CLUSTERING MULTILINGUAL ASPECTS FOR SENTIMENT ANALYSIS

Lucas Rafael Costella Pessutto
Danny Suarez Vargas
Viviane P. Moreira

⊙⊙⊙○○ Publicada há 5 dias

Great meats but also noisy

Tradução do Google

The ambience of the restaurant is very nice and authentic, it breaths the 'classic steak house' vibe. We went for a sirloin and filet mignon which were cooked to perfection. The salad we got was very nice too, the baked potato okayish. However, what really annoyed us was the noise - to the point that it was hard having a conversation. The repetitive jazz on the background in combination with the regular chatter didn't make us want to stay long. I'm not saying a restaurant should come with total silence but this was a disappointing side to the otherwise good experience.

carlobernardini
Amsterdã, Países Baixos
🔲 34 👍 6

- Users exchange their shopping experiences, by writing product reviews
- Reviews are usually raw texts
- Reviews are a valuable source of information (for buyers and companies), but hard to process

GuilhermeMa...
Mogi Mirim, SP
🔲 29 👍 4

Publicada 13 de junho de 2018
ESPETACULAR EXPERIÊNCIA
CARNÍVOROS! ESSE É O LUGAR!

Ambiente agradável, atendimento espetacular! Bebidas com p...

A carne é maravilhosa! Tudo foi muito bom! E não achei caro!
absurdamente bem, com bebidas e tudo ficou 200 dólares. Pa...
o preço excelente.

267lynn
**Manchester,
Reino Unido**
🔲 6 👍 4

Publicada há 3 semanas
Tradução do Google
毁了我的生日！

我去生日的时候去了家很贵的餐厅，我很失望。我们酒店推荐这家餐厅，我是通过...
我到了餐厅就点菜，服务非常迅速，实际上我们准备的时间不到一个小时。但我们...
后，主菜就等着了。我们餐厅可以吃点甜食，但是简单到了，小费被割了细，立...
到20%的小费，这是我们与服务员的唯一一次谈话。被我谈好，能够很不错，但...
了。

Pantels_Rho...
Rodes, Grécia
🔲 106 👍 25

Publicada 29 de julho de 2018
Πολύ καλό κρέας σε κλασσική Νεου...
Από τα καλύτερα steakhouse του Μανχάταν. Μεγάλες...
ατμόσφαιρα ρετρό. Μάλλον ακριβό εστιατόριο, αλλά αξί...
έχει Έλληνα ιδιοκτήτη, για αυτό στην είσοδο του εστιατ...
φωτογραφίες τσολιάδων.

FLB17082017
**Vimercate,
Itália**
🔲 112 👍 8

Tradizione
Tradução do Google
In una città che offre immense opportunità d...
più raffinato rispetto al passato, offre una esp...
dimensioni, patatine ottime, ambiente anni 60...
una placevolissima serata

Chema L
Chema L
🔲 20 👍 4

Publicada há 1 semana · 🔲 via dispositivo móvel
Muy buena carne
Tradução do Google
Entrada espectacular, con la nevera acristalada con la carne a la vista. Ya marca la
especialidad del restaurante. Personal muy atento y correcto, quizá con la excepción del q
atiende a la entrada, q resultó correcto pero frío. La carne espectacular. En mi caso T-bone
poco hecho. Raciones y acompañamientos generosos. Evidentemente, no es barato, pero se puede
esperar que sea barato, aunque no es de los steak house más caros.

Gurmangezgin
**Istambul,
Turquia**
🔲 245 👍 45

Publicada 22 de outubro de 2017 · 🔲 via dispositivo móvel
Manhattanın en iyilerinden
Tradução do Google
Manhatta'da times square e çok yakın. Klasik olmuş bir steakhouse. Gi...
pişmemiş etler karşılıyor. Oldukça popüler bir mekan. Etler çok güzel. Ta...
Mostrar menos

jeandenice06...
Nice, França
🔲 187 👍 65

Publicada há 1 semana
Excellent Steakhouse
Tradução do Google
Un grand moment de dégustation d'une excelle...
Certes cher, mais cela en vaut vraiment la pein...
Le service était impeccable avec le but de la sa...
Viande succulente.
Je connaissais déjà, et y retournerai !

Cepreй C
Cepreй C
🔲 40 👍 22

Publicada 25 de dezembro de 2017 · 🔲 via dispositivo móvel
Красиво но и только
Tradução do Google
Типичный американский стейк-хаус с уютным интерьє...
больше разочаровали чем порадовали - оказ ...
примечательного... Официант потом еще насто ...
недостаточно. Удовольствия по итогу не получили...

RolfundClaudia
Baar, Suíça
🔲 67 👍 25

Publicada há 2 semanas
Spezielle Ambiance, sehr gutes Essen und sehr guter Serv...
Tradução do Google
Wir besuchten nach Jahren wieder einmal NYC und das Gallaghor's. Man hat uns e...
wunderbaren Tisch reserviert, die Bedienung war ausgezeichnet und das Essen sel...

5

carlobernardini
Amsterdã,
Países Baixos
↗ 34 👍 6

◖◖◖◖◖ Publicada há 5 dias

Great meats but also noisy

| Tradução do Google |

The ambience of the restaurant is very nice and authentic, it breaths the 'classic steak house' vibe. We went for a sirloin and filet mignon which were cooked to perfection. The salad we got was very nice too, the baked potato okayish. However, what really annoyed us was the noise - to the point that it was hard having a conversation. The repetitive jazz on the background in combination with the regular chatter didn't make us want to stay long. I'm not saying a restaurant should come with total silence but this was a disappointing side to the otherwise good experience.

6

- Aspect-Based Sentiment Analysis tends to extract a large number of aspects from reviews
- People tend to use different words to express same aspect
- For example: *beef*, *filet mignon*, and *meat* represent the same aspect in a restaurant domain
- Thus, it becomes necessary to refine the results of ABSA algorithms

- Task of grouping together equivalent aspects across multiple languages
- First work to address this problem
- **Hypothesis**: The extraction and representation of contextual information of reviews combined with a good similarity measure allow us to group multilingual aspects without translation

Restaurant Reviews

- Service was just ok, it is not what you'd expect for $500.
- Single Worst Restaurant in Manhattan.
- The crust has a great bite and a good chew, the sauce has good acidity to it, the salt from the cheese is great.
- The meat is fresh, the sauce is great and service is good too.

- El servicio es muy bueno y la calidad de la comida al mismo nivel.
- Fabuloso, muy atentos a la comida excelente y un ambiente estupendo.
- Sólo puedo decir cosas buenas de este restaurante.
- Agradable ambiente, muy buen servicio y exquisitos platos.

- Een sfeervolle atmosfeer met een live band die de juiste toon vond.
- Het personeel is zeer vriendelijk en correct zonder stijf te zijn.
- Gezellige sfeer, vriendelijke bediening en het eten is TOP!
- De chef-kok is een kunstenaar die weet hoe om te gaan met de smaak en het uiterlijk van goed eten.

Aspect Phrase sets with respective Virtual Document considering a window of [-3,3] words

Service = {ok, great, good}
Restaurant = {single, worst, Manhattan}
Crust = {great}
Sauce = {good, chew, good, acidity, fresh, great}
Cheese = {salt, great}
Meat = {fresh}

Servicio = {muy, bueno, buen, exquisitos}
Comida = {calidad, mismo, nivel, atentos, excelente}
Ambiente = {excelente, estupendo, agradable, buen}
Restaurante = {buenas}
Platos = {exquisitos}

Omgeving = {sfeervolle, livebandje}
Personeel = {zeer, vriendelijk}
Sfeer = {gezellige, vriendelijke}
Bediening = {vriendelijke}
Eten = {top, uiterlijk, goed}
Chef-kok = {kunstenaar}

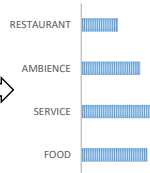### Results of Multilingual Aspect Clustering

FOOD
Crust = {great}
Sauce = {good, chew, good, acidity, fresh, great}
Cheese = {salt, great}
Meat = {fresh}
Comida = {calidad, mismo, nivel, atentos, excelente}
Platos = {exquisitos}
Etem = {top, uiterlijk, goed}

SERVICE
Service = {ok, great, good}
Servicio = {muy, bueno, buen, exquisitos}
Personeel = {zeer, vriendelijk}
Bediening = {vriendelijke}
Chef-kok = {kunstenaar}

AMBIENCE
Ambiente = {excelente, estupendo, agradable, buen}
Omgeving = {sfeervolle, livebandje}
Sfeer = {gezellige, vriendelijke}

RESTAURANT
Restaurant = {single, worst, Manhattan}
Restaurante = {buenas}

Possible use of results

RESTAURANT

AMBIENCE

SERVICE

FOOD

2 PROBLEM DESCRIPTION / SOLUTION OVERVIEW

MULTILINGUAL ASPECT CLUSTERING
OVERVIEW

**CLUSTERING MULTILINGUAL
ASPECTS FOR SENTIMENT
ANALYSIS**

- Aspect Phrase – Explicit properties of a target that occur in a set of reviews. Can be composed of one or more terms, *e.g.* lava cake, wine list

- Aspect Group – Set of aspect phrases that represent a same feature
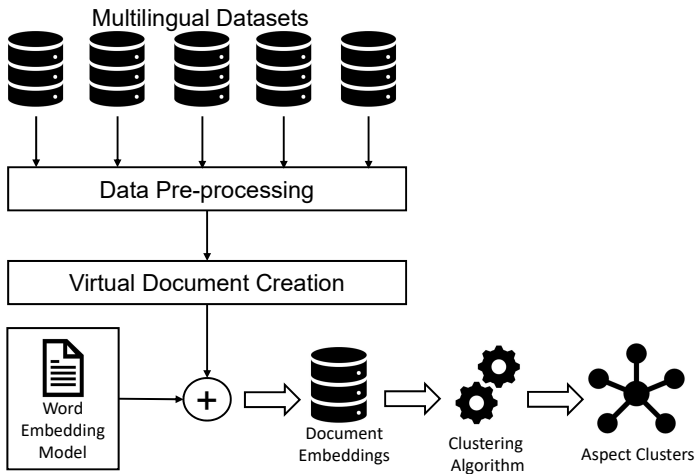
- Splitting of the review text into sentences
- Tokenization
- Converting all words to lowercase

- We followed the proposal by [Zhai et al. 2011]
- Extraction of the surrounding words of each occurrence of an aspect phrase, in a $[-t, t]$ window
- Stopwords and other aspect phrases that co-occur in the same sentence are removed

Example:
"The <u>service</u> is amazing and the <u>ambience</u> is good for a date"
Given $t = 5$,
$VD_{service} = \{amazing\}$
$VD_{ambience} = \{amazing, good, date\}$

- Representation of Virtual Documents in a language-independent form
- To achieve this representation, we employ Multilingual Word Embeddings
  - Transform monolingual word embedding spaces in a compatible multilingual space
- **Document Embeddings**
  - Composed of the word embedding representations of each word in the Virtual Document and each word in the aspect phrase

- Centroid-based clustering algorithm. (k-means)
- Input: Document Embeddings and the desired number of clusters k
- Output: Set of k clusters of related Aspect Phrases
- How k-means works:
  1. Selection of initial centroids
     - Randomly
     - Using a Heuristic
  2. Set the remaining aspect phrases to a cluster
  3. Update the centroids
  4. Repeat steps 2 and 3 until no aspect phrase changes cluster

- The distance measure used to compare two Document Embeddings is the Word Mover's Distance (WMD) [Kusner et al. 2015].

- It is a dissimilarity measure. The higher the WMV score, the more different the two document embeddings are

- WMV measures the minimum traveling cost of the words in one document to words in the other document

- It works even if the two documents have no words in common

The wine list has interesting good values

They have a good beverage menu with reasonable prices

- There are no ready datasets for this specific task
- Multilingual datasets are difficult to annotate
- The datasets used were provided in Task 5 from SemEval 2016
  - Restaurant Domain
  - Data in 5 languages: English, Dutch, Russian, Spanish, and Turkish
  - Annotated data with information of aspect phrases and their classification in six categories: Food, Restaurant, Drinks, Service, Ambience, and Location

Dataset	English	Dutch	Russian	Spanish	Turkish	Total
#Reviews	350	300	312	627	300	1,889
#Senten.	2,000	1,722	3,655	2,070	1,232	10,679
#APs	644	508	1,024	543	831	3,550

- Adaptation of [Zhai et al. 2011] monolingual technique to group aspect phrases
- This work was chosen because it is the most seminal work in this field
- Semi-supervised technique, that automatic labels part of the data before clustering
- Labelling is based on two heuristics: group aspect phrases that share some word and lexical similarity
- Clustering through EM + Naïve Bayes algorithm

- Entropy is the proportion of each gold partition present in each cluster

- Purity measures the largest portion of a cluster that contains data from a single golden partition

- The goal is to simultaneously minimize entropy and maximize purity

• ENTROPY

Method	English	Dutch	Russian	Spanish	Turkish	All
L-EM	1.748	1.753	1.974	1.591	2.076	1.932
MAC-RAND	1.654	1.801	1.814	**1.517**	**2.002**	1.858
MAC-CENT	**1.624**	**1.719**	**1.706**	1.540	2.039	**1.841**

• PURITY

Method	English	Dutch	Russian	Spanish	Turkish	All
L-EM	0.598	**0.591**	0.503	**0.644**	0.484	**0.549**
MAC-RAND	0.605	0.559	0.540	**0.644**	0.487	0.543
MAC-CENT	**0.629**	0.576	**0.571**	0.581	**0.492**	0.539

- Aspect Phrases that share Words

Aspect Phrase	Translation (EN)	Lang.
menu 'parels van india'	menu 'pearls of india'	DE
menu kaart	menu card	DE
3 gangen menu	3 course menu	DE
детское меню	children's menu	RU
блюд из меню	dishes from the menu	RU
sake menu	sake menu	EN
menu de primavera	Spring Menu	ES
menu fiyatları	menu prices	TR

- Semantically Related Aspect Phrases (sea food)

Aspect Phrase	Translation (EN)	Lang.
scallops	scallops	EN
scampi in de look	scampi in de look	EN
рыба в беконе	fish in bacon	RU
sea urchin	sea urchin	EN
fried shrimp	fried shrimp	EN
lobster knuckles	lobster knuckles	EN
oysters	oysters	EN
stir fry blue crab	stir fry blue crab	EN
fried oysters and clams	fried oysters and clams	EN
pulpo con langostinos	octopus with prawns	ES
vieira con sopa	scallop with soup	ES
soya soslu somon	salmon with soy sauce	TR

- Implement another version of our approach - in which the number of clusters does not need to be informed
- Development of heuristics that can be used with multilingual aspect phrases, in order to build a semi-supervised algorithm
- Build a visualization tool in order to summarize the results of clustering

📄 KUSNER, M. et al. From word embeddings to document distances. In: *International Conference on Machine Learning*. [S.l.: s.n.], 2015. p. 957–966.

📄 ZHAI, Z. et al. Clustering product features for opinion mining. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. [S.l.]: ACM, 2011. (WSDM '11), p. 347–354. ISBN 978-1-4503-0493-1.

Lucas Rafael Costella Pessutto
Danny Suarez Vargas
Viviane P. Moreira
Institute of Informatics — UFRGS
`lrcpessutto@inf.ufrgs.br`