Coherent Personalized Paragraph Generation for a Successful Landing Page

Yusuf Mücahit Çetinkaya
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
yusufc@ceng.metu.edu.tr

İsmail Hakkı Toroslu
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
toroslu@ceng.metu.edu.tr

Hasan Davulcu
Computer Science and Engineering
Arizona State University
Tempe, AZ, U.S
hdavulcu@asu.edu

***Abstract*—Social media has become an important place for online marketing like never before. Businesses use various techniques to identify and reach potential customers across multiple platforms and deliver a message to grab their attention. A notable post could attract potential customers to the product landing page. However, the acquisition is only the beginning. The landing page should respond to the visitor's need for persuasion to increase conversion rates. Showing every visitor the same page is far from that goal. Even if the product meets everyone's needs, their priorities may differ. In this study, we propose a pipeline that includes gathering and identifying potential customers from Twitter, determining their priorities by understanding the context of their message, and creating a coherent paragraph that addresses the issue to display on the landing page.**

***Index Terms*—coherent paragraph generation, landing page optimization, online marketing, target customer, topic modeling, Twitter**

I. Introduction

Nowadays, companies try to use online tools to reach potential customers. The three fundamental steps of online marketing are:

- **Acquisition:** Making people visit the landing page
- **Conversion:** Persuading them to take actions
- **Retention:** Deepening the relationship with clients

Each activity feeds into the next. Many companies spend large amounts of money on acquisition and retention activities. The conversion step, namely the landing page, is usually neglected, making the spent money waste [1].

A product might have multiple functions to solve a variety of problems. Putting all the solutions on the landing page would result in a messy and hard-to-follow design. Each client's priorities may be different. Identifying the issues requires listing the problems that the product solves. Domain experts should create a codebook by matching the most occurring named entities and phrases in the collected tweets and corresponding categories.

In this study, we have enhanced the online marketing activity steps for the products or services where the target customer definition is unstructured. We have example use case where the target clients are patients suffering from high drug prices, specifically chronic diabetes, and the product is pharmacy discount card. We propose a novel end-to-end pipeline that generates a personalized landing page based on the prospective clients' priorities. The pipeline consists of the following steps;

- Determining the keywords of interest that needs to be used while gathering messages from social media,
- Building a model for detecting related messages among the collected messages with less effort of labeling,
- Generating a coherent landing page paragraph that leverages the message context to extrapolate the priorities.

Generating coherent paragraphs is not a trivial task. Sentence order in a paragraph is an important part of natural language generation. If sentences are not carefully arranged, paragraphs can become confusing or difficult to read. This problem is studied in multi-document summary [2]. In addition to topical relatedness, other characteristics such as chronology, precedence, and sequence between sentences must be considered to form a summary paragraph from multiple documents. Similarly, [3] focuses on restoring the original paragraph by arranging a set of unordered sentences. However, when writing generic sentences that do not rely on precedence or successive sentences, we need to focus on the thematic proximity between sentences to form a coherent landing page paragraph.

II. Methodology

The proposed procedure consists of successive components to generate a personalized dynamic landing page. The pipeline starts by collecting the tweets that contain specific keywords and distinguishing the related ones. The associated tweets are thematically modeled to extract priorities considering the codebook provided by the domain experts. Finally, the topic probabilities generate a coherent landing page paragraph addressing the prospect's concerns.

A. Collecting Tweets using pre-determined keywords

Keywords provided by humans would be biased and limited. Instead, finding the communities for the intended issue on social media gathers the corpus without much effort. In the example use case, we have collected 36 thousand posts from the Facebook groups and pages that people with diabetes meet. Picking the posts makes the corpus large enough to give an idea of the most common phrases or words.

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

We used Yet Another Keyword Extraction (YAKE!) to determine the important words to monitor on Twitter. It is an unsupervised, language-independent, term frequency-free method that relies on the statistical features of the corpus [4]. We concatenated the posts to have a single document and considered n-grams with lengths up to 3. The top 2000 keywords are extracted for further human labeling. The list we have ended up has been labeled as keyword or not by experts.

Assuming the tweets that contain given keywords are related to our target domain is a strong bias. Related tweets might include the determined keywords; however, it does not mean all tweets that contain these phrases are part of our target tweet set. Collected tweets need to be eliminated, considering their context to remain only the related ones.

B. Detecting related messages

The target tweets list is a subset of all tweets that contain specific keywords. An automated process for detecting the proper subset is required for a successful lock-on mechanism for prospective clients. The texts must be represented by word embeddings to measure their similarity quantitatively. The main contribution of the word embedding concept is to represent words in a shared space and reveal their semantic distance. This representation also enables neural networks to learn semantic representations of the word sequences better.

There are several word embedding algorithms and pre-trained models in the literature. In this study, we used a pre-trained model since it works as well as traditional topic modeling methods as discussed in [5] where they reported BERT [6] beats the others in that sense. Additionally, BERT handles the same word differently in different sentences considering the context of the whole sentence and the word's position. Example in Table I shows that the word "sweet" in different sentences is embedded into space by the context. The cosine similarities between the word embeddings of the same word are high when the context is similar.

TABLE I: The cosine similarities between 'sweet' words' BERT embeddings in different contexts.

Sentence 1	Sentence 2	Cos. Sim.
Eating lots of **sweet** has been shown to worsen joint pain because of the inflammation.	This is a very kind and **sweet** cat sitting on the sofa.	0.46
	People with diabetes can still eat **sweet** with a healthful meal plan.	0.84

There are 75 thousand randomly collected tweets containing given keywords. Pre-trained BERT on the English language with uncased letters is used to represent the tweets in a vector space. The last two layers' average of each word's embedding is accepted as the sentence embedding, forming a 768-dimensional vector. We have applied PCA to examine the samples visually and found that, as expected, there is no single separating hyperplane for related and unrelated tweets. We have clustered the tweets using K-Means with different k parameters and used the elbow method on silhouette scores, where $k = 150$ was the result.

The samples in the clusters are semantically close, which reduces the effort of labeling. Instead of human labeling all of the instances in the dataset as related/unrelated, top-100 samples closer to the center of each cluster are labeled, and the majority of the label is assigned to all of the tweets belonging to that cluster. The human-assisted model is used for the training phase and for detecting the suspects on the fly. The tweet is converted into word embedding, and the closest cluster with its label is assigned to the tweet. If it is related, the process continues; when it is unrelated, ignored.

C. Modeling topic for user's priorities

There are 8 particular categories, and Table II gives the details. The user's message may be prone to multiple categories with different weights corresponding to the user's priorities. For instance, the tweet *"Skyrocketing drug prices are hurting patients and costing taxpayers. Millions of Americans do not take their medications as prescribed due to the price"* has complaints about costs, medication, and patients.

TABLE II: Predefined marketable aspects of the product retrieved from the codebook and corresponding generic sentence details.

Category	Sentences About
card	Definition of the discount card and its mission
service	The Brand's service capabilities and reachability
cost	The issue about overpaying for prescriptions
insurance	The drawbacks of the traditional insurance approach
health	The importance of health and its effect on life
medication	The significance of the medication
patients	The people that the Brand can help
chronic	Explanations on chronic diseases

We leverage GuidedLDA [7] to take advantage of domain experts while detecting the user's priorities. We feed the model with a set of seed words that are representative of the corpus. Since the given seed is biased, we have added two blank topics along with the 8 for the model to populate by itself in case of failing to notice.

Table III shows the top-7 words in each topic identified by the GuidedLDA using given seeds. The last two topics are populated by the model, which is expected to be orthogonal to the given topics. The model has identified two unnoticed topics that mainly focus on hope&fight against fatal diseases and healthy life. Seeded topics are consistent with the provided codebook.

Words weigh each topic where the sum of all weights makes 1 and each topic has a set of words where their weights also sum to 1. When a document arrives, each phrase's significance for the corresponding topic is calculated by adding these occurrences and normalized to get a final topic distribution. Calculated probabilities are considered as the priorities of the message and consumed while generating the paragraph.

The model results in the topic distribution for the example about "skyrocketing drug prices" covered in this section to

TABLE III: Identified topics by GuidedLDA using given seeds showing top-7 n-grams.

Topic	Words
card	gift, card, holiday, gift card, credit card, coupon
service	healthcare, ai, technology, problem, security, solution, medtech
cost	price, cost, save, saving, pricing, discount, money
insurance	insurance, plan, medicare, health insurance, coverage, pay, medicaid
health	health, health care, care, digital, mental, aging, clinic
medication	medication, med, diaetes, flu, vaccine, insulin, pill
patients	patient, pharmacist, physician, cancer, medical, therapy, clinical
chronic	disease, chronic, disability, dementia, risk, chronic pain
new topic-1	hope, family, member, dream, impact, cancer, fight
new topic-2	love, holiday, stay, healthy, book, support, eat

68% cost, 20% medication, 10% patients, and 1% card. The results are highly consistent, which detects the priorities of the message as expected.

D. Generating a coherent landing page paragraph

The landing page paragraph begins with a related sentence about the prospective client's highest priority and progresses to the lowest. However, the paragraph should not be too long as this can prevent the user from reading the paragraph at all. We have limited the paragraph length to a maximum of five sentences.

Each topic has a set of generic sentences related to that topic discussing different aspects. The most related sentences from each selected topic are put to form a coherent paragraph. To measure the relatedness between the sentences and the user's message, they are converted to vectors using the trained GuidedLDA model. Cosine similarity between candidate sentences' and the tweet's representations are calculated, and the most similar sentence is adopted.

Considering the coherency in the paragraph, we have applied two different approaches;

1) Each topic sentence is selected, independently calculating the similarity between candidate sentences and the source message. The most semantically similar sentence is selected for every sentence.
2) The first sentence is selected considering the similarity between the highest priority topic candidates and the source message. Upcoming sentences are selected by comparing the previous and candidate sentences'.

For the running example in the previous section that identifies the topic, the selected sentences among candidates for the paragraph using both approaches are displayed in Table IV. Using only the source tweet while selecting the following sentences will produce a paragraph highly related to the source sentence. However, the incoherent behavior in the paragraph can be easily sensed. Similarly, using only the previous sentence solves the coherency problem, but the latter topics are becoming slightly disconnected from the source tweet.

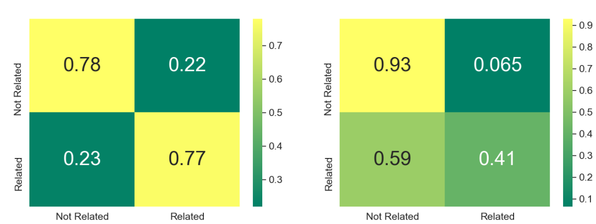
III. Experiments

To validate the relatedness model's performance, we randomly collected 174 tweets containing determined keywords

TABLE IV: Generated paragraph for the tweet *"Skyrocketing drug prices are hurting patients and costing taxpayers. Millions of Americans don't take their medications as prescribed due to the price."* using two different approaches.

Topic	Selected Sentences
	Only source tweet
cost	The cost of a prescription may differ by more than $100 between pharmacies across the street from each other!
medication	Taking the right medications at the right time is not only important financially—it is also essential to health and wellbeing.
patients	[BRAND] works for everyone: insured, uninsured, Medicare recipients.
card	[BRAND] is working to create a healthier wealthier America.
	Only previous sentence
cost	The cost of a prescription may differ by more than $100 between pharmacies across the street from each other!
medication	Taking medicine as prescribed or medication adherence is important.
patients	With the rising price of prescriptions and the increase in high deductible health plans and high copays, there's no limit to who [BRAND] can help.
card	[BRAND] will help you pay less than the cash price for prescription.

and labeled them, with 82 being related and 92 unrelated. We converted tweets into BERT embeddings using the same procedure as we developed the relatedness model and calculated the cosine similarities for each cluster center. The most similar n clusters have voted on the tweet based on their label and assigned the majority's decision. In order to determine n value, we have applied grid search on a separate set which gave 7 as the best f-score. Figure 1a shows the confusion matrix for the models prediction on the test set. The overall f-value of the model is 0.776. However, predicting unrelated tweets as related is a more serious problem since it will annoy people who are not interested in the product. We have adjusted the model to get a stronger majority where 5 out of 7 should vote as related. As displayed in Figure 1b, even though the weighted f-score is decreased to 0.665, the precision of not-related tweets increased to 93.5%, where we are still able to capture the 41.5% of the related tweets.



(a) (b)

Fig. 1: Confusion matrix of relatedness model using majority voting on the test set. The majority is defined as (a) 4 out of 7 and (b) 5 out of 7.

To validate the topic model, we wanted 3 participants to rank the top 5 categories for 20 different tweets from most to least relevant. We calculated the Fleiss' Kappa [8] score to measure the agreement between annotators, which gave the result of 0.643. We excluded the examples in which they could not compromise using the one-out technique until we reached a strong agreement where the remaining 15 tweets had a 0.824 Kappa score. BERT, GloVe, and GuidedLDA based models have extracted the topics with their probabilities, and we have benchmarked their overlaps to the user-given ranks. Table V shows the comparison for the models. The predicted labels are compared with the annotator labels, and they get 1 point if they match. The comparison is made with exact, +/-1, +/-2 rank order and without any order. As reported in [5], BERT predicts the topics close to the GuidedLDA model without the order, followed by GloVe. However, GuidedLDA is more successful in prioritizing the predicted topics. We used GuidedLDA to predict the topics of a given tweet.

TABLE V: Predicted top-5 topics' rank-oriented scores for a given tweet to the user-given ranks with different tolerance values.

Model	Exact	+/-1	+/-2	Any
BERT	0.6	1.6	2.7	3.1
GloVE	0.7	1.5	2.3	2.9
GuidedLDA	**2.7**	**3.3**	**3.5**	**3.7**

For each category, we have prepared propaganda phrases in advance. The paragraph generator created the paragraph by selecting one sentence from each category according to the order dictated by the topic model. We compared two techniques for the sentence selection process using two different models. The approach for selecting the first sentence relies on generating more related paragraphs to the target tweet, where it calculates the cosine similarity between the tweet's embed and candidate sentences for each topic. The second approach focuses more on creating a coherent paragraph by checking the similarity to the previously selected sentence. The embeddings of the BERT and GuidedLDA models are used to calculate the similarities. We generated paragraphs for 120 tweets, resulting in 600 sentences. We calculated the cosine similarity between consecutive sentences and counted the jumps that broke the flow. Table VI shows that BERT, with distance to the previous sentence, had 125 divergent sentences, which conflicts with the semantic integrity of the paragraph. Therefore, we proceed to BERT for a set selection of the categories extracted by the GuidedLDA topic model.

TABLE VI: Model's selected next-sentence embedding's divergency and cosine similarity to previous sentence average.

Model	Δ to	Divergent Sentences	Cos. Sim.
BERT	source tweet	35.3%	0.789
	previous sentence	**20.8%**	**0.834**
GuidedLDA	source tweet	45.2%	0.721
	previous sentence	27.6%	0.795

In order to measure the overall performance of the pipeline, we collected tweets from Twitter using predefined keywords, filtered related tweets using the relatedness model, extracted topic models using GuidedLDA, and generated a paragraph by selecting sentences using BERT embeddings. To five participants, we showed ten tweets and two pages for each of them, one of the currently used static landing pages and the personalized landing page. We wanted them to presume that they were the tweet owner and visited both pages. They aim to select which one caught their attention and give them a score from 1 to 5 for persuasion. The average of the raters' results 84% of the personalized landing pages caught their attention, whereas only 56% for the static landing page. The average score for addressing the issue is 4.28 for the personalized landing page, whereas this value is only 3.36 for the static landing page. Fleiss' Kappa score is 0.619 for attention grab and 0.667 for addressing the need, meaning substantial agreement among raters.

In order to prepare the survey without bias, displayed tweets were chosen to consist of different topics. Since the static landing page mainly mentions the solution for additional costs, the participants thought that was related to the tweet. Excluding the tweets complaining about costs resulted in a 32% attention grab and a 2.85 addressing need score for the static landing page. The results show that the proposed pipeline performs better in attracting attention and providing solutions to the needs.

IV. Conclusion

In this study, we propose an end-to-end unstructured target customer pipeline for a product or service that needs to be marketed. We have expanded the traditional online marketing funnel by adding two steps to make it workable for prospects with irregular characteristics. The process begins with a systematic procedure to identify the keywords of interest. A methodical way of distinguishing the suspects with less effort for data labeling is applied to these focal points. The personalized landing page, introduced as a final step of the process, has been shown to outperform the static landing page on attention-grabbing and addressing the need with substantial agreement among human evaluators.

References

[1] T. Ash, *Landing page optimization: The definitive guide to testing and tuning for conversions*. John Wiley & Sons, 2011.

[2] D. Bollegala, N. Okazaki, and M. Ishizuka, "A bottom-up approach to sentence ordering for multi-document summarization," *Information processing & management*, vol. 46, no. 1, pp. 89–109, 2010.

[3] B. Cui, Y. Li, and Z. Zhang, "Bert-enhanced relational sentence ordering network," in *EMNLP'2020*, pp. 6310–6320, 2020.

[4] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, "Yake! keyword extraction from single documents using multiple local features," *Inf. Sciences*, vol. 509, pp. 257–289, 2020.

[5] S. Sia, A. Dalmia, and S. J. Mielke, "Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too!," in *EMNLP'2020*, pp. 1728–1736, 2020.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[7] J. Jagarlamudi, H. D. III, and R. Udupa, "Incorporating lexical priors into topic models," in *EACL 2012* (W. Daelemans, M. Lapata, and L. Màrquez, eds.), pp. 204–213, The Association for Computer Linguistics, 2012.

[8] J. L. Fleiss, "Measuring nominal scale agreement among many raters.," *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.