How Creative is the Crowd in Describing Smart Home Scenarios?

Tahir Abbas
Eindhoven University of Technology, Eindhoven, the Netherlands
t.abbas@tue.nl

Vassilis-Javed Khan
Eindhoven University of Technology, Eindhoven, the Netherlands
v.j.khan@tue.nl

Daniel Tetteroo
Eindhoven University of Technology, Eindhoven, the Netherlands
d.tetteroo@tue.nl

Panos Markopoulos
Eindhoven University of Technology, Eindhoven, the Netherlands
P.Markopoulos@tue.nl

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188509

Abstract
Internet of Things (IoT) is recently attracting vendors like Google, Homey, and Samsung that have already brought to the market a plethora of devices and services supporting smart home automation. However, recent studies have shown that end-users having little knowledge of the features and possibilities of IoT devices, face difficulties in conjuring up meaningful use scenarios that combine such devices. Therefore, they fail to anticipate useful configurations beyond those provided by vendors and hence missing out on the vast potential of the IoT. We present an on-going investigation that explores the potential of sourcing IoT-relevant scenarios from a popular microtask-crowdsourcing platform, and a preliminary evaluation of such scenarios with respect to their originality and practicality. This work paves the way for the automated leverage of crowdsourced user scenarios to support IoT end-users.

Author Keywords
Crowdsourcing; IoT; smart home; end users; creativity evaluation techniques; similarity measures;

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

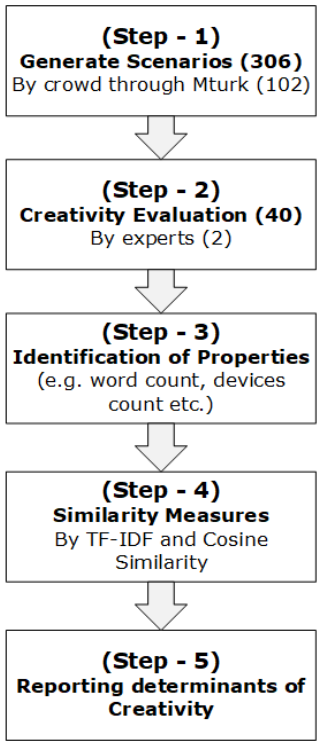

Figure 1: Research Method

Introduction & Related Work

Developments in IoT are increasingly enabling the realization of "smart home" scenarios for different purposes including health and wellness, cooking, elderly care, and communication [6]. However, this development has brought about the challenge of ensuring "interoperability" among different devices and services, which has been partly addressed by services such as IFTTT, Zapier[1], Apiant[2] to name a few. More specifically, in IFTTT, end-users make "recipes" -as they are known- to program devices and applications by defining triggers to launch certain actions. But, there is still a considerable and constantly growing un-explored design space that can be realized by programming various combinations of trigger and action devices [11]. Recent research shows that end-users face significant barriers before even facing the challenging task of programming triggers and actions [2]. As the number of devices and services increases the potential combinations of those devices and services in a smart home context increases exponentially [7]. Moreover, information provided by vendors on their websites does not help consumers to translate their high-level goals to useful scenarios [6], who consequently find it difficult to conceive combinations of devices and services that would suite their context and needs.

In an 18-month living-lab study with 14 households equipped with smart home technology, Jakobi et.al [2], investigated the different stages that smart home users go through. Those included system setup (planning), installation and configuration, domestication, daily use, reconfiguration and extension. Part of their study focused on the system setup and the planning stage that has been neglected in the past. One of the findings they reported regarding system setup, is that the participants in their study, apart from those who were technology savvy, could not articulate their needs and translate them to use cases combining more than a small number of hardware components.

In relation to end-user scenarios for smart-home technology Ogonowski et al. developed an online system called Shop&Play [9]. Using that system, end-users can select specific requirements related to security, comfort and energy saving. Then based on those requirements, they can select some pre-defined scenarios for automation of smart home systems and then order a complete pre-configured package from the vendors. However, there are several questions that arise: Where should these pre-defined scenarios come from? How can vendors anticipate and gather useful scenarios and provide them on their websites? How can inexperienced inhabitants, who are already living in smart homes and want to grow their system with new devices, get useful ideas? In this research we examine whether crowdsourcing could be a suitable approach to collecting and organizing such scenarios.

Researchers have recently shown the potential in utilizing crowdsourcing platforms to generate creative content. Whether it is sketches of chairs [12] , feedback on graphics [4] or even a user interface [3], crowdsourcing seems to be an interesting approach in creating creative content of all sorts. These crowd-based techniques do not only allow gathering quality ideas from crowd workers by showing them diverse examples [10], but also assist in assessing the

[1] https://zapier.com/

[2] https://apiant.com/

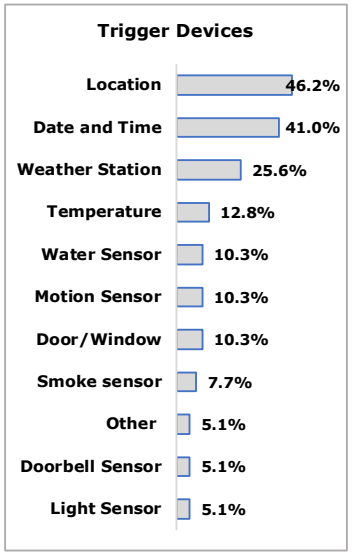

Figure 2: **Y-axis:** List of Trigger devices used in the study. The "Other" includes devices entered by users (e.g. Sleep detector and cell Phone); **X-Axis:** Usage Percentage of Trigger devices (One scenario was excluded due to ambiguity of devices)

creativity of individual ideas [12]. To the best of our knowledge, prior research has not yet examined the feasibility of using crowdsourcing to create smart home scenarios. In this study, we contribute to the existing literature by presenting the necessary steps in harnessing the crowd to generate creative content. In that way we offer a solution that can be repurposed by both IoT vendors and distributors for supporting end-user facing services.

Method

Through Amazon Mechanical Turk (MTurk) we investigated whether an anonymous crowd could assist smart home inhabitants in the creation of more diverse, creative and meaningful scenarios by combining the functionalities of given devices (see Fig. 2 & 3). We restricted the survey to U.S. workers with over 98% approval rating and 5000 HITs approved. We compensated each participant with $1.8 (US Minimum wage, 7.25$/hr) for an average of 15 minutes survey; the total cost of the study was $226.80.

On the first page of survey, we showed them a short video (1' 29'') that described the purpose of the research in detail. Inspired by affective priming theory [8], we aimed at inducing empathy in crowd workers by integrating images, videos, music etc. on the crowdsourcing platform. Next, we described the aims of the research and the researcher's profile.

In the next page, we asked crowd workers to imagine that *"you have a home with devices that are connected to the Internet. You are required to pick any combination of the devices and then write three meaningful scenarios related to the selected devices."* To avoid confusion, we used the word "input devices" for trigger devices/channels and "output devices" for action devices/channels. We also provided a field for entering devices they would wish which were not already on the list. We showed them a sample scenario and asked them to avoid creating generic, meaningless scenarios. We asked them to categorize each scenario by selecting some pre-defined categories (comfort, energy saving, home security, elderly care, child care, entertainment, health and cooking) or by adding their own category. At the end of survey, we ask them some general background questions such as, smart home experience, age, gender, family size, education and programming experience.

In total 102 MTurk workers wrote 306 smart-home usage scenarios (each wrote three). We report on a preliminary analysis of a subset of these scenarios, which we carried out to decide how best to conduct a full and detailed analysis of the scenarios obtained from the crowd. Specifically, of the 306 we randomly selected 40 scenarios which we examined to assess whether the crowd can provide meaningful and interesting scenarios and to get a feeling for the kind of scenarios they construct. Two independent experts evaluated the creativity of randomly selected 40 use case scenarios on a 7-point Likert scale. We used the same binary measures of creativity as previously used in evaluating the creativity of crowd input [12]. According to the binary measure of creativity, a creative idea should be both original and potentially practical; following the approach in [12], we also selected the mean score of 4.0 as the threshold value for both originality and practicality.

To get an overview for the complexity of the scenarios they created, we counted the number of trigger and

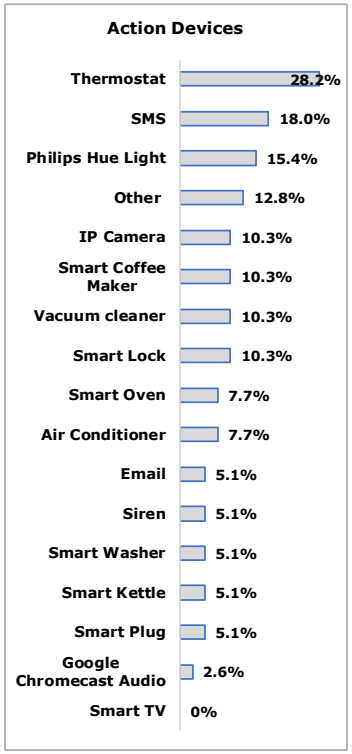

Figure 3: **Y-axis:** List of Action devices used in the study. The "Other" includes devices entered by users (e.g. Cloud Storage Service, Smart Water Tap, Smart Fridge, Cell Phone); **X-Axis:** Usage percentage of Action Devices

action devices as well as total number of triggers and actions used for each scenario. We also computed text metrics like words count, number of unique words, words per sentences and long words (> 6 letters) using Linguistic Inquiry and Word Count (LIWC)[3] tool.

We were also interested to know whether similarity measures can be applied to evaluate crowd-generated scenarios for their originality. This we assessed through cosine based similarity matrix and TF-IDF ("term frequency" and the "inverse document frequency") [5]. Cosine similarity is a vector-based measure of the similarity of two strings. In this method, we transform each string into a high dimensional vector space in which strings which are closer to each other are considered more similar and vice versa. We did that because we were also interested to see whether subjective evaluation of originality by experts could be gauged by such a similarity metric, i.e. its inverse was expected to relate to originality. We used the scikit-learn[4] Python library for implementing TF-IDF and cosine similarity. We took the mean of pairwise comparison. The result of this analysis was a score from 0 (not similar) to 1 (exactly similar) for each scenario. The research method is detailed in Fig. 1.

Results

Our sample population included 37 workers, since three participants completed the task repeatedly, so we excluded them. The sample was 43.24% male (16) and 56.76% female (21). They ranged in age from 25 to 65 (M=38.11, SD=9.56). Furthermore, they had different levels of programming experience (see Fig. 6). In total, 18 workers (48.6%) had smart home experience, which ranged from 12 to maximum of 120 months (M=12.89, SD= 24.5).

We observed that workers defined scenarios in either predefined categories or created their own categories. For example, among the predefined categories, *Comfort* and *Home Security* were the most popular (Fig. 4); the *Other* category includes user defined categories e.g. *Pet Care*, *Entertainment*, *Child Care* and *Cleaning*. These results show what IoT applications appeal to the crowd and reveal potentially unexplored categories for smart home applications. A quick search on world's leading vendor's websites like Google, Samsung etc. shows no use cases for *Pet Care* category.

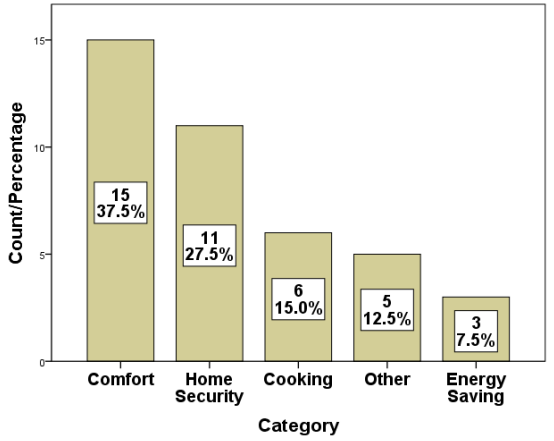

Figure 4: Prominent Categories in the sample

In total, 17(42.5%) scenarios were rated as creative by experts. See examples in Table 1. Next, we examined more closely these creative scenarios based on number of devices and their number of triggers and actions. We

[3] https://liwc.wpengine.com/

[4] http://scikit-learn.org/stable/

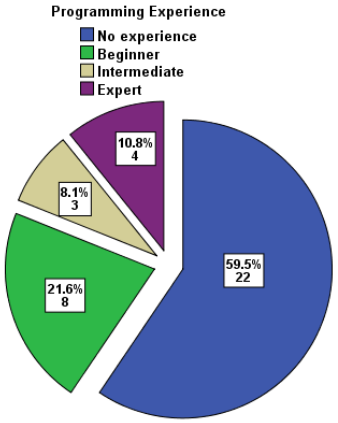

Figure 6: Programming Experience

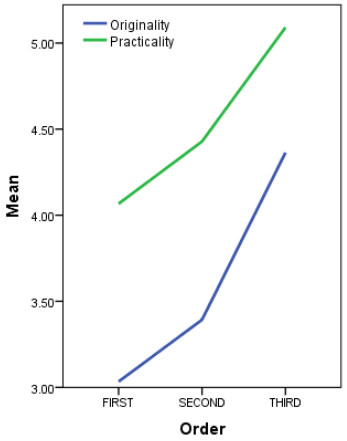

Figure 7: The mean scores for both originality and practicality increased with the order

found significant correlation between the originality and number of action devices (r=0.34, p<0.05) as well as between practicality and number of action devices (r=0.38, p<0.01). Contrary to our expectations, we found that originality and practicality were not tightly linked with number of triggers devices and their triggers. Nevertheless, we found a significant correlation between practicality and number of actions used (r=0.40, p<0.01).

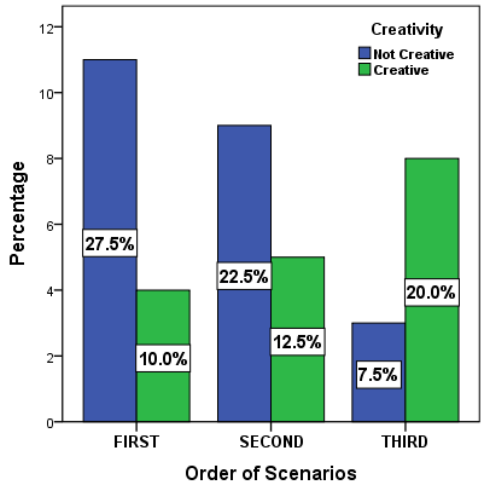

Figure 5: Impact of order on the overall creativity

Another important correlate of creativity was the total number of words used per scenario. We found that both originality (r=0.42, p<0.01) and practicality (r=0.47, p<0.01) were strongly correlated with word count. The average word count was 60 in our sample.

We found a weak inverse correlation between subjective evaluation of originality by experts and objective evaluation by similarity measures (r=-0.10, p = 0.542). This indicates that assessing originality through similarity metrics is not yet straight forward and more research is needed to find suitable AI techniques to automate the ranking of scenarios by creativity (something which is needed to scale up this process). Consequently, to analyze the full set of scenarios, we plan to crowd based evaluation techniques. This approach is supported by a recent study which has proved that crowd's evaluations were almost similar to the experts [4].

Furthermore, we were curious to know whether creativity increases with practice; in our case each worker wrote three scenarios and it is possible that they improved on the way. We found that mean score for both originality and practicality increased with the order (see Fig. 7). Determining the real impact of order on the overall creativity, we created a new dichotomous creativity variable (1=creative, 0=not creative) based on the originality and practicality scores. We encoded it 1 if both the originality and practicality scores were greater than 4.0 and 0 otherwise. We found that the overall creativity increased with each subsequent scenario χ^2 (1, N = 40) = 5.06, p<0.05 (Fig. 5).

We also found that creativity was not correlated with demographics such as: gender, programming and smart home experience which further confirms that anonymous crowd can create scenarios to support end-users.

Currently, we are looking to how these correlations could provide the basis for an automated classification of crowd-based scenarios.

Sample Scenarios:

Creative
[Male, 41] *"When I leave for work I would like the windows and doors to automatically lock themselves. I would then like this to send a signal to the vacuum cleaner which would sense that nobody is at home and it would be time to start vacuuming all the rooms. When finished it should automatically turn off. It would make a lot of comfort and sense to come back home after a long day to a well vacuumed and clean house"*
Less Creative
[Male, 36] *"When I get to within a few hundred feet of my home after 7pm, I want my lights to turn on automatically"*

Table 1: Examples of creative and less creative scenarios

Conclusion and Future Work

In this paper, we reported some preliminary results based on the analysis that we conducted on a data set of IoT application scenarios collected with MTurk. We evaluated these scenarios on various dimensions finding interesting correlations between creativity and different features of scenarios like word count, number of devices used and their actions, and the order in which workers wrote scenarios. Our results are encouraging regarding the feasibility of automatic filtering / ranking of scenarios by estimates of how creative they are based on these features as well as on TF-IDF based weighting schemes. Because only a weak inverse correlation was found between similarity measures and originality, we orient towards combining human based judgements with computational methods to assess originality [10]. Our ongoing analysis of the full data set will provide us a valuable experience for designing solutions that can provide support to end-users in constructing and evaluating IoT scenarios.

References

1. Faez Ahmed, Mark Fuge, and Lev D. Gorbunov. 2016. Discovering Diverse, High Quality Design Ideas From a Large Corpus. *Volume 7:* ASME'16, V007T06A008.
2. Timo Jakobi, Corinna Ogonowski, Nico Castelli, Gunnar Stevens, and Volker Wulf. 2017. The Catch(es) with Smart Home – Experiences of a Living Lab Field Study. *Proc. CHI'17*:1620–1633.
3. Walter S Lasecki, Juho Kim, Nick Rafter, Onkur Sen, Jeffrey P Bigham, and Michael S Bernstein. 2015. Apparition: Crowdsourced User Interfaces that Come to Life as You Sketch Them. *Proc. CHI'15*: 1925–1934.
4. Kurt L uther, Jari-Lee Tolentino, Wei Wu, et al. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. *Proc. CSCW & SC*.
5. Christopher D. Manning, Prabhakar Ragahvan, and Hinrich Schutze. 2009. An Introduction to Information Retrieval. *Information Retrieval* c:1–18.
6. Sarah Mennicken and Elaine M. Huang. 2012. Hacking the natural habitat: An in-the-wild study of smart homes, their development, and the people who live in them. LNCS, June: 143–160.
7. Sarah Mennicken, Jo Vermeulen, and Elaine M Huang. 2014. From Today's Augmented Houses to Tomorrow's Smart Homes : New Directions for Home Automation Research. *Proc*. *UbiComp '14*: 105–115.
8. Robert R. Morris, Mira Dontcheva, and Elizabeth M. Gerber. 2012. Priming for better performance in microtask crowdsourcing environments. *IEEE Internet Computing* 16, 5: 13–19.
9. Ogonowski Ogonowski, Philippe Hennes, and Maximilian Seiffert. 2016. Shop&Play Erlebnis im Smart Home: Nutzung statt Installationschaos. *Workshop Usability für die betriebliche Praxis*.
10. Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale - Leveraging Ideas from Others to Generate More Creative and Diverse Ideas Pao. *CSCW '15*: 937–945.
11. Blase Ur, Melwyn Pak, Yong Ho, et al. 2016. Trigger-Action Programming in the Wild : An Analysis of 200 , 000 IFTTT Recipes. *Proc. CHI'16*: 3227–3231.
12. Lixiu Yu and Jeffrey V Nickerson. 2011. Cooks or Cobblers? Crowd Creativity through Combination. *Proc. CHI' 11*: 1393–1402.