Developing a Method for Identifying Instances of Group Generative Interactions in Enterprise Social Media

Research in Progress

Elisavet Averkiadi
Michigan State University
averkiad@msu.edu

Wietske Van Osch
HEC Montreal
Michigan State University
wietske.van-osch@hec.ca

Yuyang Liang
Michigan State University
liangyuy@msu.edu

ABSTRACT

Companies hold particular interest in group generative interactions - the conception of novel ideas and solutions through group exchanges. They are a root-cause of innovation and thus are important to companies' survival. Enterprise Social Media (ESM) offer a unique opportunity to study generative group interactions, due to the transparent nature of activities on these platforms. In this research-in-progress paper, we conduct a preliminary analysis to develop a method that could identify the instances of ESM-based generative group interactions, where we focus on distinguishing generative versus non-generative group interactions. To do this, we used the text from all group interactions from an ESM platform of a multinational organization. We implemented machine learning models to learn and classify the text as generative or non-generative. As a result, we produced the top important term features from the best performing model. These features will help us understand the nature of discussions that occur in these interactions in future studies.

Keywords

ESM, Generative Interactions, Group Interactions, Enterprise Social Media, Team Behaviors.

INTRODUCTION

Enterprise Social Media (ESM) are web-based applications utilized for internal corporate business processes and objectives, by offering users various features that enable them to communicate, network, organize, leverage information capital, among other activities. ESM are equipped with a compatible set of affordances (Leonardi, Huysman, & Steinfield, 2013) for collaborations to occur and have the potential to foster *group generative interactions* – the creation of novel ideas and solutions through group exchanges (Avital & Te'eni, 2009). They offer a unique opportunity for researchers interested in studying team behaviors as one of ESM's unique affordances, visibility, allows all contributions to the platform to become visible and therefore available for analysis. The ability to observe how people collaborate is not only an opportunity to improve our theoretical understanding of the nature of group interactions in ESM, but also a chance to improve such interactions in the context of ESM platforms and possibly beyond. Group generative interactions, as the root-cause of innovation (Avital & Te'eni, 2009), are a class of collaboration that are of particular interest to companies, given the direct link between a company's ability to innovate and their chances to survive and thrive (Abernathy et al., 1985; Hambrick, 1983; Henderson et al., 1990; Lieberman et al., 2001; Tushman et al., 1986).

In this research-in-progress paper, we will present the results from a preliminary analysis geared toward developing a method that could identify instances of ESM-based generative group interactions, using machine-learning to classify content data from an ESM platform utilized in a multinational organization. For this preliminary analysis we took a sub-sample of 1% of all group interactions[1]. Although our ultimate aim is to develop an algorithm that not only identifies when generative interactions occur, but could help distinguish different forms of such interactions, given the small sample used for this preliminary analysis, we focus on differentiating generative versus non-generative group interactions as a first step in identifying the extent to which ESM-based interactions embody generativity. Beyond the theoretical insights produced by understanding the extent to which generative interactions occur in ESM, the resulting business analytics tool also allows managers to

[1] Ultimately the goal is to use a 10% training sample

understand the level and nature of engagement with ESM in their organization as it pertains to generative activity, and hence creativity and innovation.

THEORETICAL BACKGROUND

Generativity and Generative Collaboration

Generativity is the ability to create, originate, or produce (Avital et al., 2009; Webster, 2009). In ESM-based groups, group generative interactions will occur as group members jointly engage in the creation, origination, and production of ideas (Van Osch & Avital, 2010). As generative interactions are a root-cause of innovation, focusing on such interactions will allow us to investigate a critical precursor to innovations within organizations (Avital, et al, 2009; Lovelace, Shapiro, & Wiengart, 2001; Van Osch & Avital, 2010).

Tsoukas (2009) explores the various ways of producing novel conceptualizations and identifies different forms of group generative interactions that can be inferred from creative cognition research (Dunbar, 1997; Frinke, Ward, & Smith, 1992), namely combination, expansion, and reframing. Generativity can thus stem from *combining* already existing concepts in new ways (Wisniewski, 1997), expanding the use of an existing concept from its core use to match a new situation (i.e., expansion), or by creatively deconstructing an existing concept and reconstructing it to fit a new situation (i.e., reframing) (Bartunek & Franzak, 1988; Bateson, 1972; Van Osch & Avital, 2010; Watzlawick, Weakland, & Fisch, 1974) often by challenging the status quo (Van Osch et al., 2010).

As aforementioned, although our ultimate goal is to develop unobtrusive methods for classifying these distinct forms of generative interactions, in this research-in-progress paper we focus on developing a tool for differentiating generative versus non-generative group interactions.

Enterprise Social Media and Generative Interactions

Given the existing research highlighting the relevance of ESM to knowledge-focused interactions, there is reason to assume that ESM satisfy the criteria for supporting team-level group generative interactions (Leonardi, 2014; Van Osch and Steinfield, 2013, 2018). ESM afford the identification of relevant information and relevant individuals, or contributors. They create a wider awareness and broaden contribution in creative processes and may also spark the creative disruptions of existing concepts and work practices, as necessary, to support the creation of novel concepts. Still, there is a lack of information, in studies so far, on how these benefits may occur and the ways in which ESM can aid group generative interactions for larger groups. Given the lack of guidance for using such a tool to its full potential, it is of paramount importance to examine instances of group generative interactions in ESM to ultimately investigate ways of leveraging technology design concepts that could nudge teams towards such interactions.

DATA AND METHODS

Data

For this study to be conducted, data is provided from an ESM tool utilized by an organization that conducts research and consulting in the domain of human-computer interaction. More specifically, the organization focuses on building technology and furnishing products for a variety of clients, from corporate offices, healthcare, educational institutions, and government institutions. The case organization has over 11,000 employees, with over 80 locations around the world[2]. The ESM tool used by the case organization was launched with the purpose of supporting business connections, communications, and collaborations among employees across the organization. At the time of the data collection for this study, a stable base of 10,000 users had accumulated over five years since the launch of this tool. Of these 10,000 employees who use the tool, 91% (9,000 users) are members of groups, who participate in group discussions and activities. The primary reason for using this data is to have a relevant object of study to reach our objective of developing a model for identifying the instances, in the text from these group interactions, that pertain to generative activity.

The data has 20,000 threads, and 219 (~1%) of these were used to develop the machine learning model.

Methods

Data Preparation

To develop a training set for the machine learning model, the data was labelled with a code for the presence or absence of generative activity in the text. The coding scheme can be seen in Table 1. As aforementioned, ESM-based generative interactions belong to three different types, namely combination, expansion, and reframing, as defined previously.

Generativity Type	Code	Description
Generative Activity	1	Used either conceptual reframing, expansion, or combination.
Non-Generative Activity	0	No use of conceptual expansion, reframing, or combination.

Table 1: Generativity Coding Scheme

Human coders were trained to identify posts (embedded within threads to maintain the context of the conversation)

[2] Including the Americas, Europe, Asia, Africa, and Australia.

containing elements of one of the three types of generative interactions using a coding manual that provided definitions and examples of each of the types. Text that did not include these types was labelled as non-generative (coded as 0). Given the small sub-sample used for this preliminary analysis, the three generativity types were collapsed into a single category.

In order for the data to be processable by the machine learning models, it was pre-processed[3]. The text was then split into single terms, and lemmatized – a technique that reduces a word to its base form. Next, we extracted features from the text using the 'Bag of Words' method – a representation of the text that describes the occurrence of words in the data, by the number of times they appear. TF-IDF (Term Frequency – Inverse Document Frequency) was used to count the occurrence of the words in the data, and then used to vectorize the text. The text is subsequently in a numerical vector form.

Model Implementation

Following feature extraction, we implemented several machine learning models, and compared their performance at classifying the text to find the best one. The data was split into training and testing sets, where the training set included 175 messages and the testing set included 44. The models selected for this included: Random Forest, AdaBoost (Adaptive Boosting), Naïve Bayes (Multinomial), Support-Vector Machine (SVM), and Logistic Regression.

The data included many non-generative interactions which hindered the learning of the models. To improve learning for the models, we implemented oversampling. This is a technique that adjusts the distribution of the two classes (generative activity vs. non-generative activity), and thus improves the imbalanced dataset. The performance measures that were used to compare the machine learning models were f-1 score, accuracy, and Area Under the Curve (AUC).

RESULTS

Table 2 shows the performance of the models. Random Forest performed better than other models implemented. From its performance, we can infer that it accurately predicted the percentage of interactions that were generative and non-generative. AdaBoost performed almost as well as Random Forest, while compared to all models the Naïve Bayes model shows underwhelming results. In the data 28% of observations were generative interactions, and 72% of observations were non-generative

The results in Table 2 indicate that the Random Forest model performed well at classifying the generative and non-generative interactions in our data; with 76% accuracy at correctly classifying the data. AdaBoost had a closer accuracy score, of 71%, compared to the rest of the models; Naïve Bayes scored lowest, with 44% accuracy at correctly classifying the data. While a 7% difference in performance between the Random Forest and AdaBoost models may be minimal, Random Forest still had a higher AUC score – where the difference in performance between the two models was 10%. More specifically, according to the f-1 score for Random Forest, it outperforms the other models at correctly classifying the data. Our results show that Random Forest accurately classified the instances of generative activity in the text with 67% correct classification, and non-generative activity with 90% correct classification (detailed performance of f-1 score can be seen in Table 3). This is significant because there were indeed more instances of non-generative activity in the text for the model to learn from, which is reflected in the higher correct classification percentage.

Model	AUC	Accuracy	f-1
Random Forest	0.80	0.76	0.83
AdaBoost	0.70	0.71	0.81
Naïve Bayes	0.59	0.44	0.53
SVM	0.67	0.69	0.78
Logistic Regression	0.72	0.66	0.72

Table 2: Model Performance

Model	f-1	
	0	1
Random Forest	0.90	0.67
ADA Boost	0.88	0.67
Naïve Bayes	0.55	0.49
SVM	0.83	0.64
Logistic Regression	0.76	0.61

Table 3: Model Performance f-1 Score

We then used the Random Forest model to produce the top 20 important features in the text data (See Table 4). These terms are important to the model for identifying instances of generative activity. The work-related terms (such as 'value', 'product', 'leader', 'project', and others) are important for distinguishing between the two categories (generative and non-generative activity). These terms will help improve our understanding of the nature of the discussions that occur in these interactions in our future

[3] Including removing punctuation, stop words, and converting text to its lower-case format.

studies; it will additionally help us to develop the labeling scheme for future research. Table 5 shows examples of generative and non-generative interaction data.

Terms	Score
like	0.0601
work	0.0403
people	0.0313
way	0.0268
one	0.0254
new	0.0214
value	0.0194
product	0.0183
business	0.0181
take	0.0179
time	0.0172
hi	0.0171
place	0.0167
today	0.0162
different	0.0158
need	0.0158
feel	0.0144
right	0.0144
leader	0.0144
project	0.0143

Table 4: Top 20 Important Features

DISCUSSION

The unique opportunity that is offered by ESM to study technology-based group interactions is one that should be leveraged. The continuously generated numerous traces of team behaviors make ESM a suitable context to further improve studies on this topic and provide a reliable data source for future studies on computer-mediated communication and group generative interactions.

The results of this pilot study have shown that the Random Forest classifier handles the text data better than the other models implemented. The top important features showcase an essential set of language that will be helpful to distinguish generative activity in ESM-based group interactions.

The findings from our preliminary analysis show that about 28% of all group interactions occurring through ESM contain elements of generativity and therefore are potentially a root cause of innovation for the organization. Hence, these findings show that there is potential to and merit in developing a method for differentiating the types of group generative interactions that occur by using a 10% sub sample of the original data corpus. Doing so would allow us to not only theorize the nature of generative interactions occurring through ESM, but also develop theoretical models of the precursors, both at the group level but also in terms of ESM affordances, that result in distinct types of ESM-based generative interactions. For instance, the ways in which groups interact with each other and with the ESM in the context of these interactions might be different when groups are engaged in combination, expansion, or reframing. Such insights are theoretically important to obtain holistic understandings of the boundary conditions for different types of generative interactions as well as practically important to provide managers guidance for eliciting different types of generative interactions in an attempt to productive uses of ESM. Hereto, more data will have to be labelled for this direction, and further trial of machine learning algorithms will be needed to produce a accurate classifier for multiple categories of generative activity.

Category	Distribution	Example
Generative Interaction	28%	“Two factors which I feel either hinder or help engagement, are ‘autonomy’ which Rob & Bob mentioned, and ‘change’ or ‘impact’. There is a management concept of ‘leading with a light touch’, people want to understand the limits, the outside boundaries of the work they are asked to do.”
Non-Generative Interaction	72%	“The Steelcase interns had the opportunity to participate in Chicago yesterday. It was great to see the Steelcase show so full and have such an exciting buzz around it.”

Table 5: Examples of Generative and Non-Generative Interactions

REFERENCES

1. Abernathy, W. J., and Clark, K. B. 1985. Innovation: Mapping the winds of creative destruction. Research policy, 14(1), 3-22.

2. Avital, M., and Te'eni, D. 2009. From generative fit to generative capacity: exploring an emerging dimension of information systems design and task performance. Information Systems Journal 19(4), 345-367.
3. Bartunek, J. M., and Franzak, F. J. 1988. The effects of organizational restructuring on frames of reference and cooperation. Journal of Management, 14(4), 579-592.
4. Bateson, G. 1972. Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology. University of Chicago Press.
5. Dunbar, K. 1997. How scientists think: On-line creativity and conceptual change in science. T. N. Ward, S. M. Smith, J. Vaid, eds. Creative Thought. American Psychological Association, Washington, DC, 461–494.
6. Finke R.A., Ward, T.B., and Smith, S.M. 1992 (1996). Creative Cognition: Theory, Research and Applications. The MIT Press.
7. Hambrick, D. C. 1983. Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. Academy of Management journal, 26(1), 5-26.
8. Henderson, R. M., and Clark, K. B. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative science quarterly, 9-30.
9. Leonardi, P. M., Huysman, M., and Steinfield, C. 2013. Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. Journal of Computer Mediated Communication (19:1), 1-19.
10. Leonardi, P.M. 2014. Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. Information Systems Research 25(4), 796-816.
11. Lieberman, M.B., and Montgomery, D.B. 1988. First-Mover Advantages. Strategic Management Journal 9(1), 41–58.
12. Lovelace, K., Shapiro, D. L., and Weingart, L. R. 2001. Maximizing cross-functional new product teams' innovativeness and constraint adherence: A conflict communications perspective. Academy of management journal, 44(4), 779-793.
13. Tsoukas, H. 2009. A dialogical approach to the creation of new knowledge in organizations. Organization Science 20(6), 941-957.
14. Tushman, M. L., and Anderson, P. 1986. Technological discontinuities and organizational environments. Administrative science quarterly, 439-465.
15. Van Osch, W., and Avital, M. 2010. Generative Collectives. ICIS 2010 Proceedings, http://aisel.aisnet.org/icis2010_submissions/175.
16. Van Osch, W. 2012. Generative Collectives. Ipskamp Publishers, Netherlands.
17. Van Osch, W., and Steinfield, C.W. 2013. Boundary Spanning through Enterprise Social Software: An External Stakeholder Perspective. Proceedings of the International Conference on Information Systems (ICIS), Milan, Italy.
18. Van Osch, W., and Steinfield, C. W. (2018). Strategic visibility in enterprise social media: Implications for network formation and boundary spanning. *Journal of Management Information Systems*, *35*(2), 647-682.
19. Van Osch, W., Steinfield, C.W., and Balogh, B.A. 2015. Enterprise Social Media: Challenges and Opportunities for Organizational Communication and Collaboration. Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS), Kauai, Hawaii, USA, January 5-8, 2015.
20. Watzlawick, P., J. Weakland, and R. Fisch. 1974. Change. W. W. Norton, New York.
21. Webster-M. 2009. Generativity. In Merriam-Webster Online Dictionary.
22. Wisniewski, E. J. 1997. When concepts combine. Psychonomic Bulletin and Review, 4(2), 167-183

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1749018. Any opinions, findings, and conclusions or recommendations ex- pressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.