Dec 12th, 12:00 AM

Iterative Seed Word Generation for Interactive Topic Modelling: a Mixed Text Processing and Qualitative Content Analysis Approach

Morteza Namvar
The University of Queensland, m.namvar@business.uq.edu.au

Saeed Akhlaghpour
The University of Queensland, s.akhlaghpour@business.uq.edu.au

James Boyce
The University of Queensland, j.boyce@business.uq.edu.au

Salma Sharifi Khajedehi
The University of Queensland, s.sharifikhajedeh@business.uq.edu.au

Iterative Seed Word Generation for Interactive Topic Modeling: a Mixed Text Processing and Qualitative Content Analysis Approach

Short Paper

Morteza Namvar
The University of Queensland
Brisbane, Australia
m.namvar@business.uq.edu.au

Saeed Akhlaghpour
The University of Queensland
Brisbane, Australia
s.akhlaghpour@business.uq.edu.au

James Boyce
KPMG Australia
Brisbane, Australia
jboyce1@kpmg.com.au

Salma Sharifi Khajedehi
The University of Queensland
Brisbane, Australia
s.sharifikhajedehi@business.uq.edu.au

Abstract

Topic models have great potential for helping researchers and practitioners understand the electronic word of mouth (eWoM). This potential is thwarted by their purely unsupervised nature, which often leads to topics that are not entirely explainable. We develop a novel method to iteratively generate seed words to guide the interactive topic models. We assess the validity and applicability of the proposed method by investigating the critical phenomenon of Contact Tracing Mobile Applications (CTMAs) post-adoption during a time of the COVID-19 pandemic. The results show that constructs developed through our interactive topic modeling can capture primary research variables related to the phenomenon. Compared to existing topic modeling methods, our approach shows superior performance in explaining users' satisfaction with CTMAs.

Keywords: Online reviews, interactive topic modeling, seed words, research constructs

Introduction

The proliferation of socialized data in the form of online reviews offers a unique opportunity for improving service quality to customers and the public (Padmanabhan et al., 2022). Several text mining techniques, such as sentiment analytics (Deng et al., 2018; Namvar et al., 2022b) and topic modeling (Schmiedel et al., 2019), have been developed to understand the customer and public opinion expressed in social media (Abbasi et al., 2018). Academics also try to utilize this valuable source of data in their research (Jabr et al., 2020). Amongst text mining techniques, topic modeling has been of particular interest as algorithms such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) have proved to be a powerful tool to cluster documents.

Topic modeling techniques typically use document-level co-occurrence information to group semantically related words into a single cluster or topic. Their objective of these models is to maximize the their objective functions, and they have a tendency to improve extrinsic evaluation metrics (Blei et al., 2003). These models deliberately sacrifice performance on rare topics to achieve a better performance in modeling frequent words. This results in a lop-sided impression of the document corpus (Jagarlamudi et al., 2012), which leads

to suboptimal performance in domain-specific tasks. More importantly, the identified topics don't necessarily represent explainable and meaningful themes or constructs (Gholizadeh et al., 2021; Palese and Usai, 2018). Interactive topic modeling techniques are used as alternative to unsupervised ones as they can tackle the above issues by developing topic clusters based on the initial seed words. As the performance of these interactive techniques heavily depends on the initial seed words, our study asks, "how can text features be used to generate seed words in developing interactive topic models"?

We answer this question by developing a method for iterative seed word generation. We provide initial seed words for interactive topic modeling (Hu et al., 2014) through qualitative content analysis. Then through an iterative seed word generation, our developed algorithm updates seed word vectors from the corpus by considering document similarity. Our mixed-method iterative seed word generation, combined with interactive topic modeling, helps develop a probability vector of each document in the corpus, indicating their relevance to study constructs.

To test our proposed method's validity and applicability in practice, we investigate the critical phenomenon of the post-adoption use of Contact Tracing Mobile Applications (CTMAs) during the COVID-19 pandemic. Understanding user satisfaction towards CTMAs will help improve this technology, which further helps healthcare workers and public health authorities gather more data to understand the spread of the virus and slow down the rate of infections and hospitalizations during future pandemics. We collect and analyze a panel data set of 8,813 customer online reviews from March 2020 to June 2021. The results show the superiority of our proposed method in identifying the primary research constructs, *utilitarian benefits*, *installation issues*, *use issues* and *privacy concerns*. Our work focuses on the textual elements of online reviews and offers two contributions. First, we demonstrate that the proposed iterative seed word generation method helps interactive topic modeling algorithms extract the research dimensions or constructs hidden in online reviews. Second, we discover latent thematic structures that appropriately measure public satisfaction toward CTMAs. We show that the different dimensions have varying impacts on the overall user satisfaction.

To the best of our knowledge, this is the first study to combine advanced text mining techniques with qualitative methods of analyzing data to feed a topic modeling algorithm and extract context-specific dimensions from socialized data. Using these advanced techniques, we point to systematic differences between various constructs hidden in unstructured online data in the context of interest.

The remainder of this paper is structured as follows. The next section describes the proposed method for iterative seed word generation. The case of Australia's CTMA (COVIDSafe) is subsequently introduced. We then demonstrate the application of the proposed method and test its validity and applicability. The last section elaborates on the contribution of our research and its limitations and offers several suggestions for further research.

Proposed Method

Figure 1 shows the proposed method for developing the primary research constructs based on iterative seed word generation in interactive topic modeling. In the following, we elaborate on each step of the proposed approach.

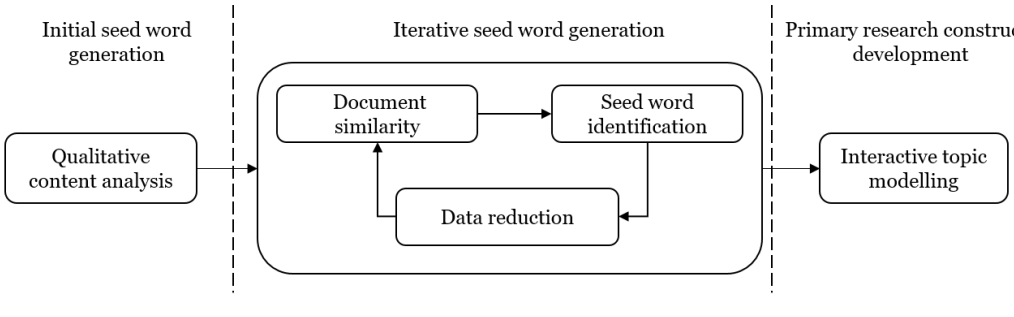

Figure 1. The Proposed Method to Develop Primary Research Constructs

Qualitative Content Analysis

We first identify the initial seed words through qualitative content analysis. In contrast with existing methods for human-in-the-loop (Grønsund and Aanestad, 2020), such as document labelling (Fügener et al., 2021), we apply qualitative content analysis to first identify the emerged themes and their first-order constructs (preliminary codes). The approach is based on analytic induction (Lapointe and Rivard, 2005), where "researchers develop hypotheses, sometimes rough and general approximations, prior to entry into the field or, in cases where data already are collected, prior to data analysis" (Gilgun, 1995, pp. 268–269). With our general hypotheses and constructs in mind, we iterate between the textual data and relevant post-acceptance constructs until we reach a point of theoretical saturation where new data were not adding significantly to our existing analysis (Eisenhardt, 1989). There can be different approaches to this qualitative analysis. One or more qualitative analysts can generate the initial codes, and then the whole team can discuss the emerging codes and reach a consensus. Alternatively, the team can perform this step collectively and rely on their readings of a sample of the data and tools such as word frequencies and word clouds to identify the most relevant seed words for each of the constructs (this was the approach adopted for this paper). At the end of this step, the initial word list of each primary research construct is created. Then a vector, including the relevant seed words, is developed for each construct.

Iterative Seed Word Generation

In this step, based on the initial seed word vectors, we iteratively generate more seed words. The objective is to identify more seed words that have not been discovered in qualitative content analysis. This is done through text processing and determining document similarity. We conduct text pre-processing and word tokenizing and generate TF-IDF (term frequency-inverse document frequency) of a corpus consisting of all the documents and seed word vectors. We then iterate on the following steps for each document:

a) we measure the cosine similarity (Namvar and Chua, 2022) of the document against each seed word vector to produce a vector, indicating document similarity percentage with each of the seed word vectors.

b) we calculate the maximum document similarity to each seed word vectors.

c) using this maximum, we temporarily relate each document to the most similar construct (or seed word vector). The document is passed to the next step if the its maximum similarity to the most related construct is above the minimum similarity percentage. Otherwise, the next document would be analyzed starting from step a.

d) we create a new list of the top frequent words for the document.

e) we compare the new word list to the relevant seed word vector, and if any word doesn't already exist, we update the seed word vector with newly added words.

This above process is repeated for all the documents in the corpus. If these documents show similarity to one of the seed words (based on the cut-off threshold for cosine similarities detection), the frequent words of the document are added to the relevant seed word vectors. Same as other text processing techniques, the cut-off threshold can be tuned by the researcher based on the study context and data to achieve the optimum results. The iteration is stopped when all the documents are analyzed, and their possible frequent words are added to the seed word vectors.

Interactive Topic Modeling

In contrast to unsupervised topic modeling, we train the topic modeling algorithm with the generated seed word vectors. This interactive (or semi-supervised) topic modeling (Hu et al., 2014) results in a vector for each document which indicates the document's relevance to the identified topics or the research constructs. The number of topics is determined based on the number of seed word vectors generated. Thes seed word vectors are identified based on the primary research constructs of interest for a particular study.

In the next section, we assess the validity and applicability of the proposed method by investigating the critical phenomenon of CTMAs post-adoption usage during a time of the pandemic.

Case Study: COVIDSafe

The nature and scale of the COVID-19 pandemic make digital tracing necessary for the mitigation of one of the most widespread pandemics the world has ever seen. Several researchers and technology vendors have proposed CTMAs (Hassandoust et al., 2021), such as TraceTogether in Singapore and COVIDSafe in Australia (Lockey et al., 2021). These apps typically use location-based data and Bluetooth to track people and provide healthcare workers and government officials with the necessary data to understand the spread of the virus and to help in slowing down the rate of infections and hospitalizations.

Hypothesis Development

Grounded in our empirical data and emergent findings, we used analytic induction and searched for appropriate theories (Gioia et al., 2013; Urquhart and Fernández, 2016). Based on the post-acceptance model of information systems (IS) continuance derived from expectation-confirmation theory (Bhattacherjee, 2001), we identified three primary constructs, namely *utilitarian benefits*, *installation issues* and *use issues*. The construct of privacy concerns was not in our a-priori model but emerged through analytic induction. However, for the sake of clarity, we employed the more traditional presentational strategy of providing all the hypotheses first before explaining the findings.

Perceived usefulness is the degree to which organizational members believe using a specific system improves their job performance (Lim and Benbasat, 2000). In IS domain, perceived usefulness refers to the expected benefit from the use of IS technologies (Bhattacherjee, 2001). In our study of CTMAs, perceived usefulness is captured by the utilitarian benefits expressed by users. The expectation-confirmation model posits that the perceived usefulness of IS systems has an impact on user satisfaction. Indeed, as the need-satisfaction perspective (Bhattacherjee and Lin, 2015) indicates, the realization of user needs is a crucial antecedent of satisfaction. Considering this perspective, the utilitarian benefit of IS systems can be linked with user satisfaction. We therefore propose:

H1: Utilitarian benefits have a positive relationship with user satisfaction with CTMAs.

According to the expectation-confirmation model, users' confirmation of expectations determines their satisfaction (Bhattacherjee, 2001) with IS systems. This model defines confirmation as users' perception of the comparison between the expectation of IS use and its performance. After interacting with and using IS, expectations are formed (Bhattacherjee and Lin, 2015). These expectations consequently shape the levels of satisfaction. In our study, confirmation is developed by measuring installation and use issues. It is reasonable that users expect CTMAs not to come with installation and use issues. Thus, we propose:

H2: Installation issues have a negative relationship with user satisfaction with CTMAs.
H3: Use issues have a negative relationship with user satisfaction with CTMAs.

Epidemic control requires sufficient use of CTMAs. However, a key concern in CTMAs' adoption is user privacy (Namvar et al., 2022a). This concern determines user decision to share their data through CTMAs and can potentially limit the CTMAs adoption. The impact of privacy concerns on CTMAs' use is reflected in the initial adoption decision and post-adoption use. Users expect assurance of their privacy when using IS, and the loss of their control over personal data affects their satisfaction. This leads us to propose:

H4: Privacy concerns have a negative relationship with user satisfaction with CTMAs.

In the following, we demonstrate how our proposed method can be applied to the available online reviews on COVIDSafe to extract the above research constructs (i.e., utilitarian benefits, installation issues, use issues and privacy concerns) and test their impact on a dependent variable, user satisfaction.

Data Collection

We used Heedzy.com to collect 8,813 reviews of the COVIDSafe app in Australia between March 2020 to June 2021 on the Google Play store. We extracted the following data items for each review: review title, review text, review date, and star rating. User satisfaction was operationalized as the star rating provided

by the users as it indicates the overall judgment of user experience of the quality of a product or service. Numerical reviewer ratings, in many instances, are available in a Likert scale format, ranging from 1 to 5, reflecting the positive, neutral, or negative evaluation of the quality of a product or service (Mudambi and Schuff, 2010). In the current literature, a review rating is considered a valid proxy for identifying a user's overall satisfaction with products and services (Salehan and Kim, 2016).

Findings

Qualitative Content Analysis

Based on the qualitative approach explained earlier, we generated the preliminary codes for each of the four constructs of our research. These codes include the frequent words associated with utilitarian benefits, installation issues, use issues and privacy concerns. Table 1 presents the four primary constructs, along with their codes. These codes are used as the initial seed words for the rest of the analysis and would be fed to the algorithm for iterative seed word generation. Representative quotes illustrate how these constructs are manifested in this context.

Construct	Initial seed words	Sample representative quotes from qualitative analysis
Utilitarian benefits	contact, safe, protection, check, support, trace, spread, optimize, initiate	"All ok for the Safety of all Australians", "I see this is a way of keeping us all safe"
Installation issues	un/re-install, invalid, fix, code, pin, problem, error, fail, connect, un/disable, verify, delete	"Disappointed that I can't install it on my wife's older phone" "Also, tried to install on my wife's phone and no matter what we tried it kept saying invalid phone number, did everything advised, still the same. Make it work properly and I'll put it back on!!"
Use issues	battery, drain, power, kill, background, Bluetooth, Wi-Fi, screen, power, sync, constant, drop, car, headphone, charge	"Uses bugger all battery, doesn't interfere with other bluetooth devices" "it interrupts the use of my Bluetooth speaker and headphones so I've had to temporarily uninstall it"
Privacy concerns	privacy, track, violate, security, permission, store, requirement, access, legal, trust, hack	"Requires postcode for unknown reasons given it tracks you. Requires Bluetooth location data because app is locating you."

Table 1. Main Constructs and their Initial Seed Words

Seed Word Generation

We used the initial seed words to find evidence of our constructs in the corpus. In an iterative approach, we took the initial seed word list and measured the cosine similarity (Namvar and Chua, 2022) of each review in the corpus to each seed word vector of the four primary study constructs. This created a vector for each review indicating its similarity percentage to each of the four primary constructs. This step allowed us to identify whether a review could be identified as having significant similarity to a construct and created the basis of our iterative approach.

Next, we filtered out the reviews that had not been identified as similar to one of the research constructs, resulting in a smaller dataset of 8,460 reviews (see the first iteration in Table 2). Overall, we had five iterations until none of the reviews showed similarity to the seed word vectors. We set the cut-off threshold for finding cosine similarity to 30% as it showed a better performance in multiple repetitions of similarity detection. However, this threshold can be tuned in a different research context with various data sizes. After completing the iterative process, 1,131 reviews showed similarity to our research constructs and their frequent words were added to the relevant seed word vectors. The remaining 7,682 reviews did not provide significant similarity percentages to the seed word vectors, as the seed word vectors had grown to a sufficient size that any new review would not be similar to all the top words combined. Also, since these

reviews consist of short sentences with an average length of 22 words, it is reasonable that our method does not add any of their words to seed word vectors. This process generated four in-depth seed word vectors that aligned closely with the primary research constructs.

Iteration	0	1	2	3	4
Remaining reviews	8,813	8,460	7,746	7,696	7,685

Table 2. The Remaining Reviews in each Iteration

Constructing the Main Research Constructs

To guide latent topic identification toward more explainable construct-specific topics (Palese and Usai, 2018), we used the final list of seed words to train the interactive topic modeling algorithm using the GuidedLDA Python package. Our interactive topic model defines a topic to be a combination of two multinomial distributions, a seeded topic and a regular topic. The seeded topic distribution is confined to creating words that come from a predefined set (Hu et al., 2014). The predefined seeded topic set for our model is the final seed words developed in the previous step. For the regular topic, the distribution can generate a topic from any word in the corpus vocabulary. Then the guided model will decide whether to choose a word from the seeded topic or regular topic distributions (Palese and Usai, 2018). The model generates a multinomial distribution for both seed and regular topics, and the confidence parameter controls the probability of drawing a word from either seed topic distribution or regular topic distribution. The model is guided to place a higher probability on related words, however, it is not forced (Hu et al., 2014). This still allows the model freedom to associate a review, with a higher probability, to another construct if it sees fit.

We constructed a vector of the main four variables for each review based on the probability that the review is associated with a construct. For example, for a review with the words such as "trust", "track", and "location", the following probability scores would be generated: 0.01% (use issue), 2% (utilitarian benefits), 0.05% (installation issues), and 97.94% (privacy concerns). This review would then have a probability vector of [0.0001, 0.02, 0.0005, 0.9794], signalling that this review is highly probable to be about construct 4 (privacy concerns). This vector was generated for each review in the corpus. Four different probabilities that are assigned to each review work as independent variables of our research.

Validation and Applicability Tests

Table 3 presents the descriptive statistics and correlations for the primary constructs or variables. The highest correlation between installation issues and utilitarian benefits stands at -0.41. As our analysis did not indicate any high correlation, removing any variables was unnecessary.

	Mean	Std	1	2	3	4
1.Utilitarian benefits	0.36	0.42	1.00			
2.Installation issues	0.2146	0.37	-0.41	1.00		
3.Use issues	0.2129	0.35	-0.40	-0.29	1.00	
4.Privacy concerns	0.17	0.30	-0.31	-0.24	-0.15	1.00

Table 3. Variables Correlations (n= 1,131)

We use negative binomial regression to test the hypotheses. Since the dependent variable "rating" is a non-negative count variable that ranges from 1 to 5, it is inappropriate to use standard multiple regression. Considering the mean of each level within our variables are larger than the variance of each level, our predictor variables are dispersed. Therefore, we apply the negative binomial regression model, consistent with prior studies investigating online reviews (Zhou and Guo, 2017).

Predictor variables were mean-centred. We also measured the variance inflation factors (VIFs) for each model to investigate the potential multicollinearity. The VIFs ranged from 1 to 4.8, significantly lower than the generally accepted threshold of 10. The Hausman statistic also showed that endogeneity does not exist.

Table 4 presents the estimation results of the negative binomial models. We only considered reviews 1,131 reviews that had values for at least one of the main four constructs.

a.Proposed interactive topic modeling		b.Unsupervised topic modeling	
Construct		Topic	
Use issues	-2.24 (0.092) ***	Topic 1	-2.75 (0.350) ***
Utilitarian benefits	0.52 (0.089) ***	Topic 2	-0.6435 (0.350)
Installation issues	-2.69 (0.091) ***	Topic 3	-2.4603 (0.349) ***
Privacy concerns	-1.89 (0.095) ***	Topic 4	0.3254 (0.349)
R Squared: 0.43		R Squared: 0.36	
Note: Robust standard errors are reported in parentheses for coefficients. ***p<0.001, **p<0.01, and *<0.05			

Table 4. Regression Analysis Explaining User Satisfaction (n= 1,131)

Robustness Tests

As it is essential in the topic modeling process to determine the optimal number of topics to be extracted for construct validity, we also compared the performance of our model with an unsupervised topic model with an optimum number of topics. These topics show the highest semantic coherence and exclusivity when applied to unsupervised topic modeling (Gholizadeh et al., 2021). We set different values for the number of topics in a model, ranging from 3 to 20. After analyzing the different models with various topic numbers, we set the extracted topic number to 13 since it generated an optimum balance between semantic coherence and exclusivity metrics. The analyses reported in Table 5 show that the performance of unsupervised topics modeling with 13 topics is significantly lower than our proposed model reported in Table 4 (a).

Topic	1	2	3	4	5	6	7	8	9	10	11	12	13
	-0.5 ***	2.41	2.41 ***	-0.13	2.98 ***	2.60 ***	0.62	1.36 *	-0.43	1.13	2.22 **	2.82 ***	2.93 ***
R Squared: 0.36 ***p<0.001, **p<0.01, and *<0.05													

Table 5. Negative Binomial Regression with Optimal Number of Topics (n= 1,131)

As another approach to the robustness test, we compare the performance of our model when its developed constructs are used for classification and numeric prediction. We used only the four developed constructs to train a Support Vector Machine (SVM) and an Artificial Neural Network (ANN). Using Python Sci-kit Learn packages, we developed the target vector, including *"user rating",* and a feature matrix, including utilitarian benefits, installation issues, use issues and privacy concerns. For SVM, we set the kernel function as Polynomial, and for ANN, we use a multi-layer perceptron (MLP) with the size of (25, 3). We evaluated the models using 5-fold cross-validation. We computed the classic classification evaluation metric, accuracy, and classic regression evaluation metric, RMSE. We compared the results with unsupervised topic modeling (with 4 and 13 topics). The presented results in Table 6 indicate that our proposed models outperform the existing models for predicting the user satisfaction of CMTAs.

	Classification (accuracy)		Prediction (RMSE)	
	SVM	ANN	SVM	ANN
Proposed interactive topic modeling	0.69	0.69	2.44	2.11
Unsupervised topic modeling (4 topics)	0.66	0.66	2.86	2.41
Unsupervised topic modeling (13 topics)	0.21	0.57	3.97	3.07

Table 6. Predictive Performance

Concluding Remarks and Future Research

Online review platforms provide open, convenient communication channels for sharing and gathering public opinion. For governments and public health authorities, these online reviews represent unique and valuable information sources to understand the public perceptions of their (digital) services. There have been earlier attempts to use topic modeling techniques for understanding public opinion. However, as reported in those studies (e.g., Gholizadeh et al., 2021; Palese and Usai, 2018), traditional approaches to topic modeling cannot effectively identify the research dimensions or constructs, identify the users' specific issues or, indeed, determine the prevalence of those issues across the datasets.

Our study demonstrated a novel approach for going beyond the existing topic modeling methods. We applied a mixed-method analysis (a sequence of qualitative content analysis, iterative seed word generation and interactive topic modeling). We showed the application of the proposed method, especially when used in time-sensitive contexts (e.g., the COVID-19 pandemic), to provide a rich and evidence-based understanding of users' attitudes and concerns. We highlighted the statistically significant relationships between these attitudes and concerns (as expressed in textual comments) and the users' overall satisfaction with a CTMA.

Our study contributes to theory by proposing a) a method of developing constructs from unstructured data (text) and b) a novel approach to model development and theory testing. In particular, while the existing literature on the expectation-confirmation model generally adopts a survey-based approach, our study examines this model using organic user review data and is inherently free from certain biases such as mono-method bias and hypothesis guessing (Burton-Jones, 2009). In addition, using analytic induction, our study contextualized the post-acceptance model of IS continuance (Bhattacherjee 2001) by adding the construct of privacy concerns that emerged from CTMAs' user reviews.

Our study also contributes to practice by a) building a rich, timely, and evidence-based understanding of users' attitudes and concerns and b) developing more effective technologies and better public relations strategies. Both our methodological approach (in unearthing users' attitudes, issues, and concerns) and findings can inform the design and deployment of the next generation of CTMAs for managing future pandemics.

This study, nevertheless, has certain limitations that can be explored in future research. The data for this study are reviews of a particular application. It consists of the reviews from users who were willing to embrace technology and share their feedback. Understanding the applicability and validity of the proposed method for more generic users requires a different study with possibly a different research context and data collection approach. To generalize the results, future research can apply the proposed method to investigate other sources of textual data (e.g., tweets) with different underlying architectures. Finally, in this study, we labeled only 1,131 reviews out of 8,813 reviews. We plan to continue this research by first enhancing our method to take more reviews into consideration. Our current algorithm did not assign 87% of the data to any of the four constructs due to the nature of the data, which consists of short reviews with an average length of 22 words. In future studies, we will apply our proposed method to other sources of unstructured data in other domains with more in-depth and lengthier documents to examine the percentage of labeled documents and the performance of the proposed method for iterative seed word generation.

References

Abbasi, A., Zhou, Y., Deng, S., Zhang, P., 2018. Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective. MIS Quarterly 42, 427-A38.

Bhattacherjee, A., 2001. Understanding information systems continuance: An expectation-confirmation model. MIS quarterly 351–370.

Bhattacherjee, A., Lin, C.-P., 2015. A unified model of IT continuance: three complementary perspectives and crossover effects. European Journal of Information Systems 24, 364–373.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research 3, 993–1022.

Burton-Jones, A., 2009. Minimizing method bias through programmatic research. MIS quarterly 33, 445–471.

Deng, S., Huang, Z.J., Sinha, A.P., Zhao, H., 2018. The interaction between microblog sentiment and stock return: An empirical examination. MIS quarterly 42, 895–918.
Eisenhardt, K.M., 1989. Making fast strategic decisions in high-velocity environments. Academy of Management journal 32, 543–576.
Fügener, A., Grahl, J., Gupta, A., Ketter, W., 2021. Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. MIS Quarterly 45, 1527–1556.
Gholizadeh, M., Akhlaghpour, S., Isaias, P., Namvar, M., 2021. Emergent affordances and potential challenges of mobile learning apps: insights from online reviews. Information Technology & People 0959–3845. https://doi.org/10.1108/ITP-05-2021-0412
Gilgun, J.F., 1995. We shared something special: The moral discourse of incest perpetrators. Journal of Marriage and the Family 57, 265–281.
Gioia, D.A., Corley, K.G., Hamilton, A.L., 2013. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organizational Research Methods 16, 15–31.
Grønsund, T., Aanestad, M., 2020. Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems 29, 101614.
Hassandoust, F., Akhlaghpour, S., Johnston, A.C., 2021. Individuals' privacy concerns and adoption of contact tracing mobile applications in a pandemic: A situational privacy calculus perspective. Journal of the American Medical Informatics Association 28, 463–471.
Hu, Y., Boyd-Graber, J., Satinoff, B., Smith, A., 2014. Interactive topic modeling. Machine learning 95, 423–469.
Jabr, W., Liu, B., Yin, D., Zhang, H., 2020. Online Word-of-Mouth. MIS Quarterly Research Curations.
Jagarlamudi, J., Daumé III, H., Udupa, R., 2012. Incorporating lexical priors into topic models, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. pp. 204–213.
Lapointe, L., Rivard, S., 2005. A multilevel model of resistance to information technology implementation. MIS quarterly 29, 461–491.
Lim, K.H., Benbasat, I., 2000. The effect of multimedia on perceived equivocality and perceived usefulness of information systems. MIS quarterly 24, 449–471.
Lockey, S., Edwards, M.R., Hornsey, M.J., Gillespie, N., Akhlaghpour, S., Colville, S., 2021. Profiling adopters (and non-adopters) of a contact tracing mobile application: Insights from Australia. International Journal of Medical Informatics 149, 104414.
Mudambi, S.M., Schuff, D., 2010. Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. MIS quarterly 34, 185–200.
Namvar, M., Akhlaghpour, S., Pool, J., Priscilia, A., 2022a. Towards explaining user satisfaction with contact tracing mobile applications in a time of pandemic: a text analytics approach. Presented at the 54th Hawaii International Conference on System Sciences (HICSS), Hawaii, USA.
Namvar, M., Chua, A., 2022. The impact of context clues on online review helpfulness. Internet Research ahead-of-print. https://doi.org/10.1108/INTR-02-2021-0093
Namvar, M., Intezari, A., Akhlaghpour, S., Brienza, J.P., 2022b. Beyond effective use: Integrating wise reasoning in machine learning development. International Journal of Information Management ahead-of-print, 102566. https://doi.org/10.1016/j.ijinfomgt.2022.102566
Padmanabhan, B., Sahoo, N., Burton-Jones, A., 2022. Machine Learning in Information Systems Research. MISQ 46, iii–xix.
Palese, B., Usai, A., 2018. The relative importance of service quality dimensions in E-commerce experiences. International Journal of Information Management 40, 132–140.
Salehan, M., Kim, D.J., 2016. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems 81, 30–40.
Schmiedel, T., Müller, O., vom Brocke, J., 2019. Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. Organizational Research Methods 22, 941–968.
Urquhart, C., Fernández, W., 2016. Using Grounded Theory Method in Information Systems: The Researcher as Blank Slate and Other Myths, in: Willcocks, L.P., Sauer, C., Lacity, M.C. (Eds.), Enacting Research Methods in Information Systems: Volume 1. Springer International Publishing, Cham, pp. 129–156.
Zhou, S., Guo, B., 2017. The order effect on online review helpfulness: A social influence perspective. Decision Support Systems 93, 77–87.