Extraction of Multicriteria Scores from Reviews

SDC

January 14, 2022

Abstract

This work is a preliminary study on transforming textual reviews into ratings on several criteria. Criteria are defined under the domain knowledge assumption. Sentences from reviews are assigned to a given criterion by comparing their embeddings and associated to a sentiment polarity score. Both similarity and polarity are processed using adjustable parameters, before being combined to compute review's multicriteria ratings. The pipeline, its optimization and the result obtained on Yelp dataset are described, with systematic links to the implementation (code available at https://github.com/sandrinedacol/multicriteria-ratings). Eventually, some improvement leads are detailled.

Contents

Introduction

The context of this work is product recommendation based on crowed-sourced reviews (database: item/user/review). In addition to review's global ratings, such systems could take advantage of ratings given on different criteria. But few available databases contain such information. The goal here is to extract multicriteria scores based on textual reviews. Therefore, this work stands in the Natural Language Processing (NLP) field.

1 Principle

1.1 Preliminary Hypothesis

- (H) Domain knowledge: criteria are defined *ab initio*.
- (H) A review is a set of sentences.
- (H) A sentence describes a sentiment polarity on one criterion[1].

1.2 Pipeline

1. Preprocess texts ;
2. For each sentence i,
   - find similarity s_{ic} to each criteria c ;
   - compute sentiment polarity score p_i ;
3. For each review, for each criterion, combine polarities of sentences, with weights according to similarities to criterion.

2 Dataset

The pipeline is tested on Yelp Academic Database[2], in which table *reviews*'s size is 6.9 Gb. The main attribute of interest is $reviews.text$. $business.categories$ and $reviews.useful$ are also used during dataset preprocessing (the two tables are linked by attribute $business\_id$). We also make use of the global rating $reviews.stars$.

Others datasets could be easily plugged in the `__init__()` method of class `Dataset` (module `review.py`), taking care that the code only deals with json inputs for now.

[1] The pipeline could be improved by assuming that a sentence potentially describe several criteria, see section 6.1.1.

[2] https://www.yelp.com/dataset

3 Pipeline

3.1 Dataset Pre-processing

Based on the fact that criteria depends on the type of business, only restaurants' reviews are selected (i.e. 'Restaurants' in $business.categories$).

In order to discard potential outliers, only best-marked reviews in terms of usefullness are kept (i.e. $review.useful \geq 4$).

The $\mathcal{R}$ filtered textual reviews are split into sentences using the TextBlob library[3] and concatenated into a single corpus of $\mathcal{S}$ sentences.

3.2 Criteria

3.2.1 Defining Criteria

Under the domain knowledge assumption, items are arbitrarily described by $\mathcal{C}$ criteria. Here, restaurants will be rated in terms of:

- Food
- Price
- Service
- Ambiance

Each criterion is represented by its own word and added to the corpus of sentences.

Criteria representation can be further refined by adding lexical fields to the words themselves (see section 6.1.2).

3.2.2 Sentence Vectorization

A transformer-based model (BERT) is used to create sentence-level embeddings.

The pre-trained model *'all-mpnet-base-v2'* maps all sentences to the same 768-dimensional-dense vector space (see section 6.1.3 for further improvement).

3.2.3 Similarities Computation

Similarity of each sentence i to each criterion c is computed using their embeddings, stored in a matrix of dimension $\mathcal{C} \times \mathcal{S}$.

Dot-product, cosine-similarity and euclidean distance are suitable functions for *'all-mpnet-base-v2'* embeddings [1]. For now, I implemented similarity $s \in [0, 1]$ based on:

[3] https://textblob.readthedocs.io/en/dev/
This library also offers the possibility to correct the spelling, but it does not seem to work very well...

- euclidian distance $d_{euc} \in [0, +\infty[$

$$s = \frac{1}{1 + d_{euc}} \tag{1}$$

- cosine similarity $s_{cos} \in [-1, +1[$

$$s = \frac{s_{cos} + 1}{2} \tag{2}$$

These similarites are defined by the parameter metric, with value *'euclidian'* and *'cosine'* respectively.

3.2.4 Similarities Processing

We assume that the similarity s of sentence i with criterion c is relevant only if it is larger than a threshold value s_θ. This threshold is implemented based on the similarity distribution and is defined as the quantile of an adjustable parameter $\theta \in [0, 1[$. For example, this threshold is the third quartile if $\theta = 0.75$.

At this point, each sentence is associated to a set of criteria with cardinality in $[\![0, \mathcal{C}]\!]$. Eventually, each sentence i is associated with its most similar criterion

$$c_i = \operatorname*{argmax}_c s_{ic} \tag{3}$$

(if the given similarity is greater than the threshold value).

The value of processed similarity

$$w_i = \max_c s_{ic} - s_\theta \geq 0 \tag{4}$$

can be seen as a weight describing how much a sentence actually refers to its associated criterion.

3.3 Sentiment

In the same way, a sentiment polarity score (from very negative to very positive) is computed on each sentence.

3.3.1 Raw Sentiment Scores

Pre-trained models are available with Python:

- *NLTK Vader (Valence Aware Dictionary for Sentiment Reasoning)* gives a score reflecting both polarity and strength of sentiment $p_{\text{Vader}} \in [-1, 1]$ (+ ratio of text for negative, neutral and positive)
- *Textblob* also gives a polarity + strength score $p_{\text{Blob}} \in [-1, 1]$ (+ subjectivity score)
- *Flair* is slightly different as it only gives a boolean prediction (polarity) + a prediction confidence.

For now, only Vader and TextBlob have been properly tested (parameter sentiment_analyzer with value *'NLTK* and *'TextBlob'* respectiveley; see section 6.2.1 for further improvement).

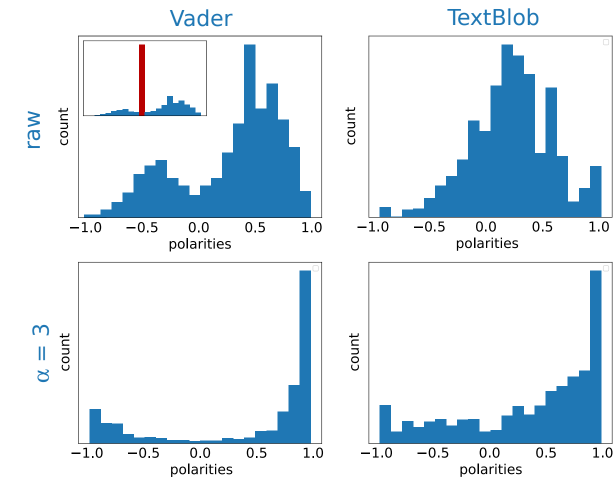

Figure 1: Sentiment scores distributions of sentences (from the first 50,000 reviews), using Vader (left) & TextBlob (right), as given by the model (up) & after refinement whith eq. (5) and $\alpha = 3$. Insert in Vader raw polarities shows the perfectly neutral polarities peak given by the model, that have been discarded.

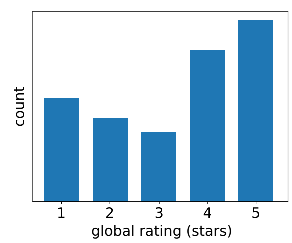

Figure 2: Review's rating distribution from the first 50,000 reviews (attribute *stars* of table *reviews*).

3.3.2 Refined Sentiment Scores

The Vader model estimates that an important number of sentences are perfectly neutral (red bar on insert of fig. 1 up left). I chose to understand that as a failure in estimating the polarity ; therfore, sentences with polarity strictly equal to 0 have been discarded.

Distributions of Vader and TextBlob sentiment scores (fig. 1 up) show that a large majority of sentences are middle-valued (± 0.5 for Vader, 0 for TextBlob). These distributions does not reflect the global ratings attributed by reviewers who tend to increase bad and good ratings in either directions (fig. 2). Hence, raw sentiment scores are stretched to extreme values by the following transformation:

$$t_p : \left| \begin{array}{ccc} [-1, 1] & \longrightarrow &]-1, 1[\\ p & \longmapsto & \tanh(\alpha p), \end{array} \right. \tag{5}$$

where α is an adjustable parameter (fig. 1 down).

Further improvement could take advantage of a second adjustable parameter (see section 6.2.2).

3.4 Multicriteria Scores

The score of a given review on a given criterion is extracted by combining polarities of all sentences of this review assigned to this criterion:

$$\sigma_{rc} = \frac{\sum_{i=1}^{k} s_{ic} \times p_i}{\sum_{i=j+1}^{j+k} s_{ic}} \tag{6}$$

is the score associated to criterion c in the r-est review composed of k sentences.

Scores are eventually converted to ratings (in star units) by:

$$t_\sigma : \left| \begin{array}{ccc}]-1, 1[& \longrightarrow &]0.5, 5.5[\\ \sigma & \longmapsto & \dfrac{5\sigma + 6}{2}, \end{array} \right. \tag{7}$$

so that ratings rounded to integers are in $[\![1, 5]\!]$ stars.

4 Tests

4.1 Pre-processing

Parameters optimization have been done using the first 50,000 reviews of the dataset, in order to minimize computation time. This number is adequate, as increasing it does not change so much parameter's optimization (fig. 3.a).

$\mathcal{R} = 1,750$ reviews remains after filtering 'restaurant' and *useful score* ≥ 4. The corpus contains $\mathcal{S} = 23,621$ sentences.

4.2 Parameters Optimization

4.2.1 Loss

As there is no labeled data, I used a pseudo-loss: multicriteria scores are averaged and compared to the global rating $\widehat{\sigma_r}$ given by the user (*reviews.stars*).

We have to keep in mind that the test results have to be treated cautiously, as it does not take into account the distribution of ratings among criteria but only their average. We can also notice that reviewers potentially do not perfectly match their global rating to the content of their text.

Average of multicriteria scores σ_r have been implemented with 2 options (parameter stars_averaging):

- *'equi'*: arithmetic mean

$$\sigma_r(\textit{'equi'}) = \frac{1}{\mathcal{C}} \cdot \sum_c \sigma_{rc} \tag{8}$$

- *'weighted'*: weighted average, taking into account the importance given to a criteria compared to the others.

$$\sigma_r(\textit{'weighted'}) = \frac{1}{\sum_c w_c} \cdot \sum_c w_c \, \sigma_{rc} \tag{9}$$

Pseudo-loss is computed with 2 options (parameter loss_type):

- *'MAE'*: mean absolute error

$$\ell_{\text{MAE}} = \frac{1}{\mathcal{R}} \cdot \sum_{r=1}^{\mathcal{R}} |\sigma_r - \widehat{\sigma_r}| \tag{10}$$

- *'rMSE'*: reduced mean squared error

$$\ell_{\text{rMSE}} = \sqrt{\frac{1}{\mathcal{R}} \cdot \sum_{r=1}^{\mathcal{R}} (\sigma_r - \widehat{\sigma_r})^2}. \tag{11}$$

4.2.2 Similarity & Sentiment Processing

Optimization of sentiment parameter α and similarity parameter θ is performed for each set of implementation options. Evolution of losses with parameters α and θ always has the shapes described in figure 3.

It seems that considering small similarities to be irrelevant does not improve sentence/criterion assignment[4].

On the contrary, sentiment scores processing is optimal for $1.0 < \alpha < 2.2$, depending on implementation options.

4.2.3 Implementation Options

Figure 4 shows the influence of (a) how multicriteria scores are averaged to be compared to the global rating, (b) the metric used to compute similarities between sentences and criteria and (c) the pre-trained model used to compute sentiment polarities. Each bar represents the smallest loss that could be obtained with the given set of options ($\theta = 0$ and optimal α).

(a) As expected, weighted scores average is slightly better than arithmetic mean.

(b) Using cosine similarity always leads to smaller loss than euclidian distance, but the difference might be seen as insignificant.

(c) NLTK Vader model outperforms TextBlob model.

[4] I can't explain why... I first thought it will be the contrary and expected to have to discuss the tradeoff between better accuracy in sentence/criterion assignment and adequate number of assigned sentences.
Nonetheless, I kept this parameter in the code, as I think it could be useful in the case of sentence assignement to multiple criteria (see secion 6.1.1).

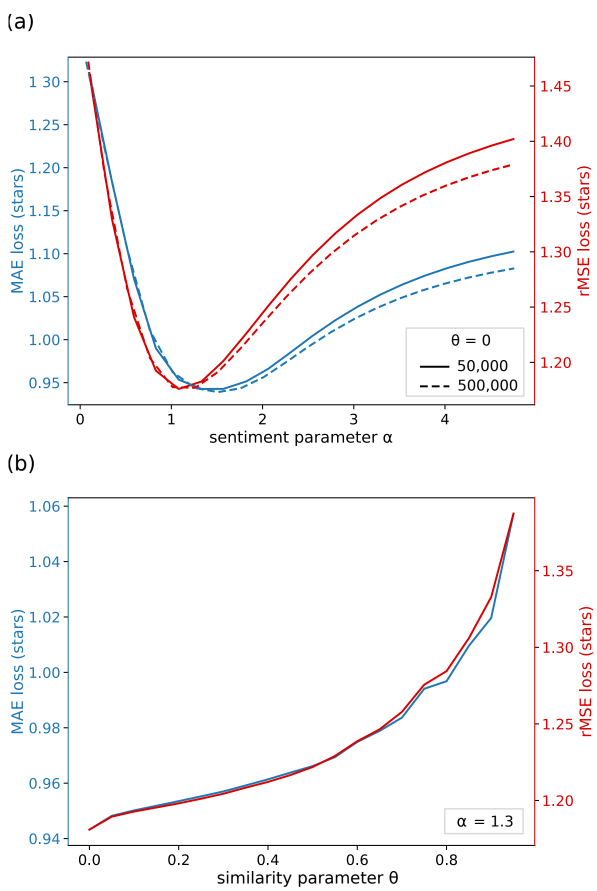

Figure 3: Parameters optimization: evolution of MAE (blue) and rMSE (red) losses with (a) sentiment and (b) similarity parameters. Other parameters are fixed: {metric=*'cosine'*, stars_averaging=*'weighted'*, sentiment_analyzer=*'NLTK'*}.

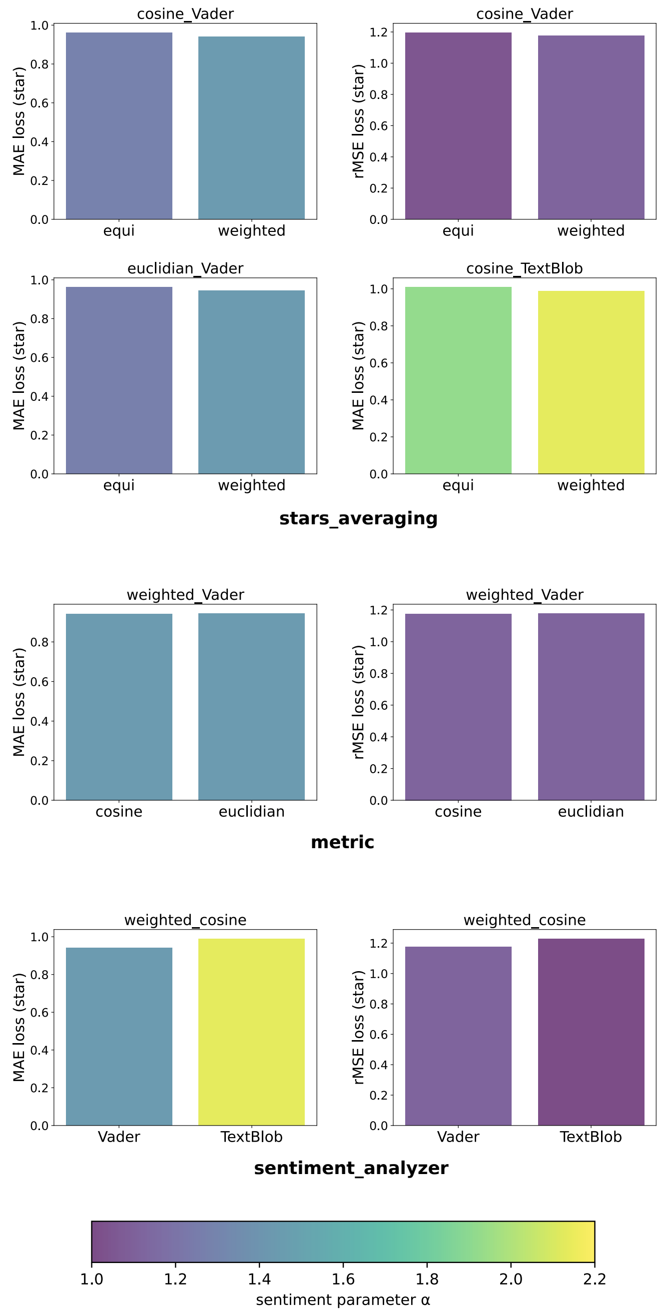

Figure 4: Influence of parameters 'stars_averaging', 'metric', and 'sentiment_analyzer' on the loss, for their optimal α (color) and $\theta = 0$.

4.2.4 Summary

The tests lead to the following parameters:

- metric $=$ *'cosine'*
- sentiment_analyzer $=$ *'NLTK'*
- $\theta = 0$.

With this set of parameters, MAE and rMSE losses are minimized for $\alpha = 1.44$ and 1.11, respectively. It could be interesting to understand the correlation between α and the type of loss (i.e. bias-variance trade-off, see section 6.2.2). As a preliminary study I made the choice of a mean value:

- $\alpha = 1.3$.

5 Results

An illustration of the process is given on appendix C, providing the intermediate computations on a given review.

5.1 Pre-processing

$\mathcal{R} = 355,780$ reviews remain after filtering the whole dataset, totaling up $\mathcal{S} = 5,082,006$ sentences.

5.2 Criteria

Corpus embeddings with *'all-mpnet-base-v2'* model is the most expansive step in terms of time computation (about 24h on my labtop, see section 6.1.3). The cache file sizes 15.6 Go.

Figure 5 .a shows the distribution of normalized cosine similarities (2), computed between sentences and criteria embeddings.

Each sentence is assigned to only one criterion and the corresponding similarity processed into a weight (4). The weight distributions (fig. 5.b) and the 2D projection of sentence embeddings colored by their assigned criterion (fig. 6) show that a majority of sentences are associated to 'Food', then 'Ambiance', then 'Service', then 'Price'. This order is consistant with mean values of raw similartities for each criterion (fig. 5.a).

5.3 Sentiment

Figure 7 shows the distribution of processed polarities.

5.4 Multicriteria Scores

The pipeline tends to overestimate the ratings (fig. 8). Considered as poorly estimated, 130,655 reviews with absolute error $|\sigma_r - \widehat{\sigma_r}| \geq 1$ star (red) are discarded, being 37%.

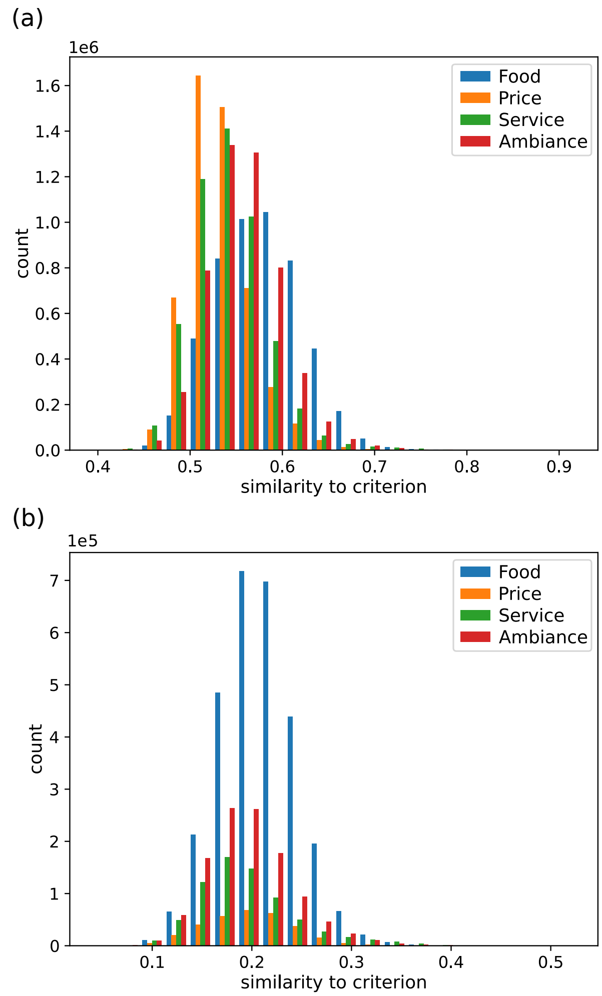

Figure 5: Distributions of (a) raw similarities and (b) processed weights of sentences to criteria.

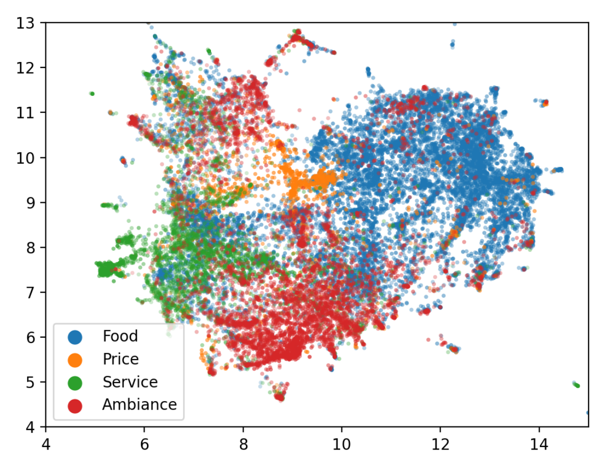

Figure 6: 2D projection of sentence emmbeddings on the 1,750 first reviews, colored by their associated criterion (transparency according to weight).

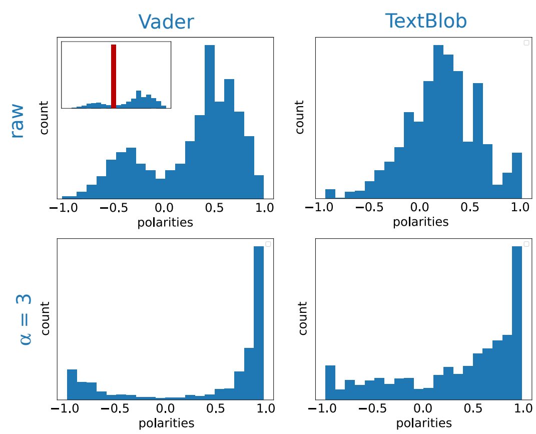

Figure 7: Distribution of processed ($\alpha = 1.3$) polarities of sentences from NLTK Vader polarities.

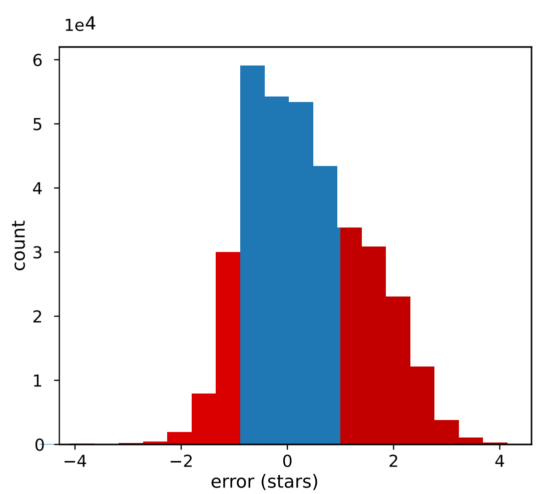

Figure 8: Error distribution on averaged multicriteria scores. Red reviews are eventually discarded.

6 Improvement Perspectives

Here are some leads I would have followed if I had more time on this work.

6.1 On Criteria

6.1.1 Releasing the Constraint '1 sentence = 1 criterion'

A sentence of the review described in appendix C

"The value is tremendous here for the amount of succulent meat and I've paid four times as much for less of an experience."

shows that some sentences can clearly refer to more than one criterion (here, 'Food' and 'Price').

Some attempts have been made to allow multiple criteria on each sentence. Despite better results regarding pseudo-loss optimization, I considered them as irrelevant. Indeed, the distribution's variance of all similarities is small (fig. 5), leading to criteria weights almost uniform on each sentence.

Therefore, I wanted to explore the avenue of comparing similarities to criteria normalized within each sentence. Normalized similarity of sentence i with criterion c_j could be defined as:

$$s'_{ic_j} = \frac{s_{ic_j} - \min\limits_{k\in[\![1,\mathcal{C}]\!]} s_{ic_k}}{\max\limits_{k\in[\![1,\mathcal{C}]\!]} s_{ic} - \min\limits_{k\in[\![1,\mathcal{C}]\!]} s_{ic}} \in [0,1]\,. \qquad (12)$$

The number of criteria $\mathcal{C}_i$ a sentence i should be assigned to is illustrated on figure 9 in the cases $\mathcal{C}_i = 1$ (here, 'Ambiance') and $\mathcal{C}_i = 2$ (here, 'Food' and 'Ambiance').

This could be implemented in 2 ways:

- In a textbook sentence, the sum of normalized similarities is equal to the number of criteria that the sentence refers to. So,

$$\mathcal{C}_i = \left|\begin{array}{lll} n_i, & \text{if} & \mathcal{C}_i - n_i \leq 1/2 \\ n_i + 1, & \text{if} & \mathcal{C}_i - n_i > 1/2 \end{array}\right., \qquad (13)$$

  where $n_i = \left\lfloor \sum_j s'_{ic_j} \right\rfloor$ is the sum's floor.

- Adding an adjustable parameter ϵ will probably increase the accuracy, despite larger testing effort. In that case, the condition for a criterion c_j to be shortlisted for a sentence i can simply be:

$$s'_{ic_j} \geq 1 - \epsilon. \qquad (14)$$

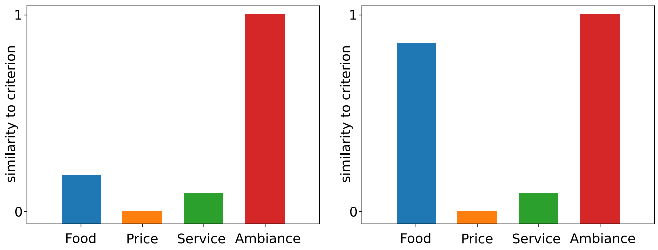

Figure 9: Qualitative shape of 1-criterion sentence (left) and 2-criteria sentence (right)

6.1.2 Refining Criteria Representation

One idea is to use a class-based algorithm derived from TF-IDF, as done in BERTopic [2]. The lexical field of a criterion would be re-defined by the top-N words, in terms of importance, in all sentences associated to the given criterion.

Here are the steps:

1. run the pipeline with criteria defined by their own word, as previously done ;
2. create $\mathcal{C}$ documents, each of them containing all sentences of the corpus that have been assigned to a given criterion ;
3. compute c-TF-IDF of all words present in the corpus[5]. The result is a matrix of dimension 'number of words'$\times\mathcal{C}$;
4. extract top-N words for each document,
5. refining by hand the words lists[6];
6. re-run the pipeline, with criteria defined by their words list.

6.1.3 Changing Sentence Embeddings

I chose *'all-mpnet-base-v2'* as it is the best-ranked model for sentence embedding in terms of performance, but it could be interesting to test other models [1].

In particular, lowering the dimensionality of vector space and increasing the vectorization speed could be useful.

6.2 On Sentiment Polarities

6.2.1 Training a Sentiment Polarity Model

I tried to run the pipeline with Flair (sentiment_analyzer=*'Flair'*), which gave bad results. But it seems that we could increase efficiency by training the model on the corpus of sentences (see tutorial in [3]).

Broadly speaking, training the sentiment model on the dataset of interest might greatly improve the final result, regardless of the model used.

6.2.2 Refining Polarities Processing

It could be interesting to understand the correlation between α and the type of loss (i.e. bias-variance tradeoff).

Starting point:

1. rMSE favors lower variance ; MAE lower bias ;
2. Optimal α is larger for MAE.

$\longrightarrow$ Does $\alpha > 1$ increases variance?
$\longrightarrow$ Do we want to favor variance minimization and add a supplementary parameter β to balance the bias?

$$(5) \Rightarrow p' = \tanh(\alpha p) + \beta \tag{15}$$

An other possibility relies on the assumption that criteria ratings and averaged criteria ratings have the same distribution. The idea is to optimize the polarity processing in a first step, the loss being, for example, an Earth Mover's Distance between polarity distribution and global rating distribution.

References

[1] Pretrained models for bert. www.sbert.net/docs/pretrained_models.html, page web.

[2] Maarten Grootendorst. Topic modeling with bert: Leveraging bert and tf-idf to create easily interpretable topics. https://towardsdatascience.com/topic-modeling-with-bert-779f7db187e6, page web.

[3] leoweber. Training your own flair embeddings. https://github.com/flairNLP/flair/blob/master/resources/docs/TUTORIAL_9_TRAINING_LM_EMBEDDINGS.md, page web.

[5] https://maartengr.github.io/BERTopic/api/ctfidf.html

[6] With N in the range of 100, I think it is acceptable.

Appendices

A Notations

notation	description	ref.
i	sentence	
c	criterion	
c_i	most similar criterion assigned to sentence i	(3)
$\mathcal{C}_i$	number of criteria assigned to sentence i	(13)
$\mathcal{C}$	number of criteria	
$\mathcal{R}$	number of reviews	
$\mathcal{S}$	number of sentences	
SIMILARITIES		
s	similarity between 2 embeddings	
s_{ic}	similarity between sentence i and criterion c embeddings	(1) (2)
$s_{\cos}$	cosine similarity	(2)
s_θ	sentence/criterion similarity threshold defined by parameter θ	(4)
s'_{ic_j}	normalized similarity of sentence i with criterion c_j	(12)
d_{euc}	euclidian distance	(1)
n_i	floor of sum of normalized similarities of sentence i to all criteria	(13)
SCORES / RATINGS		
p	sentiment polarity score	
p_i	sentiment polarity score of sentence i	3.3
σ	multicriteria score/rating	
σ_{rc}	estimated score/rating for criterion c in review r	(6)
σ_r	averaged estimated rating of review r	(8) (9)
$\widehat{\sigma_r}$	labeled rating of review r (in *review.stars*)	(10) (11)
w_i	weight of sentence i within its review	(4)
ADJUSTABLE PARAMETERS		
α	parameter transforming raw sentiment polarity scores	(5)
β	parameter balancing the bias in sentiment polarity score distribution	(15)
ϵ	parameter defining a threshold in normalized similarity	(14)
θ	parameter defining a threshold value for sentence/criterion similarities	3.2.4
LOSS FUNCTIONS		
ℓ_{MAE}	mean absolute error	(10)
ℓ_{rMSE}	reduced mean squared error	(11)

B Parameters

The table below summarizes all implementation parameters, to be set in the file parameters.yaml.

name	values	description
mode	'run'/'test'	run the pipleline or compute the loss for multiple values of parameters α or θ
dataset_name	'Yelp'	name of the dataset. Others datasets could be plugged in the __init__() method of class Dataset (module review.py).
n_reviews	$\mathbb{N}^{+*}$	used to decrease computation time. process only the first n_reviews of the dataset (before filtering)
criteria	$\{\langle \textit{criterion name} \rangle : [\langle \textit{word} \rangle]\}$	definition of domain knowledge: dictonary of items with key = name of criterion and value = list of words representing it
metric	'cosine'/'euclidian'	metric used to compute similarities between sentences embeddings and criteria embeddings. Others metric could be plugged in method compute_raw_similarities() of class Criteria (module criteria.py)
sentiment_analyzer	'NLTK'/'TextBlob'/'Flair'	pre-trained model used to compute sentiment polarity of sentences
parameter	'threshold'/'alpha'	if mode = 'test', choose the tested parameter
min_value	$\mathbb{R}$	minimum value of tested parameter
max_value	$\mathbb{R}$	maximum value of tested parameter (not included)
n_values	$\mathbb{N}^{+*}$	maximal number of values tested
significant_digits		precision on tested values : number of decimal digits
loss_type	'MAE'/'rMSE'	type of loss computed: mean absolute error or reduced mean squared error. Others losses could be plugged in method evaluate_quality() of class MultiCriteriaScoresExtractor (file main.py)
stars_averaging	'equi'/'weighted'	type of average used to compute global rating from multicriteria scores: arithmetic mean or weighted average
theta	$[0, 1]$	parameter used to filter lowest similarities
alpha	$\mathbb{R}^{+*}$	parameter used to process raw sentiment polarity scores given by pretrained model
error_max	$[0, 4[$	maximal tolerable error on average of multicriteria scores. Discards reviews with bad estimation of scores.

C Example

Let's have a look on the 12^{th} review, rated 4 stars.

C.1 Raw Similarities

The table below shows similarites computed for each sentence.

Underlined numbers are the similarities that should be sentence's maxima, according to a "human" understanding of the sentence. Colored numbers are the maximal similarities according to the model (green if in accordance with the latter ; red if not).

#	sentence	s_{Food}	s_{Price}	s_{Service}	s_{Ambiance}
1	Finally, a non-threatening and reasonably priced Brazilian BBQ!	0.604	0.558	0.530	0.562
2	You have two options here: all-you-can-eat for $12 and change or pay by the pound.	0.647	0.650	0.571	0.536
3	The unlimited option was definitely the way to go.	0.529	0.571	0.575	0.555
4	Now, buffets can be tempting, but keep in mind that the best stuff comes right to your table ON A SWORD.	0.678	0.540	0.559	0.576
5	We filled up our plates with samplings from the buffet, from various salads (the potato salad was my favorite), rice and beans, and a gigantic french fry.	0.691	0.519	0.557	0.568
6	The fun begins when the skewers of meat begin to arrive, ranging from chicken wrapped in bacon to pork ribs, chicken sausage, and top sirloin.	0.656	0.544	0.541	0.536
7	There's a definite joy in grabbing the meat as it falls off from being cut and tasting as much as possible.	0.630	0.516	0.509	0.520
8	Be sure to pace yourself; this place is extremely laid back, there's no card to flip over to tell them to bring meat and they'll keep serving you for as long as you'd like.	0.609	0.507	0.565	0.553
9	A special bonus was the live music on Friday night – a singer with a beautiful voice and a talented guitar player – offering up traditional songs, as well as their gorgeous arrangements of pop songs like Rock with You, Come Away With Me, and Hotel California.	0.531	0.533	0.518	0.557
10	Highlight of the evening was absolutely Tears in Heaven.	0.520	0.535	0.526	0.572
11	There's ample parking behind the restaurant and it's located conveniently in downtown Everett.	0.531	0.512	0.533	0.548
12	The value is tremendous here for the amount of succulent meat and I've paid four times as much for less of an experience.	0.589	0.614	0.530	0.588
13	I will definitely be back.	0.536	0.508	0.533	0.546

The result is fine (bad evaluations: 23%).

Notice that the process does not provide the possibility to successfully assign 2 types of sentences :

- global opinion, like

  "I will definitely be back."

- opinion on a criterion that have not been included under the domain knowledge assumption.
  For example,

  "There's ample parking behind the restaurant and it's located conveniently in downtown Everett."

  clearly refers to something like 'location' or 'accessibility'.

C.2 Multicriteria Scores

The table below shows the values used to compute multicriteria scores (6).

relative weight	polarity	sentence
FOOD		
0%	0	Finally, a non-threatening and reasonably priced Brazilian BBQ!
28%	0.767	Now, buffets can be tempting, but keep in mind that the best stuff comes right to your table ON A SWORD.
0%	0	We filled up our plates with samplings from the buffet, from various salads (the potato salad was my favorite), rice and beans, and a gigantic french fry.
26%	0.671	The fun begins when the skewers of meat begin to arrive, ranging from chicken wrapped in bacon to pork ribs, chicken sausage, and top sirloin.
24%	0.642	There's a definite joy in grabbing the meat as it falls off from being cut and tasting as much as possible.
22%	0.459	Be sure to pace yourself; this place is extremely laid back, there's no card to flip over to tell them to bring meat and they'll keep serving you for as long as you'd like.
PRICE		
54%	-0.133	You have two options here: all-you-can-eat for $12 and change or pay by the pound.
46%	0.415	The value is tremendous here for the amount of succulent meat and I've paid four times as much for less of an experience.
SERVICE		
100%	0.480	The unlimited option was definitely the way to go.
AMBIANCE		
33%	0.851	A special bonus was the live music on Friday night – a singer with a beautiful voice and a talented guitar player – offering up traditional songs, as well as their gorgeous arrangements of pop songs like Rock with You, Come Away With Me, and Hotel California.
36%	0.639	Highlight of the evening was absolutely Tears in Heaven.
0%	0	There's ample parking behind the restaurant and it's located conveniently in downtown Everett.
31%	0.480	I will definitely be back.

D Output

Output in json file. Here are the 10 first reviews:

"review_id": "LBJJWJ0uNlIybMfPnQGS0A",
"user_id": "bUHweiErUJ36WGeNrPmEbA",
"business_id": "6Hm2FmfLcU_M91TrZI5htA",
"stars": 5.0,
"useful": 5,
"funny": 3,
"cool": 3,
"text": "I loved everything about this place. I've only been once but I keep meaning to go back as it was so great (just a bit out of my way). I went with a fairly large group so we'd all ordered something different (back when it was $6 for everything, it looks like they've changed that now). At the time I got the chickpea fries, which were delicious but watch out for the fiber because I had to sit out of an entire board game. The board game selection was great, and I love that they serve mead (though I didn't get any at 12 pm haha). The molten lava chocolate cake was so, so good. Seriously. The real topping to the experience (other than the great service and the fact that they put lemons in their complimentary ice water) was that they were playing Siouxsie and the Banshees in the background! That + board games + good food + mead = I'm a fan forever.",
"date": "2014-02-05 21:09:05",
"multicriteria_stars": {"food": 4, "price": 4, "ambiance": 5 }

"review_id": "3c4LFWiVUHPDQzAd8uxc3A",
"user_id": "jySmPCkEkJR3cWJlkEs9cw",
"business_id": "ZW7aI5FO_3q_vSzI4_zx-Q",
"stars": 5.0,
"useful": 5,
"funny": 2,
"cool": 4,
"text": "Definitely 5 stars for the donuts. Our family has been coming to Dandy Donuts for years! Their donuts can't be beat and are a must-try. The same Cambodian family has owned this shop for years. They are kind, hard-working folks that put out a consistently fresh and delicious line of donuts as well as croissants and apple fritters their customers love. Subs are also good. I notice they've added a new low-price lunch special of hot dog, chips and drink. The surroundings are definitely ḧole-in-the-wall,̈ but go and try the donuts.",
"date": "2013-08-15 14:47:40",
"multicriteria_stars": {"food": 5, "ambiance": 4}

"review_id": "oNNTEpc2PmB4w_vy9pbsJw",
"user_id": "s4NgvdIfBH3UQdccWCyARg",
"business_id": "IdXHHEUH4ebcxdRxCo3JNw",
"stars": 3.0,
"useful": 7,
"funny": 4,
"cool": 1,
"text": " "A Bit Embarrassed" I'm not exactly sure where to start with this update, other than saying that my intention was absolutely NOT to get my - or everyone's- meal comped last night. Yes, I'm doing the Eat On $30 Challenge. Yes, I was swayed by another Yelper to go to Rosebud, even while fully aware that I didn't want the $5 Mystery Monday meal (biscuits and gravy with a fried egg - looked good when I saw James H.'s, Emily R.'s and Alain L.'s orders, but just not something I really like at all) so I was probably getting myself into a grossly-tempting land of foodiness, given all the other things I know or think sound goody good on their menu. I went anyway!I wanted the company and I'm so used to dining out, I couldn't resist, I suppose!! Long story short, the service was Capital-B Bad. BAD. Slow. Inattentive. As Lauren S. noted, no refills on water, plus other drinks were slow and things like James H.'s grits side and Emily R.'s fries were either forgotten (the former) or delayed and without steak sauce (the latter). Then, the first few folks to cash out clearly asked for separate checks, fiiiiiiinally got one that was all together, had to ask for it to be corrected . . . only to find that each one had a 20% s̈ervice chargeẗacked on. Huh?! We all sat in the bar area, the group just kinda grew as people filtered in, and nowhere on the menu does it mention anything about an auto-add for b̈ig parties.B̈esides, I thought it was insulting because pretty much everyone at the table goes there a lot or at least often-enough to have talked with Ron and be a pretty familiar f̈ace-̈ and we're Yelpers, so we're gonna tip you (oh, and PS? Not everyone ordered the $5 special, like Kathleen M. with her shrimp-n-grits or Lauren S. with her breakfast bowl . . . no one was gonna stiff them!). Still, our first three dining companions left without much fanfare, and I figured, ḧey, bad night.T̈hen

. . . the rest of us got ready to leave and a) our server took for-freakin'-ever to bring checks, and b) (the last straw in my book) dropped Emily R.'s to-go box on the floor, picked it up and resealed it, and dropped it in front of her like nothing had happened. Oh, did I mention that while we were waiting on our checks, we got to see several other servers (note: not ours) eating some of Ron's wings at the bar with r̈egulars(̈as he later described them)?? Yeah, it was a BAD night at Rosebud. So when our server finally came back with the checks, I spoke up. My intention was to let him know that we were n̈onplussedänd, as I stated to him, he should remove the äutomatic 20% service charge if he'd included it.Ä couple minutes later, Ron came out to c̈hat.Ï have to say . . . I'm ä little bit embarrassedb̈ecause I'm not the diner who wants her meal comped for any little reason, or, really, ever, unless it's a gracious gesture on management's part. And when Ron came over, it felt, well, like our server had tattled to Daddy . . . and that Ron didn't really want to hear what had happened. He said most of the right things, and ägreedẅith comments, but suddenly I felt like I was the only one speaking up, he didn't *really* care, he noted he'd worked 8̈0 or 90 hoursïn the last week (ok, but you're the owner, right? The resto business is nothing close to easy . . . what does that have to do with our conversation??), and he tried to defend his server as well as justify the service charge . . . and then he said Ï can see where this is going . . . let me just take care of all of your dinners.Ẅell, sh*t. Not really what I wanted to have happen. Honestly, it just seemed like an easy out to get *me* (us) to STHU. And then . . . how do I tip on a non-dinner? Especially one where I just had Bloody Mary's once noting how sloooow everyone's food orders came, despite being tempted? And do I just take it as a f̈reebieg̈iven my week of Eat On $30 challengedom?? It was all too much. It seemed we all left awkwardly (not to mention that Kathleen, James, and Lauren had to pay, including a s̈ervice chargeänd we didn't) and while Emily ẗipped for mes̈ince I didn't eat, but she knew I felt weird, I left with the overall thought . . . Ḧuh?? What just happened??M̈onday. I will chalk it all up to Monday. I really like Rosebud and want to continue going often and recommending it to folks, but I left feeling -and still feel - like I need to *digest* my most recent experience . . .",
"date": "2009-10-13 22:20:10",
"multicriteria_stars": {"food": 3, "price": 2, "service": 3, "ambiance": 4}

"review_id": "Z89mKeSpOD8BrfrnZ9B8Xg",
"user_id": "kV7kmmBw_8IPLyI5M6wgTw",
"business_id": "4cgE4DUDSH3CjqsRwX_LOA",
"stars": 5.0,
"useful": 4,
"funny": 1,
"cool": 2,
"text": "I was in Austin recently and searched for a place to eat on this fine website called Yelp. We were looking for hard tacos and came upon Evita's Botanitas and the menu looked pretty great. So we GPS'd ourselves there and found ourselves in a small strip mall, nothing fancy. And the food did not disappoint!! First of all, the salsa they give you is above and beyond. They bring a two-tiered production with a basket of fresh tortilla chips and 5 or 6 different salsas, all of which were delicious and not at all 'omg my mouth, it's on fire' I ordered a house margarita (rocks with salt please!) which was ass kicking in a good way. For my dinner I ordered the vegetarian quesadillas, which I admit were served to me in a unfamilliar way. Maybe I was just Dumb White Girl but when they brought my food I thought they brought soft tacos instead of a quesadilla. Granted, I'd ordered the corn tortillas rather than flour, and it looked like a plate with several soft tacos. And oh man, were those some seriously delicious quesadillas! I don't care what they looked like, they tasted like heaven. Our waitress was really friendly and I would highly recommend eating here. I will make it a point to come back the next time I'm in town.",
"date": "2009-10-14 01:45:09",
"multicriteria_stars": {"food": 4, "ambiance": 2}

"review_id": "zdPkKh6pGnNEH21AeiKBqg",
"user_id": "suBpZTEXih253-0s_2GWCw",
"business_id": "G61xj19TPnbzB7oQuoWTgQ",
"stars": 3.0,
"useful": 4,
"funny": 1,
"cool": 0,
"text": "Let me start by saying this place is lucky my girlfriend isn't the one reviewing it because she is far harsher. Decor works. You feel like you're in a down home bbq joint. Mission accomplished. I liked the menu and its simplicity as well. (although there was some confusion on the combo in that ribs and wings were not an option although they appeared to be.) My first issue was with the burnt ends. I came here for the burnt ends, because FYI, it is very hard to find burnt ends around Atlanta. this place offered them so I came. They were sub par in my opinion. They were 1.5 inch cubes of dry brisket black on all sides. This is not the burnt ends I know and love. Definite disappointment when that was my sole purpose for coming there. Other lackluster food: the brunswick stew was just okay... tasted and looked like something you would get out of a can. The potato salad had pretty much no flavor and the potato seemed slightly undercooked.

The good food: I really enjoyed the okra. I love okra, and whenever I order fried okra I assume it's going to be sliced and in corn meal. However, they fry it whole and in almost a tempura type batter. Very different and very tasty. We ordered some wings and they were very good. They're smoked and then flash fried so they are flavorful and juicy. However, I recommend getting them in a sauce because the dry rub just isn't enough. The pulled pork was also very good, but a little dry like it was pulled long before I ordered it. On top of food shortcomings, there were several service issues. The server, although very nice and polite,. was the only server working the lunch shift. So she was obviously overwhelmed by five or six tables. On top of that, our bill was mixed with the table next to us so when I got my bill it was over 80 dollars for the two of us. She fixed it, but it caused a longer delay than there already was. The service issues really weren't her fault, but management for only having one server, that was also acting as the bartender. If I'm in the area and needing some bbq I would definitely come here again. Just a disappointment when I came out of the way for burnt ends that were not great.",
"date": "2013-10-24 02:03:22",
"multicriteria_stars": {"food": 4, "price": 4, "service": 4, "ambiance": 4}

"review_id": "hPtBn9NcTDogbPhbkTNLbA",
"user_id": "wekw0bbIv8o61WkTg7GPkg",
"business_id": "jhFi0oxdyiyCYK7i-CW59A",
"stars": 4.0,
"useful": 6,
"funny": 1,
"cool": 1,
"text": "I ate here recently on the recommendations of some Muslim brothers at the Islamic Society of Boston (beautiful masjid). They said it was halal and the Arab owners confirmed that it was. This is a nice, downtown spot across the street from the park and right next to the Park Street subway station. We enjoyed the chicken kabob and baked haddock dinners, which both came with rice and salad. Very enjoyable, hot and fresh food. When I went to buy some rice pudding for dessert, the owner gave it to us for free. What a nice touch. InshaAllah my wife and I will return soon.",
"date": "2011-05-27 08:07:23",
"multicriteria_stars": {"food": 5, "ambiance": 4}

"review_id": "nw6dMN4dVYza32t4fOYJWw",
"user_id": "5kFcxNT_ngVMu2fIEtFQNA",
"business_id": "8e-xCj4xWGoYknADpovxVQ",
"stars": 4.0,
"useful": 4,
"funny": 0,
"cool": 1,
"text": "Super fan of this spot in downtown! Came out for a lunch meeting and was surprised at how good it truly was. Ordered the Lamb Gyro platter with side of Hummus and Babaghanoush. Platter has a good portion of lamb on top of a small amount of rice... served with pita and pita chips. Great Lunch spot. My new fav Mediterranean food place in Orlando!",
"date": "2013-08-03 15:32:43",
"multicriteria_stars": {"food": 4, "ambiance": 5}

"review_id": "lcTzmBYSJsTNTtaoY6Zlnw",
"user_id": "OIgKgNgdnT8uXD7EmI3IDQ",
"business_id": "j9-xoRo9U2ugRXJjWdmXmQ",
"stars": 4.0,
"useful": 4,
"funny": 1,
"cool": 3,
"text": "I've always liked Flip (I've been to both the Buckhead and Howell Mill locations several times over the past few years), but of the two Richard Blais conceptions - Flip Burger and HD1, HD1 was my favorite; so I was quite saddened when I learned they were turning this location into a Flip. This past Sunday was my first visit since Flip Burger opened, the space is fairly identical with the exception of additional seating (two large booths on either end of the restaurant). I had previously heard that management was planning on keeping a few of HD1's mainstays on the menu, but unfortunately only the 'Haute Dog' made it. No biggie, the BF and I both decided on burger options; he had the Öaxaca Burger(̈beef patty topped with sliced avocado, pico de gallo, and smoked mayo). I was a litle hesitant trying it - it sounded much too spicy for my wimpy taste buds - but the spice was surprisingly mild. Good news for me, but bad news if you're looking for some kick. The burger itself was well-cooked, very tender, and the toppings were a very refreshing contrast to the beef. I chose their special of the day: the S̈urf and TurfB̈urger (beef patty, topped with sauteed lobster knuckles, g̈arlic

herb white wine butter; pea shoots, and green goddess dressing) which came with a side of seasoned fries - served with mayo and ketchup ($16). The adjectives were just too much for me to resist, I had to have it. I wasn't disappointed it was quite tasty, it was the same beef patty of the 'Oaxaca' - the lobster was buttery and well cooked, also more plentiful than I thought there'd be. Always a plus! We finished our meal by sharing the Pistachio L̈iquid Nitrogenm̈ilkshake. No surprise, I loved it. I really like that the milkshakes have a more complex sweetness to them, there's a muti flavor component to them as well. The service was excellent, our server was friendly and attentive. There was a slight delay in our entrees coming out - but the manger came out straightaway and apologized profusely. My one tiny complaint/tip regarding Flip (all locations) - the burgers are quite small - hella tasty and gourmet - but small. Just having a burger will leave you still a bit hungry, definitely go for an app, a side - and how could you resist one of their milkshakes?",
"date": "2014-02-05 18:02:28",
"multicriteria_stars": {"food": 4, "price": 1, "service": 5, "ambiance": 4}

"review_id": "xwNlRAyJ1tLQQcN8IQmenA",
"user_id": "ywf1G9IvDLg_KytwLimFCw",
"business_id": "9JtY3GKP0QJhYzsEzFugOg",
"stars": 4.0,
"useful": 4,
"funny": 0,
"cool": 4,
"text": "Finally, a non-threatening and reasonably priced Brazilian BBQ! You have two options here: all-you-can-eat for $12 and change or pay by the pound. The unlimited option was definitely the way to go. Now, buffets can be tempting, but keep in mind that the best stuff comes right to your table ON A SWORD. We filled up our plates with samplings from the buffet, from various salads (the potato salad was my favorite), rice and beans, and a gigantic french fry. The fun begins when the skewers of meat begin to arrive, ranging from chicken wrapped in bacon to pork ribs, chicken sausage, and top sirloin. There's a definite joy in grabbing the meat as it falls off from being cut and tasting as much as possible. Be sure to pace yourself; this place is extremely laid back, there's no card to flip over to tell them to bring meat and they'll keep serving you for as long as you'd like. A special bonus was the live music on Friday night – a singer with a beautiful voice and a talented guitar player – offering up traditional songs, as well as their gorgeous arrangements of pop songs like Rock with You, Come Away With Me, and Hotel California. Highlight of the evening was absolutely Tears in Heaven. There's ample parking behind the restaurant and it's located conveniently in downtown Everett. The value is tremendous here for the amount of succulent meat and I've paid four times as much for less of an experience. I will definitely be back.",
"date": "2010-08-21 04:34:46",
"multicriteria_stars": {"food": 5, "price": 3, "service": 4, "ambiance": 5}

"review_id": "Zul1x6eu4VPaqwVOdiTuqg",
"user_id": "-zqGTvnQt9IFgTsUvuYcSw",
"business_id": "lyhNDfX8UatlRO5H3Kfccg",
"stars": 5.0,
"useful": 4,
"funny": 2,
"cool": 2,
"text": "As many people before have said I shall too say these are some of the best Persian kabobs you can get in the city! Truly juicy koobideh. The owner is so humble and says öh I just thought I would try cooking for more than my familyb̈ut I can't imagine him not having thought that because where would we downtown Portlander's be without his delicious Persian food?!?! He gives you healthy portion, accepts credit cards or cash and conveniently stays open until 4 o'clock which is later than a lot of the other food carts in the area. It's quickly become such a downtown staple. Plus the owner is so adorable and lovable that he was even featured in a local commercial, that's pretty cool.",
"date": "2015-09-23 23:14:02",
"multicriteria_stars": {"food": 4, "price": 3, "ambiance": 5}