Proceedings of the ASME 2022
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC-CIE2022
August 14-17, 2022, St. Louis, Missouri

DETC2022-90108

ASPECT-SENTIMENT-GUIDED OPINION SUMMARIZATION FOR USER NEED ELICITATION FROM ONLINE REVIEWS

Yi Han
Mohsen Moghaddam*
Department or Mechanical & Industrial Engineering
Northeastern University
Boston, MA 02115
{han.yi1}{m.moghaddam}@northeastern.edu

Meet Tusharbhai Suthar
Gaurav Nanda
School of Engineering Technology
Purdue University
West Lafayette, IN 47907
{msuthar}{gnanda}@purdue.edu

ABSTRACT

Extracting and analyzing informative user opinion from large-scale online reviews is a key success factor in product design processes. However, user reviews are naturally unstructured, noisy, and verbose. Recent advances in abstractive text summrization provide an unprecedented opportunity to systematically generate summaries of user opinions to facilitate need finding for designers. Yet, two main gaps in the state-of-the-art opinion summarization methods limit their applicability to the product design domain. First is the lack of capabilities to guide the generative process with respect to various product aspects and user sentiments (e.g., polarity, subjectivity), and the second gap is the lack of annotated training datasets for supervised learning. This paper tackles these gaps by (1) devising an efficient and scalable methodology for abstractive opinion summarization from online reviews guided by aspects terms and sentiment polarities, and (2) automatically generating a reusable synthetic training dataset that captures various degrees of granularity and polarity. The methodology contributes a multi-instance pooling model with aspect and sentiment information integrated (MAS), a synthetic data assembled using the results of the MAS model, and a fine-tuned pretrained sequence-to-sequence model "T5" for summary generation. Numerical experiments are conducted on a large dataset scraped from a major e-commerce retail store for sneakers to demonstrate the performance, feasibility, and potentials of the developed methodology. Several directions are provided for future exploration in the area of automated opinion summarization for user-centered product design.

Keywords: Abstractive summarization; Sentiment analysis; Need finding; User-centered design

NOMENCLATURE

C	The review corpus
A	The Aspect word set
a_n	The nth Aspect word in A
r_i	ith review in a review batch
w_i	ith word in a sentence
e	The encoding from pretrained model
P_t	Token level prediction in the MAS model
P_s	Sentence level prediction in the MAS model
P_{sa}	Sentence level aspect prediciton in the MAS model
P_{ss}	Sentence level sentiment prediciton in the MAS model
h	The attention head in the MAS model
a	The element wise product of each key and query
k	The kth head in the MAS model
P_h	The head prediction in the MAS model
P_r	The review level prediction in the MAS model
$\mathscr{L}_{loss}$	The loss used in the MAS model
y	The actual label
$\hat{y}$	The prediction label
$\mathscr{L}_{sum}$	The loss used in the sequence to sequence model

*Corresponding author.

1 INTRODUCTION

User feedback plays an important role in product design as it provides vital information about user experiences of interacting with various product aspects to designers and manufacturers. With the increasing use of e-commerce platforms, a large collection of user feedback in the form of online product reviews is becoming available [1]. One of the main advantages of analyzing online product reviews is that we can obtain detailed and nuanced feedback from a large number of diverse users on different aspects of the product [2, 3], which is not the case in pilot launch, small-scale usability studies, or focus-group studies involving product design and development teams [4, 5, 6]. On the flip side, it is also challenging to comprehend a large collection of textual reviews where a single review typically involves varied user experiences associated with various aspects of the product.

Natural Language Processing (NLP) approaches such as text summarization, sentiment analysis, and topic modeling can be used to extract the prominent themes from a collection of user reviews [7, 8]. Among these, topic models need qualitative interpretation of generated topics which often require significant effort and time. On the other hand, text summarization approaches provide a compiled summary of important points covered in a large collection of reviews which can be used directly by the product designers for further analysis [9]. There are mainly two types of text summarization approaches: extractive [10, 11, 12] and abstractive [13, 14, 15]. The former extract and concatenates key sentences or paragraphs from the original text without necessarily capturing their meaning, while the latter leverages language models to generate text in a more advanced fashion, similar to human interpretation.

Previous studies have examined the performance of approaches for text summarization of online reviews to summarize overall reviews and summaries of different aspects mostly involving services such as hotels and restaurants [16, 17]. One of the limitations of summarizing overall reviews or aspect-based reviews is that it can mix up the positive and negative opinions of users about that particular aspect. To address these gaps in the current literature, this paper develops a novel model for abstractive summarization of user reviews to generate sentiment-based, aspect-based summaries. The performance of the model is demonstrated and evaluated on a large review dataset of sneakers scraped from multiple e-commerce platforms.

For sentiment analysis, polarity and subjectivity are considered as two main dimensions, and determined these for various product aspects as well as for the overall product. The sentiment polarity score indicates the intensity of emotions expressed by the user, e.g., extremely negative/unhappy, neutral, moderately positive, or highly positive. The sentiment subjectivity score indicates whether the review was largely a subjective opinion, e.g., "I did not like the shoe sole" or it was objective in nature, e.g., "The shoe sole was very narrow". Both these sentiment dimensions hold important information about user experience which are mostly complementary in nature. As this is a pilot study, the sentiment intensity at the aspect level was used to group data into unique combinations of aspect and sentiment polarity, and the reviews were summarized for each group using an abstractive summarization approach. The findings from this study would be useful for researchers in the engineering design domain as well as product designers and manufacturers.

The remainder of this paper is organized as follows. Section 2 presents the NLP framework for abstractive opinion summarization. Section 3 provides the details of the computational experiments on the large review dataset for sneakers, and analyzes the computational results. Section 4 presents discussions and concluding remarks.

2 METHODOLOGY

To develop an abstractive summarization model through supervised learning, a labeled dataset that includes reviews and summary pairs is required. Yet, such dataset are rare and very hard to generate. In such cases, several studies have attempted to train supervised learning models through creating synthetic datasets, which has demonstrated remarkable performance [18, 19, 20]. Building on this idea, this paper conducts abstractive opinion summarization through a three-stage process as follows:

1. Training a multi-aspect and sentiment (MAS) model with a review-based dataset.
2. Generating synthetic dataset with the output of the MAS model.
3. Fine-tuning a state-of-the-art sequence-to-sequence model with the synthetic dataset to generate abstractive summaries for specified aspect and sentiment polarity.

The proposed model is an extension work of the aspect-controllable summarization model AceSum [21], with the following additional features:

1. AceSum only includes aspect controller, while the proposed model integrates with sentiment polarity (i.e., the sentiment controller). Further, the AceSum instance model may yield no output, yet the proposed model always predicts at least one label which is the sentiment.
2. The multi-instance model of AceSum creates the synthetic dataset using less than 10 seed words, while the proposed model generates the synthetic dataset using a rich aspect lexicon previously developed by the authors [22].
3. AceSum uses a soft-margin loss function for the multi-instance model because their label set was binary with -1 and 1. The proposed model, however, uses the Sigmoid binary cross entropy loss function for training to reduce the influence of the unbalanced dataset with respect to aspects and sentiments.

4. In the synthetic data creation process, AceSum assumes that in a review set, each review that fulfills some constraints to be a summary and all the rest to be the training corpus. The proposed model does not use the entire review as a summary, and instead only assembles a set of model selected sentences as a summary.

The multi-instance model is a machine learning framework in which labels correspond to a bag of instances that have not been labeled [23]. The goal of the model is to identify the bag's labels from those unlabeled instances. In this work, a hierarchical model structure has been used to predict the review labels from sentence and word predictions. The reason for choosing the multi-instance model is the similarity of the model structure to the human summary generation process. In the data labeling process, the first step was also filtering useful sentences from a bunch of reviews, when creating the aspect-related summary, annotators generate content from those sentences which related to a specific aspect. They then summarize those sentences to a single summary, which must have the same label as those sentences. In the MAS model, the sentiment label was added along with the aspect label with three types of polarities associated with the review: positive, neutral, and negative. In this case, the stars provided for the reviews on the e-commerce platform were used to induce user sentiments. Specifically, 5 stars denote positive sentiment, 3-4 stars indicate neutral sentiment, and 1-2 stars means negative sentiment.

2.1 Multi-instance model

The multi-instance sentiment model can be formulated as follows (Figure 1). Let C denote the corpus which includes user reviews with the stars provided by the user and $A = a_1, a_2, ...a_n$ denote the aspect set [22]. Each review r_i can be formulated as a list of words $w_1, w_2...w_n$. For a given review with word list w_n, RoBERTa [24] tokenizer RB is utilized for encoding. Thus, the encoding process can be expressed as $e = RB(W_n)$. The proposed model uses label $\{0,1\}$, instead of $\{-1,1\}$ to indicate the results. Thus, the token-level prediction P_t can be obtained using a non-linear transformation:

$$P_t = ReLU(W_e + b) \tag{1}$$

The model then uses token-level predictions to induct sentence level predictions P_s. The induction process uses the multiple attention mechanism [25] which has been implemented in the multi-instance model. The proposed model also utilizes 12 attention heads. The multi-instance model structure is depicted in Figure 1. Specifically, the batched result of P_t is split into 12 heads h. Each key key_h is transformed with a non-linear transformation. To enable better differentiation, $ReLU$ activation functions are used in the attention mechanism instead of $tanh$:

$$key_h = ReLU(W_{he} + b_h) \tag{2}$$

Other settings for the attention mechanism follow the original AceSum model. Each attention output is calculated as

$$a_h = softmax(key_h \cdot query_h) \tag{3}$$

and the head attention prediction is calculated as:$P_h = \sum_k (p_t * a_h[k])$. Each attention head in the model represents a semantic space of the review. Thus, the sentence level prediction for an aspect is calculated as follows:

$$P_{sa} = maxpooling(P_h) \tag{4}$$

Similarly, the predictions for sentiments are calculated as follows:

$$P_{ss} = maxpooling(P_h) \tag{5}$$

Analogously, the model uses sentence level prediction to induct the review level prediction P_r.

2.2 Training process

During the training process, for each review in the corpus, binary labels are used to indicate both the aspect and sentiment. Specifically, if the aspect is mentioned in a review r, the label P_r is 1. Otherwise, it is 0. For the sentiment labels, a one-hot formatting label is used in which three sentiment types are assigned: positive ('POS'), neutral ('NEU'), and negative ('Neg'). If the review has a positive sentiment, the label would be $(1,0,0)$, if it is neutral, the label would be $(0,1,0)$, and if it is negative, the label would be $(0,0,1)$. During the training process, it was observed that the dataset was highly imbalanced in terms of both aspects and sentiments. Around 50% of review did not mention any aspects from the word lexicon[22], and around 85% of reviews expressed positive sentiments. Hence, Sigmoid binary cross entropy function was used as the loss function to mitigate this imbalanced dataset issue:

$$\mathscr{L}_{loss}(y, \hat{y}) = -w_n[\hat{y_n} * logy_n + (1 - \hat{y_n}) * log(1 - y_n)] \tag{6}$$

2.3 Synthetic data creation

The MAS model yields three level predictions for aspects and sentiments. Such controllers provide a flexible way to assemble the synthetic dataset. With document level prediction,

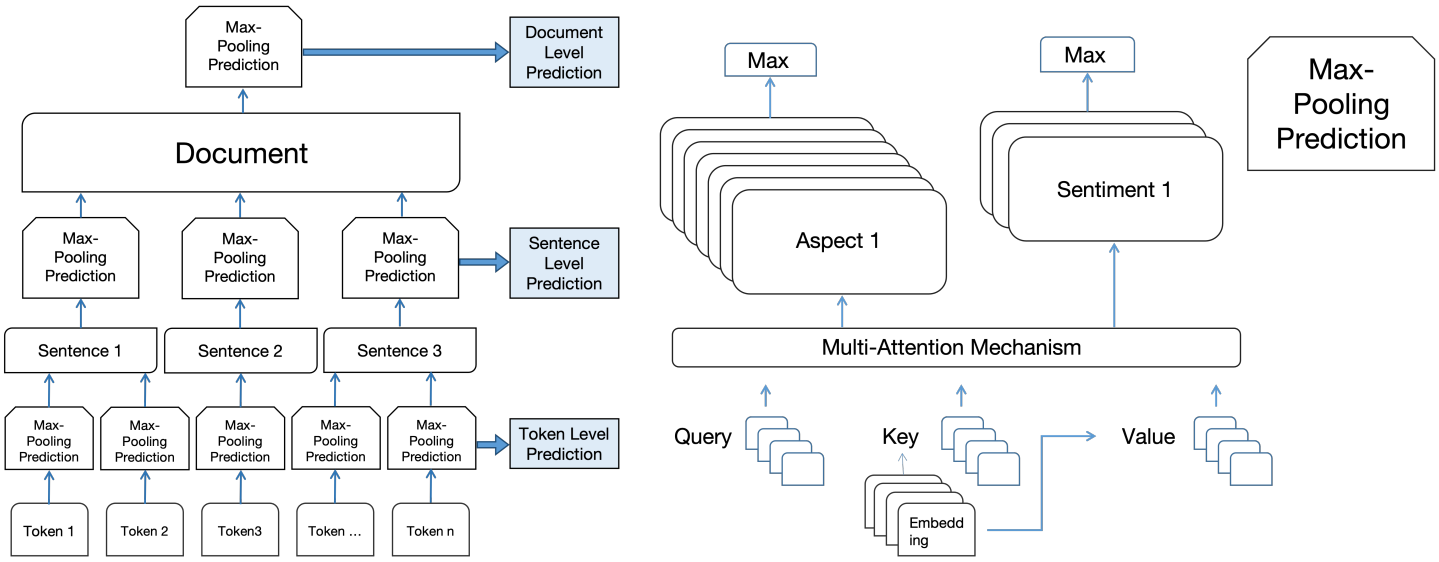

FIGURE 1: THE MULTI-INSTANCE MODEL.

P_{ra} can control the overall aspect of the summary, and sentiment controller P_{rs} can further control the polarity of the summary. Similarly, sentence level predictions P_{sa} and P_{ss} can control the sentences entering the training corpus that are aspect relevant. Token level predictions P_{ta} and P_{ts} were compared with the overall document prediction by a soft margin function:

$$\mathscr{S} = \sum log(1 + exp(-\hat{y_n} * (y_n))) \tag{7}$$

The top sentences and tokens are selected to enter the next step in the sequence-to-sequence model.

2.4 Abstractive summarization model

After building synthetic dataset, the state-of-the-art pretrained transformer based model Text-To-Text Transfer Transformer (T5) [26] for generating the opinion summaries. During the fine-tuning process, the outputs of the MAS model were also assembled in the following format:

[Aspect][Aspect1][Aspect2][Aspect..][Sentiment] [KEY] keyword1, keyword2, keyword3 ... [SNT] sentences ...

$[Aspect]$ indicates the current summary related to a certain aspect, $[sentiment]$ indicates sentiment polarity of the current model, $[KEY]$ and $[SNT]$ correspond to the selected keywords and sentences. P is used as the input and the encoding E is produced, then the decoder outputs a token distribution $p(y_t)$ conditioned on the T5 attention mechanism. Next, the model was fine-tuned with a maximum-likelihood function:

$$\mathscr{L}_{sum} = -\sum \hat{y_n} log p(y_n) \tag{8}$$

During the training process, the output of the model could be controlled by manipulating the aspect controller and the sentiment controller. Moreover, the model also has the ability to induct the overall summary by selecting all aspects during training.

3 EXPERIMENTS AND RESULTS

This section presents and analyzes the results of the computational experiments on a large dataset of reviews scraped from multiple online sneaker stores.

3.1 Dataset

A large dataset of reviews from the online websites of Finish Line [27], New Balance [28], and Asics [29] is used for demonstrating and evaluating the performance of the proposed model. The dataset includes 80k user reviews along with the star ratings corresponding to the reviews. The sneaker aspect lexicon includes 200+ aspect words and 7 categories [22]. Through the MAS model, a total of 8k training instances were generated.

3.2 Implementation

For fine-tuning the pretrained model, the weights and training options from the Hugging Face library [30] were used. The MAS model was trained with the learning rate of $1e-5$ for 10k

TABLE 1: DOCUMENT- AND SENTENCE-LEVEL F1 SCORES FOR THE MAS MODEL.

Metric	Loss function	Label type and activation in layers	Performance score
Document F1	BCE Mean Weight loss	(1, -1) label with tanh activation	79.41
Document F1	BCE Mean Weight Loss	(0, 1) label with relu activation	80.86
Document F1	BCE Sum Weight Loss	(0, 1) label with relu activation	74.84
Document F1	BCE Calculated Weight Loss	(0, 1) label with relu activation	**83.56**
Document	BCE Mean Weight loss	(1, -1) label with tanh activation	78.27
Sentence F1	BCE Mean Weight Loss	(0, 1) label with relu activation	80.21
Sentence F1	BCE Sum Weight Loss	(0, 1) label with relu activation	76.39
Sentence F1	BCE Calculated Weight Loss	(0, 1) label with relu activation	**83.41**

steps with 12 attention heads. For fine-tuning the T5 model, a learning rate of $1-e^6$ and 20k steps were set. At each training process, Adam was used with weight decay [31] as the optimizer and a linear learning rate scheduler was used for half of the step size. The summary model generates summaries with beam search size 2 and refrains from repeating grams with size 3.

3.3 Results

The results presented in this paper are based on an ongoing effort and can be improved with further experiments and refinements. Currently, the model's best performance of the ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation-Longest Common Subsequence) score is 0.17 for the 'overall' aspect with positive polarity, and the MAS model with different model structure performance could be found in Table 1. The rest of aspect training results are presented in Table 3. It is likely that different parameter settings lead to better performance results. Moreover, the model performance will be compared with other baseline models such as AceSum and MeanSum [32]. Some examples of generated summaries are provided below:

Aspect: Exterior. Sentiment: POS.

"I love the color way and the look of the shoe. I have a pair of black and white sneakers and they are very comfortable." "I love the color and the look of these shoes. They are so comfortable and I have a lot of compliments on them." "I love the look and style of these shoes. I have a pair of black and white sneakers and they are very comfortable." "I love the color combo and the color is amazing. I have a pair in a different color and they are very comfortable."

TABLE 2: THE ASPECT-SENTIMENT ROUGE-L SCORES.

Aspect	Polarity	ROUGE-L (%)
General	Positive	17.2
General	Negative	14.3
Aspect(exterior)	Positive	15.9
Aspect(exterior)	Negative	15.3
Aspect(Fit)	Positive	15.1
Aspect(Fit)	Negative	14.2

Aspect: Exterior. Sentiment: NEG.

"I've just about given up on Skechers. You never have my size.....wide, 11 1/2; if you do it's an ugly shoe and not the color I'm looking for !!!" "The right shoe had a color flaw on the right front outside and sole. A grey mark was embellished in the tan suede and orange front right side of the shoe. I didn't want flawed shoes." "I was so excited about this shoe only to be let down when I received the shoe in the mail. Color is not what I expected. Color is dull & looks worn out. Not the look I was going for." "The shoes are described as b̈lack on black' but are actually brown and black. I'm not sure if I was sent the wrong pair or is this is the color that they are only supposed to be but clearly they are brown

and not attractive at all in that style. very disappointed Because otherwise they fit perfectly."

Aspect: Overall. Sentiment: POS.

"I have small feet. And they are very comfortable. I love them.. and and they fit great." "If you're a little bit of a fan of the style, and the color of the shoe is very comfortable. And they're very comfortable to wear. They look great and comfortable too. They have a great fit.. so comfortable.." "I love the color and color of the shoes. A great color so comfortable. They're a good pair of shoes. They were very comfortable and I loved the color. They look great. Oh and they are super comfortable." "I'm a tall woman. And I have narrow feet. And they are very comfortable. I love the color.. They are comfortable, and they fit well.. very comfortable."

Aspect: Overall. Sentiment: NEG.

"The show itself is comfortable, but the band around the ankle has no stretch to it so it's almost impossible to get your foot in the shoe. Then after you finally do, it compresses your ankle and rubs." "I gave this rating due to receiving my order almost a month later. It was a Christmas gift, so it was a late gift. Initially they sent me two different sizes. I was not contacted in regards of the situation." "I ought the Valentine's pair. The tongue is stiff, and rubs against my ankle. They are so cute, but absolutely unwearable." "I LOVE the looks of these shoes and I can't find anything similar - looks wise. HOWEVER, I have only had these for a brief period of time and there are already toe holes in the top. I feel like maybe I should've gone 1/2 size up because they are tight feeling (not in the toe box but around my foot). I am super disappointed because I really like them otherwise. I am a frontline worker and got lots of compliments from my co-workers on them so i'm bummed that they haven't held up." "I am unable to wear the shoes because the way they are designed. They are extremely tight around the ankles and the heel, causing them to cut into your heel and cause blisters. They are extremely wide and loose everywhere else causing you to curl your toes to try to keep them on your feet while they cut the back of your ankles up"

These summaries were generated using a random set of reviews. The output was constrained to the aspect 'Exterior' controller (includes lots of color-related aspect words) and the sentiment 'Positive' controller. From the generation, one can induct that users with positive attitude of this shoe is especially comfortable with the color in the overall exterior aspect. The output of the model generated by all all aspects obviously includes multiple aspects as expected. And the subjectivity analysis results of the aspect 'exterior' was:

Polarity= Negative, Average Subjectivity=0.46
Polarity= Positive, Average Subjectivity=0.67

The closer Subjectivity score is to 1, the more likely it is to be an opinion rather than fact. While the Subjectivity score of Negative reviews was relatively less as compared to Positive reviews, detailed statistical analysis would be required to assess whether the negative reviews were relatively more objective in nature as compared to positive reviews. From the range of Subjectivity scores, it seemed that the reviews about 'exterior' aspect were reasonably objective in nature.

4 CONCLUSIONS

The summaries generated by the proposed abstractive opinion summarization model can directly provide potentially useful feedback for product design teams. In the exterior color summaries, for example, we can induct that the users are at least satisfied by the 'combo color', especially with the 'white and black' combo. In the overall summary, 'comfortable' as one of the aspect has been generated with an obvious higher possibility in the output, meanwhile, in the training process, the highest possible token prediction in also the word "comfortable". In the results part, to better illustrate the comparison of the users feedback, we perform the experiment on both positive controller and negative controller, summary may not be derived from the same product, but it comes from the same brand, because in the corpus dataset, the reviews was first sorted by the brand, and the result are collected from the first 50 summaries. In the results of aspect "exterior" with negative attitude, some users are complaining about the color flaw on the shoe side, this flaw may appear in the manufacturing process, and a group of users complaining about the color bias, the color are not they expected or not as they think in the color description, designers may consider also be involved in the product sale process, they could provide more accurate descriptions. In the overall summaries, the complaints become more various, the most perti6nent aspect they mention in the "overall" summaries was this sneaker does not fit well, more specifically, in the summary, lots of users are complaining the sneaker was hurting the ankle, if this is a common problem around unsatisfied users, designer may consider a more flexible design of the ankle area.

Future research can enhance the performance of the proposed opinion summarization model in multiple ways. More extensive experiments should be performed on each individual aspect to better test and validate model performance in aspect-guided summary generation. The MAS model can be improved by exploring other pooling mechanisms besides max pooling (e.g., attention pooling, mean pooling). The model performance can also be improved by better tuning of the hyper-parameters (e.g., exploring other learning rates) and increasing the train-

ing steps (10k steps used for training the model in this paper). Regarding the pretrained language model, other state-of-the-art models besides T5 (or an improved version of it) can be investigated. Further, the current study is based on a synthetic dataset for training, validating, and testing. The model performance is expected to improve if a more accurate validation dataset is used in the training process. To this end, we plan to develop a human annotated dataset that will be used in next phase.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under the Engineering Design and System Engineering (EDSE) grant #2050052. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

A Appendix A: The aspect lexicon used in the MAS model.

TABLE 3: THE ASPECT LEXICON [22].

Aspect	Part of attributes inside
Permeability	“ventilation”, “breathable”, “mesh”...
Impact absorption	“air”, “gel”, “strap”, ...
Stability	“flytrap”, “ankle”, “support”, ...
Durability	“durable”, “ripple”, “haptic”, ...
Shoe_Parts	“tonal”, “bucket”, “bottom”, ...
Exterior	“gold”, “blocking”, “metallic”, ...
Fit	“dapper”, “comfy”, “adjustable”, ...

REFERENCES

[1] Robert G. Cooper, Scott J. Edgett, and Elko J. Kleinschmidt. Benchmarking best NPD practices - III, 2004.

[2] A.F. Osborn. *Applied Imagination*. Scribner’s, New York, 1953.

[3] Tucker J. Marion and Sebastian K. Fixson. *The innovation navigator: Transforming your organization in the era of digital design and collaborative culture*. University of Toronto Press, jan 2018.

[4] Stacey, Mcfadzean, and J ’ Sketch. Managing effec-tive communication in knitwear design. Technical Report 3, 1997.

[5] Golnoosh Rasoulifar, Claudia Eckert, and Guy Prudhomme. Communicating consumer needs in the design process of branded products. *Journal of Mechanical Design, Transactions of the ASME*, 137(7), 2015.

[6] Nikolaus Franke, Martin Schreier, and Ulrike Kaiser. The “I Designed It Myself” Effect in Mass Customization. *Management Science*, 56(1):125–140, jan 2010.

[7] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving Language Understanding by Generative Pre-Training. Technical report, 2018.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.

[9] Yi Han and Mohsen Moghaddam. Eliciting Attribute-Level User Needs From Online Reviews With Deep Language Models and Information Extraction. *Journal of Mechanical Design*, 143(6), 11 2020. 061403.

[10] Hao Zheng and Mirella Lapata. Sentence Centrality Revisited for Unsupervised Summarization. *ACL Anthology*, pages 6236–6247, Jul 2019.

[11] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel S. Weld. TLDR: Extreme Summarization of Scientific Documents. *ACL Anthology*, pages 4766–4777, Nov 2020.

[12] Stefanos Angelidis and Mirella Lapata. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. *ACL Anthology*, pages 3675–3686, 2018.

[13] Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. *ACL Anthology*, pages 340–348, Aug 2010.

[14] Giuseppe Di Fabbrizio, Amanda Stent, and Robert Gaizauskas. A Hybrid Approach to Multi-document Summarization of Opinions in Reviews. *ACL Anthology*, pages 54–63, Jun 2014.

[15] Yang Liu and Mirella Lapata. Text Summarization with Pretrained Encoders. *ACL Anthology*, pages 3730–3740, Nov 2019.

[16] Stefanos Angelidis and Mirella Lapata. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. *arXiv*, Aug 2018.

[17] Stefanos Angelidis, Reinald Kim Amplayo, Yoshihiko Suhara, Xiaolan Wang, and Mirella Lapata. Extractive Opinion Summarization in Quantized Transformer Spaces. *arXiv*, Dec 2020.

[18] Eric Chu and Peter Liu. MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive Summarization. In *International Conference on Machine Learning*, pages 1223–1232. PMLR, May 2019.

[19] Reinald Kim Amplayo and Mirella Lapata. Unsupervised Opinion Summarization with Noising and Denoising. *ACL Anthology*, pages 1934–1945, Jul 2020.
[20] Hady Elsahar, Maximin Coavoux, Jos Rozen, and Matthias Gallé. Self-Supervised and Controlled Multi-Document Opinion Summarization. *ACL Anthology*, pages 1646–1662, Apr 2021.
[21] Reinald Kim Amplayo, Stefanos Angelidis, and Mirella Lapata. Aspect-Controllable Opinion Summarization. *arXiv*, Sep 2021.
[22] Yi Han and Mohsen Moghaddam. Analysis of sentiment expressions for user-centered design. *Expert Syst. Appl.*, 171:114604, Jun 2021.
[23] spaCy · Industrial-strength Natural Language Processing in Python, Jan 2022. [Online; accessed 23. Jan. 2022].
[24] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*, Jul 2019.
[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv*, Jun 2017.
[26] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv*, Oct 2019.
[27] The. Finish Line: Shoes, Sneakers, Athletic Clothing & Gear, Feb 2022. [Online; accessed 11. Feb. 2022].
[28] Athletic Footwear and Fitness Apparel - New Balance, Feb 2022. [Online; accessed 11. Feb. 2022].
[29] ASICS | Official U.S. Site | Running Shoes and Activewear | ASICS, Feb 2022. [Online; accessed 11. Feb. 2022].
[30] t5-base · Hugging Face, Feb 2022. [Online; accessed 11. Feb. 2022].
[31] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. *arXiv*, Nov 2017.
[32] Eric Chu and Peter J. Liu. MeanSum: A Neural Model for Unsupervised Multi-document Abstractive Summarization. *arXiv*, Oct 2018.