

Open Domain Suggestion Mining Leveraging Fine-Grained Analysis (Workshop Paper)

Shreya Singal*, Tanishq Goel*, Shivang Chopra*, Sonika Dahiya

Delhi Technological University, Delhi

Email: {shreyashivam20, tanishqgoel, shivangchopra11, sonika.dahiya11}@gmail.com

Abstract—Suggestion mining tasks are often semantically complex and lack sophisticated methodologies that can be applied to real-world data. The presence of suggestions across a large diversity of domains and the absence of large labelled and balanced datasets render this task particularly challenging to deal with. In an attempt to overcome these challenges, we propose a two-tier pipeline that leverages Discourse Marker based oversampling and fine-grained suggestion mining techniques to retrieve suggestions from online forums. Through extensive comparison on a real-world open-domain suggestion dataset, we demonstrate how the oversampling technique combined with transformer based fine-grained analysis can beat the state of the art. Additionally, we perform extensive qualitative and quantitative analysis to give construct validity to our proposed pipeline. Finally, we discuss the practical, computational and reproducibility aspects of the deployment of our pipeline across the web.

Keywords—Suggestion Mining, Fine-grain Classification, Transformer, Oversampling

I. INTRODUCTION

A. Context and Original Scope

The rising popularity of online review forums like Yelp¹, and Tripadvisor² has spawned a popular line of study involving suggestion mining. Suggestion Mining can be defined as the classification of reviews as Suggestion or Non-Suggestion which helps firms enhance their services according to the customers' needs. Manually going through the large number of reviews and filtering out the relevant suggestions is a very cumbersome process. Therefore companies now emphasize on trying to find robust automated suggestion mining mechanisms.

Apart from general suggestions, decision makers in big corporates aim to extract focused suggestions to improve their brands or organizational practices [1]. This in turn leads to a much less studied sub-task of open-domain suggestion mining [2]. Unlike most of the work done currently in in-domain suggestion mining, open-domain mining involves a multi-tier classification wherein the model not only segregates the suggestions from the non-suggestions, but also predicts the specific domain to which the review belongs.

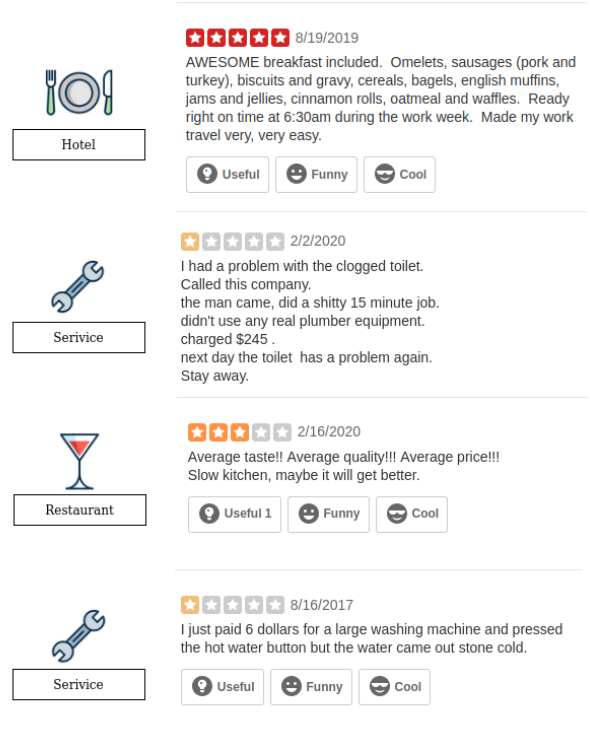


Figure 1. Manually provided multi-domain reviews on Yelp

A few examples of the kinds of reviews to be analysed are shown in Figure 1. As clear from the examples, the reviews might or might not contain valuable suggestions. Many times reviews contain positive feedback of the customers' experiences which are not very valuable to organizations in terms of their utility. The first example in the figure, although pretty detailed, doesn't add any valuable information which the hotel can make use of to improve its services. On the other hand, the other reviews are fairly clear in the problems with the respective products or services and hence give very clear indication to the respective firms of the improvements needed.

B. Challenges

Analysis of the current work done in the domain reveal several challenges to the correct classification of a review. The foremost being the sparse nature of suggestions in such a large pool of reviews. In-depth study of the various review platforms reveal a huge imbalance in the number of suggestions as compared to the number of non-suggestions. This imbalance induces unintended bias in the final classification pipeline which need to be eliminated. Apart from that, another significant challenge is the very long nature of reviews. In most of the cases, the length of a review is very big as compared to the actual suggestion presented in the review. Filtering out the required text from the review poses a hurdle to the correct classification of the respective reviews. The multi-domain nature of the data poses yet another challenge to the baseline classifiers. Correctly classifying suggestions and the domain they belong to proves to be a herculean task for those classifiers.

C. Motivation

The goal of our experiments is to gain insights into the way people formulate opinions regarding certain products or services and the differentiating dynamics of the cross-domain suggestions. Recent advancement in the field of Natural Language Processing and its application in the domain of Suggestion Mining has allowed companies to leverage customer reviews in order to enhance their products and services. The long process of manually segregating suggestions from the non-suggestions is financially taxing for the parent firms. An automated mechanism for open-domain suggestion mining significantly cuts down on these time and cost overheads. Furthermore, in most scenarios, a clear demarcation on the domain of the review is not present. A domain-independent system can therefore be effectively leveraged in such cases.

D. Contribution

The main contributions of our work are as follows:

- Propose a novel two-phase transformer-based architecture for open-domain suggestion mining using fine-grain analysis.
- Leverage Attention mechanism to achieve state-of-the-art performance in suggestion mining.
- Employ discourse marker based oversampling to handle the imbalance nature of the data.
- Use an Adapter-based transfer learning mechanism in conjugation with Transformer to improve training efficiency and performance.
- Perform an in-depth qualitative and quantitative analysis to study the practical, reproducibility and deployment aspects of our proposed pipeline.

Baseline	ID	OD	DB	AM
Ramanand et al. [3]	✓	✗	✗	✗
Brun et al. [4]	✓	✗	✗	✗
Wicaksono et al. [5]	✓	✗	✗	✗
Negi et al. (a) [6]	✓	✗	✗	✗
Negi et al. (b) [2]	✓	✓	✗	✗
Negi et al. (c) [7]	✓	✓	✗	✗
Jain et al. [8]	✓	✓	✓	✗
Ours	✓	✓	✓	✓

Table I
RELATIVE COMPARISON OF VARIOUS BASELINES. **ID**: IN-DOMAIN SUGGESTION MINING **OD**: OPEN-DOMAIN SUGGESTION MINING **DB**: DATA BALANCING **AM**: ATTENTION MODELLING

II. RELATED WORK

A. Suggestion Mining

The initial attempts at suggestion mining involved manual feature and linguistic rule modelling. Ramanand et al. [3] proposed lexicon-based rules to extract wishful comments from within suggestions. Further, Brun et al. [4] used semantic and morphological parsers to mine patterns and model linguistic rules to perform the task of suggestion mining. Following these, the focus shifted towards statistical machine learning approaches. Supervised classification was performed using Hidden Markov Models and Conditional Random Fields [5], Factorization Machines [9] and Support Vector Machines [6] to capture statistical-linguistic traits in the data. The use of neural networks for suggestion mining was first proposed by Negi et al. (a) [6]. Very recently, the SemEval challenge [10] saw several researchers perform the in-domain suggestion mining tasks using approaches like LSTM, BiLSTM, BERT [11], ULMfit [12]. All these approaches were however constrained to a single domain.

B. Open Domain suggestion Mining

Negi et al. (b) [2] first introduced the concept of open-domain suggestion mining. They used neural networks and pre-trained word embeddings to perform cross-domain training and also released the Open-Domain Suggestion Mining dataset which is subsequently used in our experiments. Building on this work, Negi et al. (c) [7] used an LSTM-based architecture to induce distant supervision in the Open-Domain Suggestion Mining task. This work was further extended by [8], [13] where they perform suggestion mining as a multi-task approach to classify the nature as well as the domain of the given reviews. Our work follows the footsteps of [8], [13] in modelling the nature of our downstream task.

C. Imbalanced Classification

Data imbalance has been a pressing issue in Natural Language Processing tasks for a long time, such as in Ghosh Chowdhury et al. [14]. This is a particularly grave issue in the

case of suggestion mining due to the sparsity of the number of suggestions in the reviews. Imbalance data can be handles using two types of techniques:

- **Random Undersampling:** Random undersampling involves randomly removing instances of the majority class to make the class-distribution even. However, this approach leads to a huge loss of data and is therefore not usually preferred.
- **Random Oversampling:** Traditionally, oversampling techniques like SMOTE (Synthetic Minority Over-Sampling Technique) [15] have been used to curb the problem of data imbalance. Further extensions of SMOTE like MSMOTE (Modified Synthetic Minority Over-Sampling Technique) [16] and MWMOTE (Majority Weighted Minority Oversampling Technique) [17] have also been proposed.

In the case of random oversampling, all the above approaches work in the euclidean space and hence, interpretability is compromised which is not acceptable in Natural Language Analysis tasks. This problem was effectively tackled by Jain et al. [8] where they propose LMOTE (Language Model-based Oversampling Technique) which leverages a Hidden Markov Model to oversample minority classes. Following this, Zhang et al. [18] used a discourse marker-based technique to perform oversampling on the Sentiment Analysis of imbalanced datasets to yield significant better performance in an interpretable fashion.

III. PRELIMINARIES

A. Problem Formulation and Notations

Let $\mathcal{R} = \{r_1, r_2, \dots, r_n\}$ be the set of n reviews that belong to d domains $\mathcal{D} = \{d_1, d_2, \dots, d_d\}$. The purpose of our experiments is to perform multi-level classification to predict the nature $n_{r_i} \in \{\text{'suggestion'}, \text{'non-suggestion'}\}$ at the first level and then the domain d_j of each review r_i such that $n_{r_i} \in \{\text{'suggestion'}\}$.

We use V to denote the vocabulary of English tokens present in our dataset. A d -dimensional word embedding e_i is associated with each token $\mathbf{v}_i \in V$ such that $e_i \in \mathbb{R}^d$. The set of all non zero embedding vectors is represented as E . Vectors are displayed in boldface. The embedding vector e_i for each token v_i can be mathematically defined as:

$$e_i = \text{Word2Vec}(v_i) \quad (1)$$

B. Transformers

Previously, LSTMs were the only way to effectively capture the semantic cues in various sentiment analysis tasks. However, [19] proposed a novel architecture Transformer, which replaced the complex recurrent computations with attention mechanism. Transformers play a pivotal role in our architecture and have the added advantage of attention-directed classification of the reviews. The attention function

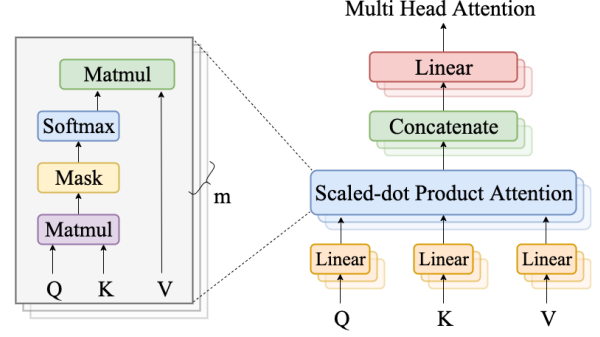


Figure 2. Internal structure of the Multi-Headed Self-Attention Mechanism in a Transformer block.

is a mapping from the d -dimensional input Values V onto the output (\mathcal{A} which is parameterised by a set of Key - Value (K, V) pairs. The Attention can essentially be defined as a weighted sum of Values $V = \{v_i | i \in \{1, 2, \dots, d\}\}$ such that,

$$\text{Attention}(K, Q, V) = \text{SOFTMAX}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

where $k, Q, V \in \mathbb{R}^d$. Furthermore, in order to capture multiple perspectives, instead of averaging the attention scores, we use multi-head self-attention mechanism. Linear projections of Queries, Keys and Values are taken in order to incorporate the different perspectives of the m attention heads which are then concatenated to form the final output.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_m) \mathbf{W}^O \quad (3)$$

where $\text{head}_l = \mathcal{A}(Q\mathbf{W}_l^Q, K\mathbf{W}_l^K, V\mathbf{W}_l^V)$, such that the project matrices $\mathbf{W}_l^Q, \mathbf{W}_l^K, \mathbf{W}_l^V \in \mathbb{R}^{\frac{d}{m} \times d}$, $\forall l \in \{1, 2, \dots, m\}$, and $\mathbf{W}^O \in \mathbb{R}^{d \times d}$. The parallel computation of these heads is analogous to performing a single attention. Hence, allowing the model to capture positional and temporal information without increasing time complexity.

Due to the very long nature of reviews and the sparse suggestions present in them, the attention layers lead to significant enhancement in the performance of our classifier over the baselines tested. Attention-directed classification allows our model to filter out the necessary parts of the review to perform further analysis.

C. Parameter-Efficient Transfer Learning

Traditionally transfer learning in Natural Language Processing can be done via either of the two techniques, feature-based transfer and fine-tuning. Both these techniques require a new set of weights for each downstream task. Adapter-based transfer learning however allows us to overcome these drawbacks by providing a parameter efficient solution.

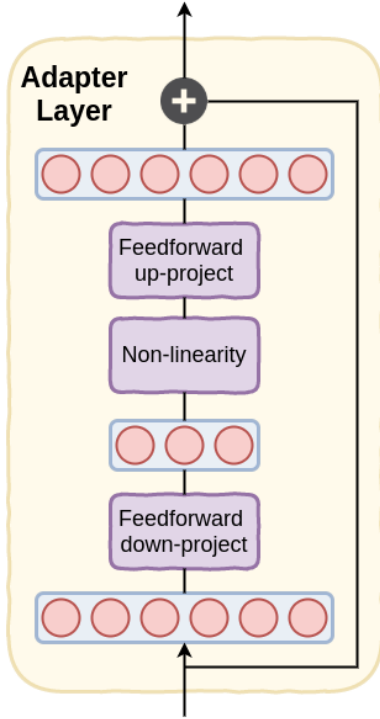


Figure 3. Internal structure of the Adapter layer used for Transfer Learning.

The internal structure of Adapter layer is shown in Figure 3. Parameter tuning involves adding these adapter layers between an existing architecture and training its parameters. Mathematically, we can represent a neural network with a function $\phi_w(x)$, parametrised by w . Adapter tuning involves defining a new function $\psi_{w,v}(x)$, where the parameters w are copied from the pre-trained network. The initial parameters v_0 are chosen such that they resemble the original function $\psi_{w,v_0} \approx \phi_w(x)$.

As shown in Figure 3, the purple layers are fine-tuned during the training process. Internally, the adapter employs a bottleneck architecture. It projects the given d -dimensional input to a smaller m -dimensional input, applies a non-linearity and then projects it back to the original d dimensions. Therefore, a total of $2md + d + m$ parameters are added per Adapter layer which is a significant improvement over training the original architecture end-to-end.

IV. DATA AND PREPROCESSING

To validate the proposed hypothesis we use the dataset curated by Negi et al. (b) [2]. The dataset consists of reviews across four domains: Hotel, Electronics, Travel and Software annotated as Suggestion and Non-Suggestion. The details of the dataset can be found in Table II. Initial analysis reveals a huge imbalance in the data. This is due to the inherent deficit nature of suggestions on the various sources from where the reviews were collected.

Dataset Identifier	Suggestion : Non-Suggestion
Travel Train	1314/3869 (0.34)
Travel Test	229/871 (0.26)
Hotel Train	448/7086 (0.06)
Hotel Test	404/3000 (0.13)
Electronics Train	324/3458 (0.09)
Electronics Test	101/1090 (0.09)
Software Train	1428/4296 (0.33)
Software Test	296/742 (0.39)

Table II
DETAILS OF THE OPEN DOMAIN SUGGESTION MINING (ODSM)
DATASET.

The following steps were involved in the preprocessing of the reviews before feeding them into the pipeline:

- **Tokenization:** A tokenizer was used to parse the reviews into tokens.
- **Spelling Correction:** Closer analysis of the text revealed a lot of spelling errors in the reviews. The spellings were corrected using a sequence matcher on the words and cross-referencing them to the words in the NLTK English corpus.
- **Lemmatization:** The tokenised tweets were then passed to the WordNet Lemmatizer to convert the words to their root forms.
- **Data Balancing:** Post the Lemmatization of the reviews, the imbalanced nature of the data is handled by the discourse marker-based minority oversampling as explained in phase 1 of the methodology.
- **Embedding Generation:** Word2vec model was then fine tuned on the balanced data to get custom embeddings for the reviews.

V. OVERALL PIPELINE

A. Phase 1: Discourse Marker based Over-Sampling

The first phase of the pipeline involved handling the imbalanced nature of the dataset. This involved a three step process as described in Algorithm 1. The steps are as follows:

- **Training Baseline Classifier:** The first step of phase 1 involves training a baseline sentiment analysis classifier. We use the architecture proposed by [8] for the same.
- **Discourse Marker Enhancement:** For each review in the domain d_i , we check for the presence of traditional discourse markers like “and”, “but”, “because”. Once a discourse marker is found, the review is split into three parts $\{s_{h_i}, m, s_{t_i}\}$ where s_{h_i} , and s_{t_i} represent the head and tail discourse respectively. Once we obtain the head and tail discourse, we perform two operations namely

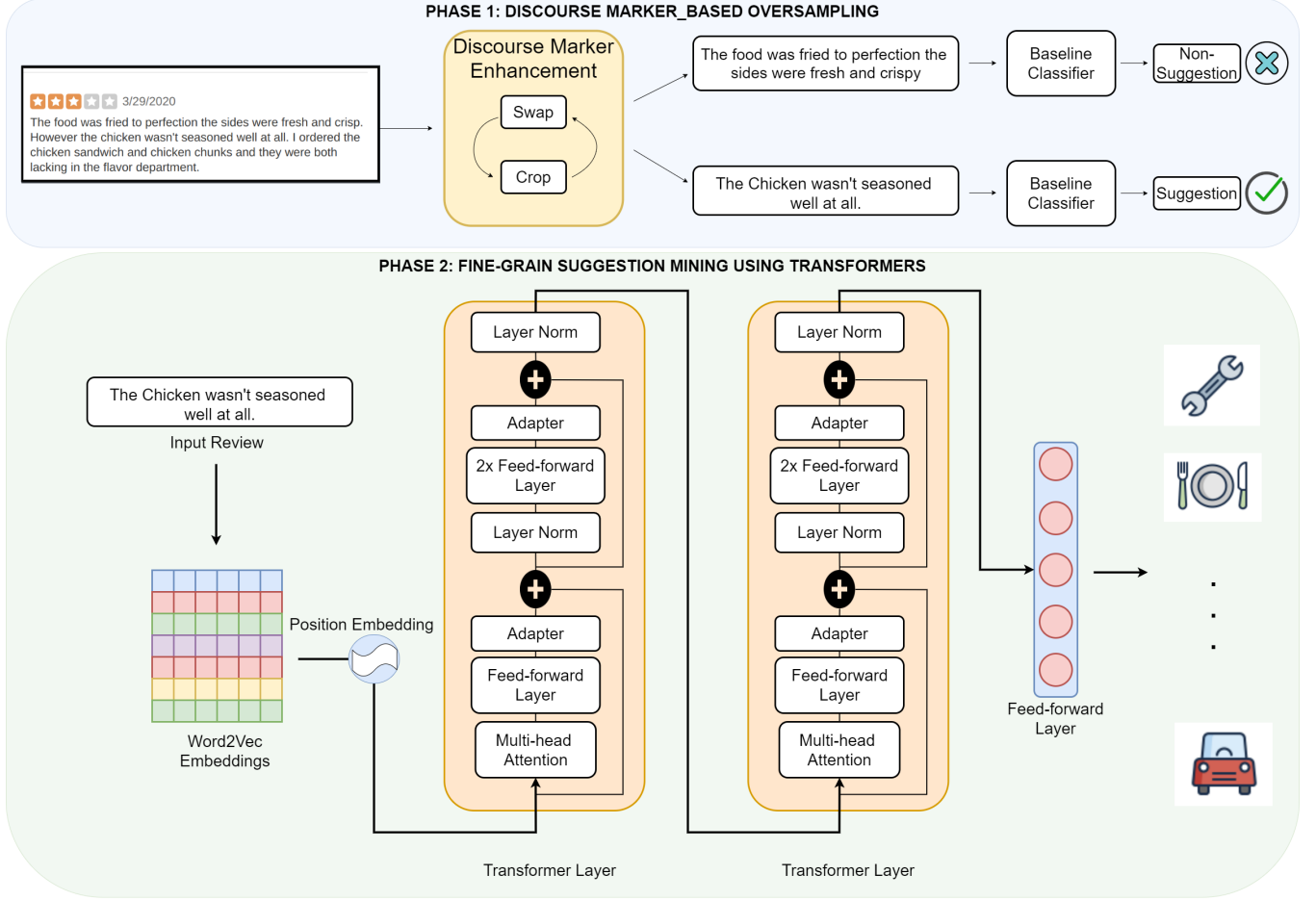


Figure 4. Overall architecture demonstrating the various phases of the proposed pipeline.

SWAP and **CROP** on the segmented review. The SWAP operation involves swapping the head and tail discourse of the review. The CROP operation however involves cropping of the head and tail discourse from the compound review.

- **Inference-based Pruning:** Once the SWAP and CROP operations are applied on the review, they are pruned using the trained baseline classifier and the new sub-reviews classified as suggestion by \mathcal{C} are added to the respective dataset.

This process is repeated for all the domains $d_i \in \mathcal{D}$ to obtain a balanced augmented dataset.

B. Phase 2: Fine-grain Suggestion Mining using Transformers

The second phase involves training the Transformer-based architecture on the augmented data. The initial step in this phase involves preserving the temporal nature of the input data. We use *position embedding* (\mathcal{P}) for this purpose. $\mathcal{P} \in \mathbb{R}^{x \times d}$ is obtained by generating a sinusoidal position

embedding with $1 \leq i \leq h$ positions for each vector-dimension $1 \leq j \leq d$:

$$\mathcal{P}_{i,2j} = \sin\left(\frac{i}{10000^{2j/d}}\right) \quad (4)$$

$$\mathcal{P}_{i,2j+1} = \cos\left(\frac{i}{10000^{2j/d}}\right) \quad (5)$$

To train the Transformer module, we include adapters with dimensionality $d = 32$ and scale the learning rate by a factor of 10 to train the newly added uninitialised parameters effectively. Gradients are accumulated over two steps to simulate larger batch sizes, which helps bring down the categorical cross-entropy loss faster.

VI. EXPERIMENTS AND RESULTS

The impact of our proposed pipeline and the oversampling strategies was tested by conducting several experiments and ablation studies on the dataset. The same test-train split was used as provided in the original dataset, the details of which can be found in Table II. Similar to the prior work including Negi et al. [2] and Jain et al. [8], we use F1 score for the

Method	Hotel	Electronics	Travel	Software	Pooled (Fine-Grain)
Baseline (SVM) [6]	0.79	0.78	0.66	0.72	0.33
Baseline (LSTM) [2]	0.79	0.77	0.64	0.75	0.68
FastText	0.85	0.85	0.57	0.78	0.81
FLAIR	0.86	0.86	0.68	0.82	0.82
Casual Transformer	0.88	0.86	0.80	0.91	0.84
SMOTE + FastText	0.85	0.83	0.73	0.87	0.85
Jain et al. [8]	0.86	0.83	0.71	0.88	0.85
SMOTE + FLAIR	0.87	0.84	0.76	0.89	0.87
SMOTE + Casual Transformer	0.89	0.88	0.78	0.88	0.90
Discourse Marker + FastText	0.89	0.87	0.78	0.89	0.87
Discourse Marker + FLAIR	0.91	0.87	0.77	0.91	0.89
Discourse Marker + Casual Transformer	0.91	0.88	0.80	0.90	0.91

Table III

PERFORMANCE EVALUATION USING F1 SCORE. IN ALL THE CASES, THE DISCOURSE MARKER-BASED OVERSAMPLING LEADS TO A SIGNIFICANT IMPROVEMENT OVER THE BASELINE CLASSIFIERS. FURTHERMORE, QUALITATIVE ANALYSIS REVEALS THE CONFUSING NATURE OF THE REVIEWS OF TRAVEL DOMAIN AS THE REASON FOR THE UNUSUALLY LOW F1 VALUES.

```

1 Input: Reviews (  $\mathcal{R}$  )
2 Pretrain a baseline suggestion classifier  $\mathcal{C}$ 
3 Let  $\mathcal{M}$  be the set of traditional Discourse Markers
4 for  $d$  in  $\mathcal{D}$  do
5   for  $r_i$  in  $\mathcal{R}$  do
6     for  $m$  in  $\mathcal{M}$  do
7       if  $\mathcal{C}(r_i) = 1 \wedge d \in r_i$  then
8          $r_i = \{s_{h_i}, m, s_{t_i}\}$ 
9         // SWAP Operation
10        Add  $\{s_{t_i}, d, s_{h_i}\}$  to the dataset.
11        if  $\mathcal{C}(s_{h_i}) = 1$  then
12          // CROP Operation
13          Add  $s_{h_i}$  to the dataset.
14          if  $\mathcal{C}(s_{t_i}) = 1$  then
15            // CROP Operation
16            Add  $s_{t_i}$  to the dataset.
17        end
18      end
19    end
20  end

```

Algorithm 1: Discourse Marker based Over-Sampling

comparison of the various models experimented upon. The results of the various experiments performed are summarised in Table III.

A. Baseline Classifiers

The initial experiments involved replicating the baseline results on individual domains. These experiments were modelled as in-domain binary classification problems. Our Transformer-based architecture, used in phase 2 of our pipeline, was able to outperform all the other machine learning and deep learning based architectures including

the ones used by Negi et al. [2], [6]. We improved on the existing baseline architectures by significant margins across all the domains of suggestion mining. Jain et al. [8] pointed out the confusing nature of suggestions in the Travel domain and the performance reduction it led to. However, our Transformer-based backbone was able to effectively mitigate this limitation.

B. Fine-grained Analysis

The experiments performed for in-domain suggestion mining were then extended towards open-domain suggestion mining via fine-grain analysis. In most of the cases, the results of the corresponding fine-grain analysis followed trends similar to those of in-domain analysis. However, in the case of simple classifiers like SVM and vanilla LSTM, the lack of sufficient parameters and inability to map complex decision boundaries. Approaches like FLAIR [20] and Transformer [19] were able to effectively learn the complex decision boundaries and yield significantly better results in the fine-grain analysis. However, the imbalance nature of the dataset was still driving the models away from optimal results.

C. Oversampling Minority Classes

Finally, the performance of our in-domain and multi-domain was enhanced using various data augmentation techniques. Minority class oversampling yielded a significant improvement in performance of our classifiers. Initially, SMOTE [15] was used as the standard minority oversampling technique. Though led to an improvement in the performance of our classifiers, we were unable to explain the enhanced results qualitatively due to the fact that SMOTE operates in euclidean space and is not interpretable. We then moved onto the Discourse marker based over-sampling

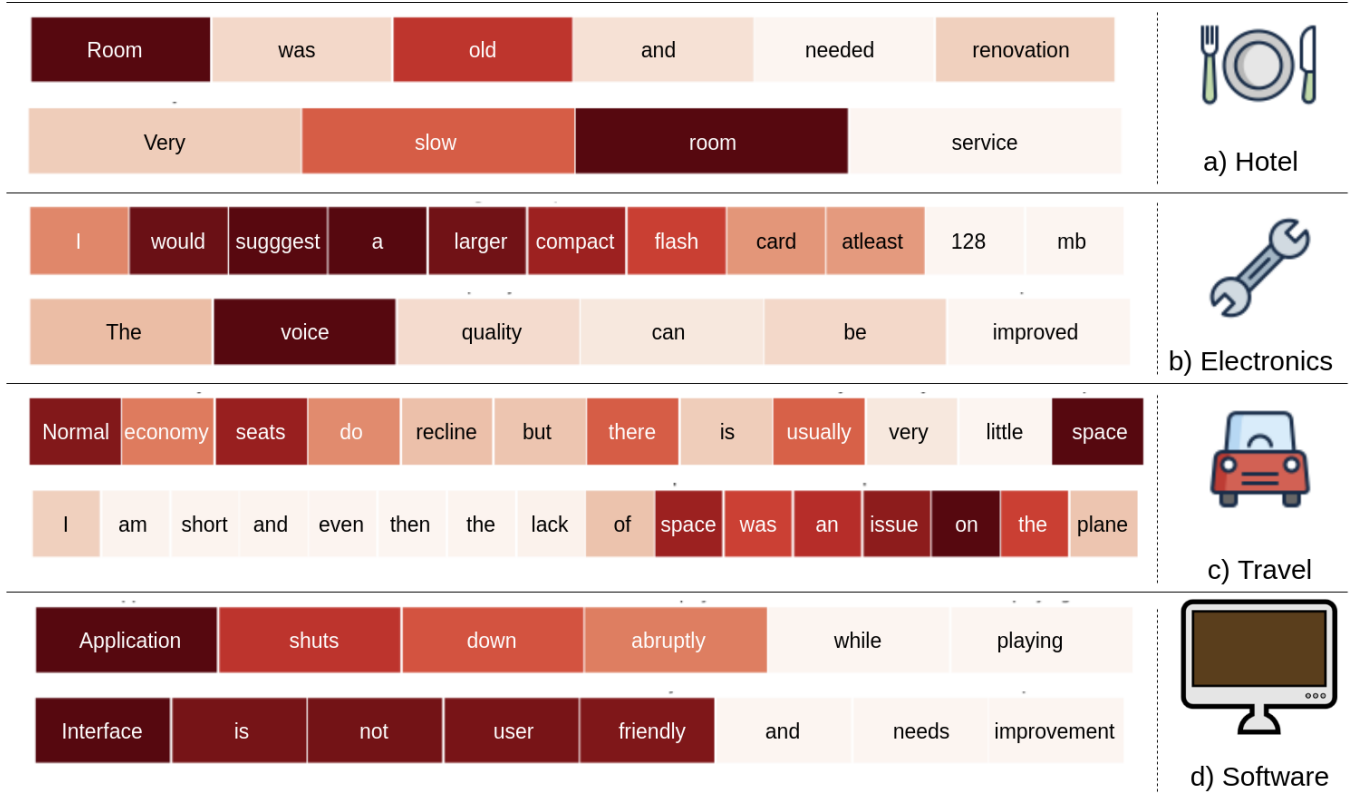


Figure 5. Qualitative Analysis of the word-wise heatmaps representing the relative attention scores for the suggestions in various domains under study.

approach explained in phase 1 of the pipeline. This approach turned out to perform comparably to SMOTE, without compromising model interpretability along the way.

VII. DISCUSSION

A. Qualitative Analysis

Due to the high dependence of our approach on the multi-head Attention mechanism in our pipeline, it was necessary to perform an extensive qualitative analysis of the Attention scores and the semantics they were capturing. The analysis involved analyzing the post-training attention inferences of complete reviews in the dataset, especially the very long reviews. Figure 5 shows the sentence based analysis along with the corresponding heatmaps. It was seen that the list of words with the highest attention values had a high overlap with the words obtained via the SAGE analysis which provided construct validity to the usage of the attention-based transformer mechanism. The words and their relative attention scores were summarised in the form of domain-specific word clouds as shown in Figure 6.

B. SAGE Analysis

We employ a Sparse Additive Generative Model (SAGE) to identify discriminating tokens between the suggestions across the various domains. The magnitude of the SAGE value of linguistic token signals the degree of its uniqueness,

and in our case a higher SAGE value denotes representativeness of the word for the respective domain. Table IV reports the top 10 tokens for suggestions in each domain. Many of the discriminating keywords for each domain were the ones with the highest attention values found using the multi-head attention mechanism of the Transformer. The SAGE analysis therefore validated the attention modelling done by the Adapter-augmented Transformers.

C. Limitations

The primary limitation of our work lies in the inability of our proposed models to effectively capture suggestions modelled as assertions in very long reviews. As shown in Figure 7, cases where the suggestions in the form of assertions like “I had to unplug fridge since it was noisy.” are presented in a very long review tend to get mis-classified. This limitation can be tackled by taking into account multiple modalities while classification of the reviews. For instance in the above example, incorporating the accompanying picture of fridge along with the textual content will help improve the performance of the model significantly. This can be done by an architecture similar to the ones proposed in Agarwal et al. [21], [22].

Hotel		Electronics		Travel		Software	
Token	SAGE	Token	SAGE	Token	SAGE	Token	SAGE
staff	1.68	zen	2.37	coach	1.96	app	-2.95
bathroom	1.67	apex	2.36	shorts	1.94	api	1.52
breakfast	1.65	g3	2.34	Egypt	1.94	developers	1.50
lobby	1.64	Canon	2.34	Ireland	1.93	buffering	1.48
renovated	1.64	Nikon	2.32	Prague	1.91	Android	1.46
beds	1.63	warranty	2.29	customs	1.89	iOS	1.43
decor	1.62	DVD	2.28	plane	1.88	emulator	1.42
spacious	1.60	viewfinder	2.27	jacket	1.87	feeds	1.41
buffet	1.58	players	2.25	dress	1.87	browser	1.40
desk	1.55	LCD	2.23	tour	1.85	notification	1.36

Table IV

TOP 10 DISCRIMINATING TOKENS USED IN THE REVIEWS OF VARIOUS DOMAINS EXTRACTED USING ALONG WITH THEIR RESPECTIVE SAGE SCORES.

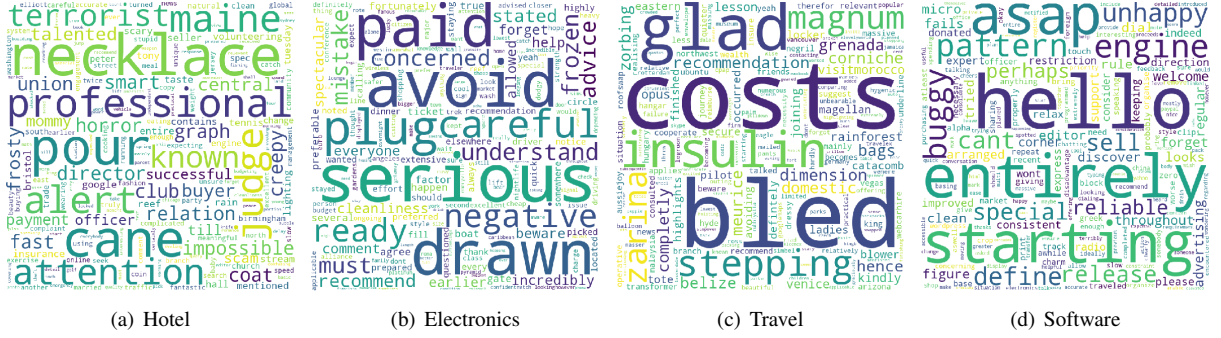


Figure 6. Attention-based word clouds for the various domains. The size of the words represent the relative attention values on the words as obtained from the multi-head self attention layer of the Transformer.

VIII. CONCLUSION AND FUTURE WORK

In this work, we propose a Transformer-based fine-grain analysis approach to tackle the task of open domain suggestion mining. We further mitigate the problem of imbalance dataset via a discourse marker based oversampling approach. Our experiments showed that our proposed pipeline outperforms the current state of the art by significant margins. We further perform extensive qualitative and quantitative experiments so as to prove the construct validity of our proposed pipeline. The future work involves incorporating reviewer profiling information into our classification models as proposed in works like Sinha et al. [23], Ghosh Chowdhury et al. [24]. Furthermore, reviews are often associated to the geographical conditions and the corresponding issues with the products or services. Incorporating geographical information of the reviews as done in Gautam et al. [25] can thus help improve the performance of our model. Another possible direction can be using mixup techniques like the ones introduced in Jindal et al. [26] instead of the discourse marker based technique to handle the data imbalance. This will lead to an increased adversarial robustness of the trained models thereby better downstream classification results.

★★★★☆ 7/11/2019

The hotel room I was in was nicely renovated however there was no bathtub just a standing shower. I had to unplug the mini fridge with how noisy it was, and the complimentary breakfast was just wheat toast, apples and oranges.



Figure 7. Review demonstrating the limitation of our approach in capturing suggestions modelled as assertions.

REFERENCES

- [1] V. Jijkoun, W. Weerkamp, M. de Rijke, P. Ackermans, and G. Geleijnse, “Mining user experiences from online forums:

- An exploration,” in *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*. Los Angeles, California, USA: Association for Computational Linguistics, Jun. 2010, pp. 17–18. [Online]. Available: <https://www.aclweb.org/anthology/W10-0509>
- [2] S. Negi, M. de Rijke, and P. Buitelaar, “Open domain suggestion mining: Problem definition and datasets,” *CoRR*, vol. abs/1806.02179, 2018. [Online]. Available: <http://arxiv.org/abs/1806.02179>
 - [3] J. Ramanand, K. Bhavsar, and N. Pedanekar, “Wishful thinking - finding suggestions and ’buy’ wishes from product reviews,” in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Los Angeles, CA: Association for Computational Linguistics, Jun. 2010, pp. 54–61. [Online]. Available: <https://www.aclweb.org/anthology/W10-0207>
 - [4] C. Brun and C. Hagège, “Suggestion mining: Detecting suggestions for improvement in users’ comments,” *Res. Comput. Sci.*, vol. 70, pp. 199–209, 2013.
 - [5] A. Wicaksono and S.-H. Myaeng, “Automatic extraction of advice-revealing sentences for advice mining from online forums,” 06 2013, pp. 97–104.
 - [6] S. Negi and P. Buitelaar, “Towards the extraction of customer-to-customer suggestions from reviews,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 2159–2167. [Online]. Available: <https://www.aclweb.org/anthology/D15-1258>
 - [7] —, “Inducing distant supervision in suggestion mining through part-of-speech embeddings,” *CoRR*, vol. abs/1709.07403, 2017. [Online]. Available: <http://arxiv.org/abs/1709.07403>
 - [8] M. Jain, M. Leekha, and M. Goswami, “A multi-task approach to open domain suggestion mining (student abstract).” in *AAAI*, 2020, pp. 13 817–13 818.
 - [9] L. Dong, F. Wei, Y. Duan, X. Liu, M. Zhou, and K. Xu, “The automated acquisition of suggestions from tweets,” in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, ser. AAAI’13. AAAI Press, 2013, p. 239–245.
 - [10] S. Negi, T. Daudert, and P. Buitelaar, “SemEval-2019 task 9: Suggestion mining from online reviews and forums,” in *Proceedings of the 13th International Workshop on Semantic Evaluation*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, Jun. 2019, pp. 877–887. [Online]. Available: <https://www.aclweb.org/anthology/S19-2151>
 - [11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *CoRR*, vol. abs/1810.04805, 2018. [Online]. Available: <http://arxiv.org/abs/1810.04805>
 - [12] J. Howard and S. Ruder, “Fine-tuned language models for text classification,” *CoRR*, vol. abs/1801.06146, 2018. [Online]. Available: <http://arxiv.org/abs/1801.06146>
 - [13] M. Leekha, M. Goswami, and M. Jain, “A multi-task approach to open domain suggestion mining using language model for text over-sampling,” in *European Conference on Information Retrieval*. Springer, 2020, pp. 223–229.
 - [14] A. Ghosh Chowdhury, R. Sawhney, R. R. Shah, and D. Mahata, “#YouToo? detection of personal recollections of sexual harassment on social media,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 2527–2537. [Online]. Available: <https://www.aclweb.org/anthology/P19-1241>
 - [15] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res. (JAIR)*, vol. 16, pp. 321–357, 01 2002.
 - [16] S. Hu, Y. Liang, L. Ma, and Y. He, “Msmote: Improving classification performance when training data is imbalanced,” *Computer Science and Engineering, International Workshop on*, vol. 2, pp. 13–17, 01 2009.
 - [17] S. Barua, M. M. Islam, X. Yao, and K. Murase, “Mwmote—majority weighted minority oversampling technique for imbalanced data set learning,” *IEEE Trans. on Knowl. and Data Eng.*, vol. 26, no. 2, p. 405–425, Feb. 2014. [Online]. Available: <https://doi.org/10.1109/TKDE.2012.232>
 - [18] T. Zhang, X. Wu, M. Lin, J. Han, and S. Hu, “Imbalanced sentiment classification enhanced with discourse marker,” *CoRR*, vol. abs/1903.11919, 2019. [Online]. Available: <http://arxiv.org/abs/1903.11919>
 - [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
 - [20] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, “FLAIR: An easy-to-use framework for state-of-the-art NLP,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 54–59. [Online]. Available: <https://www.aclweb.org/anthology/N19-4010>
 - [21] M. Agarwal, M. Leekha, R. Sawhney, and R. R. Shah, “Crisis-dias: Towards multimodal damage analysis - deployment, challenges and assessment,” in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 346–353. [Online]. Available: <https://aaai.org/ojs/index.php/AAAI/article/view/5369>
 - [22] M. Agarwal, M. Leekha, R. Sawhney, R. Ratn Shah, R. Kumar Yadav, and D. Kumar Vishwakarma, “Memis: Multi-modal emergency management information system,” in *Advances in Information Retrieval*, J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, Eds. Cham: Springer International Publishing, 2020, pp. 479–494.

- [23] P. P. Sinha, R. Mishra, R. Sawhney, D. Mahata, R. R. Shah, and H. Liu, "Suicidal - a multipronged approach to identify and explore suicidal ideation in twitter," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 941–950. [Online]. Available: <https://doi.org/10.1145/3357384.3358060>
- [24] A. Ghosh Chowdhury, A. Didolkar, R. Sawhney, and R. Ratn Shah, "Beyond hostile linguistic cues: The gravity of online milieu for hate speech detection in arabic," in *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, ser. HT '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 285–286. [Online]. Available: <https://doi.org/10.1145/3342220.3344930>
- [25] A. Gautam, P. Mathur, R. Gosangi, D. Mahata, R. Sawhney, and R. R. Shah, "#metooma: Multi-aspect annotations of tweets related to the metoo movement," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, no. 1, pp. 209–216, May 2020. [Online]. Available: <https://aaai.org/ojs/index.php/ICWSM/article/view/7292>
- [26] A. Jindal, D. Gnaneshwar, R. Sawhney, and R. R. Shah, "Leveraging BERT with mixup for sentence classification (student abstract)," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 13 829–13 830. [Online]. Available: <https://aaai.org/ojs/index.php/AAAI/article/view/7186>