Transactions on Replication Research

Open Data: Evaluating Topic Modeling Interpretability Using Topic Labeled Gold Standard Sets

Biagio Palese

Operations Management and Information Systems, Northern Illinois University, DeKalb, IL, USA

bpalese@niu.edu

Gabriele Piccoli

Stephenson Department of Entrepreneurship & Information Systems, Louisiana State University, Baton Rouge, LA, USA and University of Pavia, Pavia, Italy

gpiccoli@lsu.edu

Abstract:

This paper provides the gold standard sets described and used by Palese and Piccoli (2020) which advance a scalable method, Human Interpretable Topics (HIT), for assessing the interpretability of topic modeling results. Gold standard set are a small collection of documents extracted from a dataset and manually labeled by humans. The role of such a gold standard dataset is to establish a "ground truth" against which researchers can benchmark the results generated by algorithmic topic models. We first provide a detailed description of the classification procedure used to create the labeled sets. Then, we make available descriptive statistics of the three different gold standard sets named respectively: inclusive, full agreement and partial agreement. These gold standard sets can be used to benchmark different/new models built in research analyzing online customers' reviews in the context of the lodging industry. We hope a larger number of researchers will follow our example and use the *AIS Transactions on Replication Research* journal to share open access gold standard sets in different areas of interest.

Keywords: human interpretable topics, gold standard set, open data text mining, topic evaluation, topic interpretability measure, topic modeling.

1 Gold Standard Set Classification Procedure

The classification procedure is composed of two distinct parts: In the first part, we provide descriptions and examples of the seven different topics that we want to identify. In the second part, using them, the human raters will need to classify a list of documents (available in Excel).

The classification is completed when the human raters have assigned the topics to all the documents in the list. Keep in mind that each document can contain a different number of topics. It is the human raters' job to report all the topics that according to them are discussed in each document.

1.1 Part 1

To complete this procedure, it is very important to understand each topic. Please, read the topics descriptions and examples below very carefully before starting the classification of the documents available in the Excel file.

•**Service**: Comments that describe the service provided by the hotel. It includes comments that describe the performance of the hotel staff in helping guests with any element of their stay (e.g., checkin, checkout, problem resolution, etc.).

Examples: 1) "The service at this hotel is very good, and they are very polite and friendly." 2) "We will strengthen the personnel training to provide the impressive service for all guests." 3) "Checkin is quick and the staff friendly."

•**Value**: Comments that refer to the economic value proposition of the hotel. It includes comments that describe guests' perceptions of the experience they received for the price they paid. It also contains general assessment of the stay without including the price.

Examples: 1) "The price-value is excellent." 2) "I think this hotel is fantastic value for money and would recommend to anyone." 3) "Totally worth it for not much more money." 4) "Overall, everything was good."

•**Location**: Comments that describe the hotel's location and its surroundings, including the view from the hotel or local attractions.

Examples: 1) "The location is great as it is close to two subway stations and is located in the nice area of the city." 2) "Close to restaurants and bars." 3) "Huge shopping malls are within walking distant."

•**Room**: Comments that relate to physical aspects of the room, room amenities and room areas (e.g., the bathroom).

Examples: 1) "The rooms are spacious but noisy." 2) "LED TV and a comfortable bathroom." 3) "The bed is custom made by Simmons and is very comfortable."

•**Food**: Comments that describe food or drinks served in the hotel and its restaurants, including the quality of the breakfast buffet or the quality of the food/drink delivered to the room.

Examples: 1) "They only had 2 kinds of dressings for salad." 2) "The food was surprisingly delicious for breakfast and dinner." 3) "The Chinese restaurant in the hotel is also wonderful."

•**Greetings and salutations**: Comments that contain polite words or signs of welcome or recognition. These are often included in hotel responses, but they could also be in reviews (e.g., Dear GM of the Royal Hotel…).

Examples: 1) "Best regards, General Manager" 2) "Thank you for your continued support." 3) "With warm regards, Hotel."

•**Hotel Amenities**: Comments that relate to hotel facilities not in the room. For example, the gym, the pool, the spa, etc.

Examples: 1) "Pool: There is an outdoor pool on the 5th floor with ropes to swim laps." 2) "Gym: This is one of the best hotel gyms ever." 3) "There is a yoga/pilates/spin studio." 4) "WIFI is not free"…

•**Unclassifiable**: It includes comments that you are not able to confidently put in any other category.

Examples: 1) "I was very happy!" 2) "As usual, staying at here is like coming to a 2nd home." 3) "All the wonderful things I've read in this forum are all true."

1.2   Part 2

Keep in mind the above descriptions and examples while completing the classification in the Excel named "goldSetRevRes". If you think a comment contains more than one topic, you should report all of them in order of dominance. However, remember to insert only one topic per column. For example, we have a comment that includes 4 different topics: topic 1 = value; topic 2= service; topic 3= location; topic 4= unclassifiable. When you have classified all the documents, please save the file as "yourfullname_goldSetRevRes" (e.g. joshadams_goldSetRevRes) and email it to the lead researcher. You have now completed the classification. Thank you.

2   Gold Standard Sets Descriptive Statistics

Table 1. Gold standard sets descriptive			
	Inclusive set	Full agreement set	Partial agreement set
Total number of documents	500	280	499
Total number of topics	1,695	790	1,363
Average number of topics per document	3.39	2.82	2.73
Fleiss' kappa between human raters	0.87	1	1

Table 2. Labeled topics descriptive									
Predefined Topics	Inclusive set			Full agreement set			Partial agreement set		
	Count	Topics %	Documents %	Count	Topics %	Documents %	Count	Topics %	Documents %
Value	280	16.52	56.00	128	16.20	49.23	213	15.63	42.68
Service	340	20.06	68.00	170	21.52	60.71	277	20.32	55.51
Room	261	15.40	52.20	117	14.81	41.78	228	16.73	45.69
Location	251	14.81	50.20	115	14.56	41.07	215	15.77	43.09
Hotel amenities	193	11.39	38.60	66	8.35	23.57	113	8.29	22.64
Greetings and salutations	255	15.04	51.00	158	20.00	56.43	246	18.05	49.30
Food	115	6.78	23.00	36	4.56	12.86	71	5.21	14.23

3   Conclusions

Gold standard sets enable the systematic and rigorous evaluation of different topic models' parametrizations and the measurement of their precision in identifying humanly interpretable topics. They are required to adopt HIT. In this paper, we contribute three gold standard sets that can be used in research analyzing online customers' reviews in the context of the lodging industry and a protocol for creating reliable gold standard sets in different areas of interest. We encourage IS researchers to follow our example and use the AIS Transactions on Replication Research journal to share open access gold standard sets.

References

Palese, B., & Piccoli, G. (forthcoming). Evaluating topic modeling interpretability using topic labeled gold standard sets. *Communications of the Association for Information Systems.*

About the Authors

Biagio Palese. is an Assistant Professor for Information Systems in the College of Business at Northern Illinois University. He earned his PhD in Business Administration with a concentration in Information Systems at Louisiana State University. His teaching and research interests embrace introduction to management information systems, data analytics, effective use, customer service, digital data streams and text mining. His research has appeared in journals such as *MISQ Executive*, the *International Journal of Information Management*, the *European Journal of Information Systems*, *Information & Management* and in conference proceedings, including the International Conference of Information Systems and Americas Conference of Information Systems. The potential of his research has been recognized with his selection at the ICIS 2018 Doctoral Consortium. He has presented at various conferences, including ICIS, AMCIS, SIM Connect Live and BIG XII MIS Symposium.

Gabriele Piccoli. is the Edward G. Schleider chair for information systems in the E. J. Ourso College of Business at Louisiana State University and is on leave from the University of Pavia. He is the director of the Digital Data Streams Lab at LSU. His research, teaching, and consulting expertise are in strategic information systems and the use of advanced IT to support customer service. His most recent research focus is on digital data streams and their potential for value creation. He is author of the book *Information Systems for Managers: Text and Cases*. His research has appeared in both academic and applied outlets such as *MIS Quarterly*, *Journal of AIS*, *European Journal of Information Systems*, *Decision Sciences*, *California Management Review*, *MIS Quarterly Executive*, *Communications of the ACM*, and *Harvard Business Review*. He has published 16 full-length teaching case studies through *Communications of the Association for Information Systems* as well as *Harvard Business School Publishing*.