011	”turn”	0.016
”router”	0.008	”easy”	0.009	”easy”	0.014
”home”	0.007	”great”	0.009	”home”	0.013
”bridge”	0.007	”smart”	0.009	”setup”	0.008
”turn”	0.007	”turn”	0.009	”smart”	0.008
”good”	0.007	”kit”	0.007	”room”	0.006

TABLE XXIII: Tile Mate

Topic 1 Words	Topic 1 Magnitude	Topic 2 Words	Topic 2 Magnitude	Topic 3 Words	Topic 3 Magnitude
”battery”	0.034	”phone”	0.056	”key”	0.064
”year”	0.021	”great”	0.034	”time”	0.028
”tile”	0.013	”find”	0.034	”find”	0.020
”new”	0.011	”key”	0.026	”phone”	0.018
”month”	0.010	”love”	0.021	”lost”	0.016
”product”	0.010	”time”	0.012	”wallet”	0.014
”replace”	0.007	”bluetooth”	0.011	”easy”	0.013
”location”	0.007	”product”	0.011	”lose”	0.009
”time”	0.006	”bought”	0.011	”bought”	0.008
”buy”	0.006	”gift”	0.010	”month”	0.008