Aug 10th, 12:00 AM

Using Deep Learning for Identifying Social Media Customer Service Opportunities

Richard Gruetzemacher
Wichita State University, ross.gruetzemacher@wichita.edu

Sue Abdinnour
Wichita State University, sue.abdinnour@wichita.edu

Using Deep Learning for Identifying Social Media Customer Service Opportunities

Emergent Research Forum (ERF)

Ross Gruetzemacher
Wichita State University
Ross.Gruetzemacher@wichita.edu

Sue Abdinnour
Wichita State University
Sue.Abdinnour@wichita.edu

Abstract

Social media is commonly used by organizations to address customer service issues like complaints. For example, this is common in the airline industry, but not all organizations utilize it. However, the latest natural language processing (NLP) techniques can enable organizations to identify complaints—and other social media customer service opportunities—automatically. This exploratory work demonstrates a system for identifying complaints in social media using these NLP techniques that significantly improves upon the current state-of-the-art. We discuss implications for research, practice and future work.

Keywords

Social media, customer service, deep learning, business analytics.

Introduction

Organizations now widely use social media for responding to customer service complaints. A prime example of this is the airline industry, where some firms heavily rely on this approach. This is understandable as air travelers often encounter delays, which can result in long waits for customer service agents, both in the airport and in call centers. Social media offers an opportunity for restless travelers to vent their frustrations to an ever-present online audience. However, complaints by influencers can quickly propagate through social networks to the detriment of airlines (Gunarathne et al. 2018). Thus, it is critical that airlines identify and address social media complaints quickly; oftentimes complaining on social media works out better for customers than complaints directed through other channels (Gunarathne et al. 2015), especially for influencers (Gunarathne et al. 2017).

Airlines' use of social media customer service (SMCS) is not difficult to observe. Table 1 contains data from two Twitter interactions collected in November of 2021 that are demonstrative of airline SMCS in practice. Tweets depicted in the first two rows illustrate an interaction between a disgruntled customer and United Airlines. In this interaction we see that United responded quickly – within three minutes – and resolved the issue. In the last two rows of this table, we see an interaction between Neil Druckman, a social media influencer, and Delta Airlines, about his bags being lost by the airline. In this interaction we observe a response that is very different because Delta takes nearly 12 hours to reply. It is interesting that the Tweet was retweeted hundreds of times and liked thousands of times before Delta was able to respond. Ultimately, Druckman did thank Delta for the quality customer service when his bags finally arrived at his destination, but not before the damage was done. Responding earlier could have successfully mitigated the negative publicity that this Tweet created.

Airlines are not alone in using social media as a channel for customer service; reputable organizations across a wide variety of industries have recognized the value of using social media to this end (Guo et al. 2020). To explore this topic more deeply, big data analytics techniques could be very useful (Gunarathne et al. 2015), such as for classifying complaints, compliments or other customer service opportunities. Previous work has explored the use of machine learning techniques for classifying complaints in social media data, and surprisingly deep learning was found to be inferior to simple feature extraction and logistic regression (Preotiuc-Pietro et al. 2019).

In this preliminary study, we demonstrate that the latest deep learning techniques for natural language processing (NLP) achieves a significant increase in performance over the previous state-of-the-art techniques (Preotiuc-Pietro et al. 2019) for SMCS complaint classification. Drawing on the strong capabilities of these powerful techniques, we discuss the implications for applying these techniques to SMCS information systems (IS) research.

User	Date	Time	Tweet
@TanayLakhani	12/19/21	9:51 PM	@united flight UA684 is delayed by 2 and half hours. We missed our hotel checkin window. Please provide whom to contact for reimbursement.
@united	12/19/21	9:54 PM	Hello, Tanay. We are sorry to hear about this travel disruption. Once your travels are complete, you can file a formal claim with our Customer Care team for possible comepnsation options. When you have a moment, you can file your claim, here: fly.united.com/jThXh9.
@Neil_Druckman	12/9/21	5:04 PM	Me: Let's fly @delta. It can't be as bad as last time. All baggage lost Hours on hold Delta: Yes, we can see you checked your bags in, but we have no clue if they were loaded onto the original flight, the connecting one or neither. Me: 🙁 Delta: Thanks for choosing Delta!
@Delta	12/10/21	4:59 AM	Hi! This is Robin in the Delta Baggage Service Center. Please use this DM link to send us a private message and I'll be able to further assist you.

Table 1. Airline Social Media Customer Service Examples on Twitter

Methodology

Deep transfer learning has been an effective method for a variety of tasks for over a decade (Bengio 2012). More recently, these techniques have become widespread in the area of NLP due to the advent of the transformer architecture (Vaswani et al. 2017) and other advances in NLP involving deep transfer learning (Howard and Ruder 2018). Gruetzemacher and Paradice (2022) describe these recent advances, and their implications for IS research, in depth. This study demonstrates the power of these techniques for a practical business application with simple, baseline transformer language models.

Statistical language models are an NLP method that essentially functions to predict the next word in a sentence. Therefore, they can be used for tasks like autocomplete, but also for text generation tasks including in conversational agents, for report generating, for question-and-answer tasks, etc. Transformer language models first emerged in the summer of 2018 from OpenAI with the Generative Pretrained Transformer (GPT; Radford et al. 2018) and have rapidly become one of the most widely used NLP techniques. A more recent version of OpenAI's GPT, GPT-3 (Brown et al. 2020), has led to dramatic changes in thinking about AI capabilities that has resulted in the emergence of essentially a new paradigm described as foundation models.

Google's BERT language model (Devlin et al. 2019) is widely thought to be a good baseline for benchmarking transformer language model performance. BERT is an autoencoding transformer language model that is well-suited for discriminative tasks and has been adapted for a wide variety of tasks ranging from science (Beltagy et al. 2019) to finance (Araci et al. 2019) to conversational question and answering systems (Reddy et al. 2019). Because BERT is a standard baseline, and the most widely used transformer language model, we selected it as the first model to use for our SMCS classification task. RoBERTa (Liu et al. 2019) is a robustly optimized version of BERT resulting from a replication study that explored using training data and emphasized the significance of hyperparameters tuning, outperforming BERT.

BERT is a relatively large language model (Devlin et al. 2019) and requires the use of graphics processing units or other co-processors due to the large amount of computation required for training and fine-tuning. The bulk of the heavy computation is conducted during pretraining by large tech firms, and then pretrained models are released to the public. These pretrained models contain a lot of knowledge about the world because they are trained on very large corpora, like Wikipedia and Google Books corpora. Using the process of fine-tuning, this knowledge can be transferred to downstream, domain-specific tasks. Here, the knowledge is being transferred to the task of classifying social media complaints.

In this study, we use the standard model sizes of BERT (Devlin et al. 2019) and RoBERTa (Liu et al. 2019). BERT-large and RoBERTa-large typically perform better but require greater computational resources for training and fine-tuning. Since this is an exploratory study, we selected to simply use the base models. BERT and RoBERTa are no longer state-of-the-art on any common NLP benchmarks (Wang et al. 2019), but they are still some of the most widely used language models for many applications because they are well-suited for fine-tuning tasks due to their strong adaptability. While newer models may lead to increased performance, a systematic comparison of transformer language models is beyond the scope of this study, and BERT and RoBERTa are effective for demonstrating the strengths of models relative to previous deep learning techniques for NLP.

Fine-tuning of the models was conducted using Google Colab notebooks utilizing graphics processing units for the deep learning computation. Fine-tuning is implemented in a supervised learning fashion, by further training the pretrained model on the labeled, complaint classification dataset. We used 10-fold cross-validation and default hyperparameters without any hyperparameter tuning, so the results approximate those of test data.

Preliminary Results

Preotiuc-Pietro et al. (2019) presented a labeled dataset for classifying complaints collected from Twitter. These complaints were taken from the Twitter API based on 93 usernames of firms that actively used their accounts to reply to complaints. In total, this dataset contains 3,499 Tweets that are labeled for binary classification as either a complaint or not a complaint, and the data is available online. The results previously demonstrated by Preotiuc-Pietro et al. are representative of the previous state-of-the-art. Both techniques used in the 2019 study involve deep learning, but the use of deep learning for feature extraction was slightly better.

The results are shown in Table 2. The highest results are emphasized with bold. BERT outperforms the previous state-of-the-art for all metrics, and by a significant margin. However, as expected, RoBERTa outperforms BERT by a significant margin. We report precision and recall because they illustrate that much of the overall gains in performance for RoBERTa are due to improved recall.

	ROC AUC	Accuracy	F1	Precision	Recall
BERT-base (this study)	0.935	0.867	0.795	0.865	0.753
RoBERTa-base (this study)	**0.952**	**0.893**	**0.847**	**0.87**	**0.83**
Logistic regression w/ deep features (2019)	0.873	0.805	0.78	-	-
Long short-term memory (2019)	0.864	0.802	0.77	-	-

Table 2. Transformer Language Model Results for SMCS Complaint Classification

Discussion, Future Work & Implications

With simple fine-tuning of the RoBERTa-base model we were able to achieve a new state-of-the-art on the task of SMCS complaint classification. The performance of this model could be trivially improved by simply using more computational resources to fine-tune RoBERTa-large. Further improvement could be obtained in a straightforward manner with different pretraining strategies or better suited language models (Gruetzemacher and Paradice 2022).

Future Work

With respect to IS theory, we see future work using similar deep learning tools as offering opportunities to extend existing theory related to SMCS (Gunarathne et al. 2018). Specifically, we see strong potential for investigating differential customer service on social media across industries and across cultures. This has not previously been possible, but these new techniques, including multilingual transformer language models (Conneau et al. 2020), offer unique opportunities for IS researchers to leverage big data and machine learning, and particularly for continuing this project.

With respect to the methodology, we see future work exploring the unique characteristics of different transformer language models for the task of SMCS complaint classification, particularly in the context of using these tools for decision support or expert systems. This would make for a valuable contribution to the literature as we are aware of no existing work documenting the pros and cons of these different models for applications in business.

Implications for Research and Practice

Although it is only beginning to be applied in IS research (Samtani et al. 2020), deep learning clearly has tremendous promise for both research and practice. Similarly, the advanced NLP techniques demonstrated here have especially strong potential for IS research (Gruetzemacher and Paradice 2022) and practice (Gruetzemacher 2022).

The application of the big data driven techniques described in this paper—like transformer language models—to other, IS research problems offers significant implications for future IS research in applications well beyond SMCS. Particularly, we feel that multilingual language models (Conneau et al. 2020) hold tremendous potential for future IS research because they unlock multilingual textual analysis that is a very promising area, but which has been scarcely tapped into (Ebrahim et al. forthcoming; George et al. 2018).

This work also carries significant implications for practice. As the second example in Table 1 illustrates, organizations like Delta are not effectively utilizing decision support systems and expert systems for identifying customer service opportunities involving social media influencers. Consequently, the development and deployment of SMCS decision support systems and expert systems based on the techniques demonstrated in this work could have immediate significant advantages for firms that do not use them. Such systems can make SMCS interactions—like that demonstrated in the first example of Table 1—possible with minimal resources; it may even be possible to automate many interactions like this using expert systems to support social media bots.

Conclusion

In this preliminary work we demonstrate the power of using transformer language models and deep transfer learning to achieve state-of-the-art performance on SMCS complaint classification. This is an important task, and given our choice of a modest, baseline transformer language model, it is clear that further improvements over the results presented here can be obtained in a straightforward fashion. This work has significant implications for future work involving social media customer service because it shows that deep learning can be applied to big data with strong enough performance to be used for more practical applications to gain greater insight into firms' practices involving customer service via social media.

REFERENCES

Araci, D. 2019. *Finbert: Financial Sentiment Analysis with Pre-trained Language Models*. Master's Thesis. University of Amsterdam.

Beltagy, I., Lo, K. and Cohan, A. 2019. "SciBERT: A Pretrained Language Model for Scientific Text," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (EMNLP), Hong Kong, China, pp. 3615-3620. Association of Computational Linguistics.

Bengio, Y. 2012. "Deep Learning of Representations for Unsupervised and Transfer Learning," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Edinburgh, Scotland, pp. 17-36.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D. 2020. "Language Models are Few-shot Learners," in *Advances in Neural Information Processing Systems* 33 (NeurIPS).

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L. and Stoyanov, V. 2019. "Unsupervised Cross-lingual Representation Learning at Scale," in *Proceedings of Annual Meeting of the Association of Computational Linguistics*, Florence, Italy, pp. 8440-8451. Association of Computational Linguistics.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the Annual Meeting of the North American Chapter of the Association of Computational Linguistics*, Minneapolis, United States, pp. 4171-4186. Association of Computational Linguistics.

Ebrahimi, M., Chai, Y., Samtani, S. and Chen, H. 2021. "Cross-lingual Cybersecurity Analytics in the International Dark Web with Adversarial Deep Representation Learning," *MIS Quarterly*, forthcoming.

George, J. F., Gupta, M., Giordano, G., Mills, A. M., Tennant, V. M. and Lewis, C. C. 2018. The Effects of Communication Media Culture on Deception Detection Accuracy," *MIS Quarterly*, 42(2), 551-575.

Gruetzemacher, R. 2022. "The Power of Natural Language Processing," in *Harvard Business Review* (digital article). Retreived from https://hbr.org/2022/04/the-power-of-natural-language-processing.

Gruetzemacher, R. and Paradice, D. 2022. Deep Transfer Learning & Beyond: Transformer Language Models in Information Systems Research. *ACM Computing Surveys*. Forthcoming.

Gunarathne, P., Rui, H. and Seidmann, A. 2015. "Customer Service on Social Media: The Effect of Customer Popularity and Sentiment on Airline Response," in *Proceedings of the 48th Hawaii International Conference on System Sciences* (HICSS), pp. 3288-3297. IEEE.

Gunarathne, P., Rui, H. and Seidmann, A. 2017. "Whose and What Social Media Complaints Have Happier Resolutions? Evidence from Twitter," *Journal of Management Information Systems*, (34:2), pp.314-340.

Gunarathne, P., Rui, H. and Seidmann, A. 2018. "When Social Media Delivers Customer Service: Differential Customer Treatment in the Airline Industry," *MIS Quarterly*, (42:2), pp.489-520.

Guo, Y., Fan, D. and Zhang, X. 2020. "Social Media–based Customer Service and Firm Reputation," *International Journal of Operations & Production Management*, (40:5), pp. 575-601.

Howard, J. and Ruder, S. 2018. "Universal Language Model Fine-tuning for Text Classification," *arXiv:1801.06146*. Retrieved from https://arxiv.org/abs/1801.06146.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv:1907.11692*. Retrieved from https://arxiv.org/abs/1907.11692.

Preotiuc-Pietro, D., Gaman, M. and Aletras, N. 2019. "Automatically Identifying Complaints in Social Media," in *Proceedings of Annual Meeting of the Association of Computational Linguistics*, Florence, Italy, pp. 5008-5019. Association of Computational Linguistics.

Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I. 2018. "Improving Language Understanding by Generative Pre-training," Blog, OpenAI, retrieved Dec. 11, 2021, from https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/languageunsupervised/language_understanding_paper.pdf.

Reddy, S., Chen, D. and Manning, C.D. 2019. "COQA: A Conversational Question Answering Challenge," *Transactions of the Association for Computational Linguistics*, 7, pp.249-266.

Samtani, S., Zhu, H., Padmanabhan, B., Chai, Y. and Chen, H. 2020. "Deep Learning for Information Systems Research," *arXiv:2010.05774*. Retrieved from https://arxiv.org/abs/2010.05774.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. 2017. "Attention is All You Need," in *Advances in Neural Information Processing Systems* (NIPS), Long Beach, United States, pp. 5998-6008.

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O. and Bowman, S.R. 2019. "SuperGlUE: A Stickier Benchmark for General-purpose Language Understanding Systems," in *Advances in Neural Information Processing Systems* 32 (NeurIPS), Vancouver, Canada.