

# Open Domain Suggestion Mining using Fine-Grain Analysis

**Shreya Singal: Delhi Technological University**

**Tanishq Goel: Delhi Technological University**

**Shivang Chopra: Delhi Technological University**

**Ms. Sonika Dahiya: Delhi Technological University**

# Objectives

- The primary objective of the work is 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.
- This objective is further decomposed into a multi-task problem which classifies the domain of the reviews classified as suggestions. This is also known as Open-Domain Suggestion Mining.

# Dataset

The **ODSM** Dataset by Negi et al. has been used, which has reviews across the following domains:

- Hotel
- Electronics
- Travel
- Software

The screenshot displays three Yelp review cards. Each card features a domain icon (fork and plate for Hotel, wrench for Service, martini glass for Restaurant), a star rating, a date, and a text review. Below the text are three buttons: 'Useful' (with a question mark icon), 'Funny' (with a smiley face icon), and 'Cool' (with a tongue-out icon).

Domain	Rating	Date	Review Text	Useful	Funny	Cool
Hotel	5 stars	8/19/2019	AWESOME breakfast included. Omelets, sausages (pork and turkey), biscuits and gravy, cereals, bagels, english muffins, jams and jellies, cinnamon rolls, oatmeal and waffles. Ready right on time at 6:30am during the work week. Made my work travel very, very easy.	1	0	0
Service	1 star	2/2/2020	I had a problem with the clogged toilet. Called this company. the man came, did a shitty 15 minute job. didn't use any real plumber equipment. charged \$245 . next day the toilet has a problem again. Stay away.	0	0	0
Restaurant	4 stars	2/16/2020	Average taste!! Average quality!!! Average price!!! Slow kitchen, maybe it will get better.	1	0	0
Service	5 stars	8/16/2017	I just paid 6 dollars for a large washing machine and pressed the hot water button but the water came out stone cold.	1	0	0

**Fig. 1:** Manually provided multi-domain reviews on Yelp.

# Related Work

Related work across the following three domains have been investigated:

- Suggestion Mining
- Open-Domain Suggestion Mining
- Imbalanced Classification

<b>Baseline</b>	<b>ID</b>	<b>OD</b>	<b>DB</b>	<b>AM</b>
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 1:** Relative comparison of various baselines. **ID:** In-Domain Suggestion Mining **OD:** Open-Domain Suggestion Mining **DB:** Data Balancing **AM:** Attention Modelling

# Challenges

- **Highly imbalanced data:** The suggestion:non-suggestion distribution is very low leading to biased results.
- **Very long reviews:** In most of the cases, the length of a review is very big as compared to the actual suggestion presented in the review.

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 2:** Details of the Open Domain Suggestion Mining (ODSM) dataset.

# Oversampling Algorithm

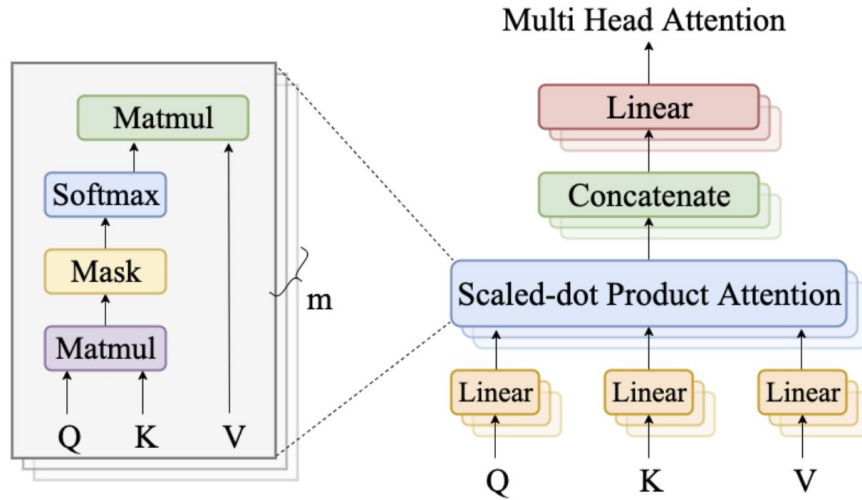
The Discourse-Marker based oversampling algorithm involves the following 3 steps:

- Training baseline classifier
- Discourse marker enhancement
- Inference based pruning

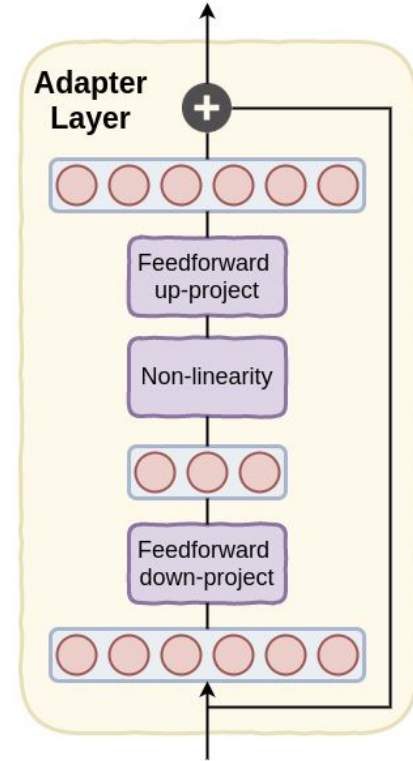
```
1 Input: Reviews ( R )
2 Pretrain a baseline suggestion classifier C
3 Let M be the set of traditional Discourse Markers
4 for  $d$  in  $\mathcal{D}$  do
5   for  $r_i$  in  $\mathcal{R}$  do
6     for  $m$  in  $M$  do
7       if  $\mathcal{C}(r_i) = 1 \wedge d \in r_i$  then
8          $r_i = \{s_{h_i}, m, s_{t_i}\}$ 
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

# Preliminaries

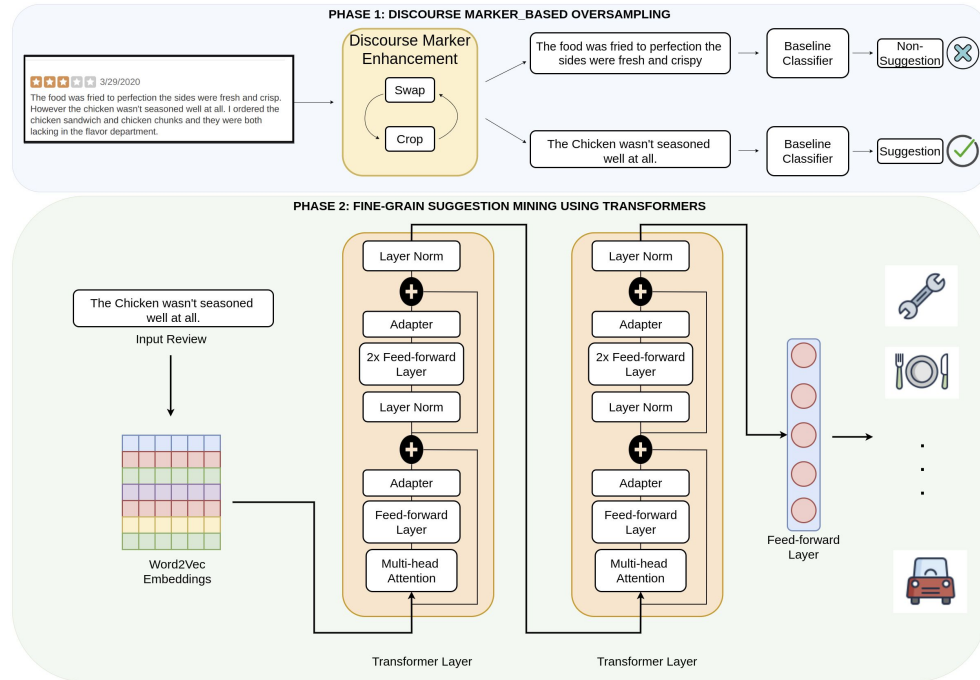


**Fig. 2:** Internal structure of the Multi-Headed Self-Attention Mechanism in a Transformer block explaining the various operations involved in obtaining the multi-head attention heat-maps .



**Fig. 3:** Internal structure of the Adapter layer used for Transfer Learning.

# Proposed Architecture



**Fig. 4:** Overall architecture demonstrating the various phases of the proposed pipeline.

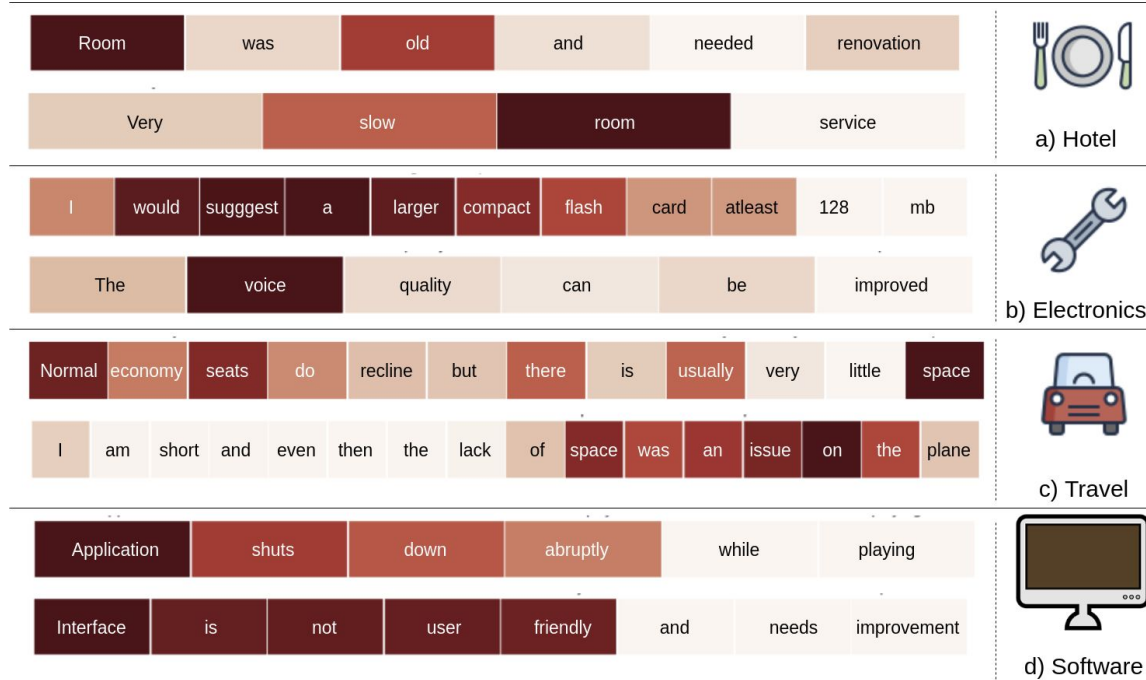


# Results

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	<b>0.80</b>	<b>0.91</b>	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	<b>0.88</b>	0.78	0.88	0.90
Discourse Marker + FastText	0.89	0.87	0.78	0.89	0.87
Discourse Marker + FLAIR	<b>0.91</b>	0.87	0.77	<b>0.91</b>	0.89
Discourse Marker + Casual Transformer	<b>0.91</b>	<b>0.88</b>	<b>0.80</b>	0.90	<b>0.91</b>

**Table 3:** 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.

# Qualitative Analysis



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

# Limitations

★★★★☆ 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.



**Fig. 5:** Review demonstrating the limitation of our approach in capturing suggestions modelled as assertions.