

---

# Extracció de sentiments i opinions

---

PID\_00257770

Joaquim Moré

---

Temps mínim de dedicació recomanat: 4 hores

---



Universitat  
Oberta  
de Catalunya

**Joaquim Moré**

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats pel professor: Jordi Casas (2019)

Primera edició: setembre 2019  
Autoria: Joaquim Moré  
Llicència CC BY-SA d'aquesta edició, FUOC, 2019  
Av. Tibidabo, 39-43, 08035 Barcelona  
Realització editorial: FUOC



*Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-Compartir igual (BY-SA) v.3.0 Espanya de Creative Commons. Podeu modificar l'obra, reproduir-la, distribuir-la o comunicar-la públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), i sempre que l'obra derivada quedi subjecta a la mateixa llicència que el material original. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca>*

# Índex

<b>Introducció</b> .....	5
<b>1. Elements d'una opinió</b> .....	7
1.1. El <i>target</i> o <i>topic</i> .....	7
1.2. Els aspectes .....	7
1.3. La polaritat .....	7
1.4. L' <i>opinion holder</i> i el moment .....	7
1.5. Exemple dels elements d'una opinió .....	8
<b>2. Detecció automàtica dels elements d'una opinió</b> .....	9
2.1. Detecció del <i>target</i> .....	9
2.1.1. Detecció de les diferents denominacions del <i>target</i> amb una ontologia .....	10
2.1.2. Detecció de les diferents denominacions del <i>target</i> amb <i>word embeddings</i> .....	10
2.1.3. Detecció de la referència anafòrica al <i>target</i> .....	11
2.2. Detecció dels aspectes .....	12
2.2.1. Detecció a partir de les <i>infoboxes</i> .....	13
2.2.2. Detecció a partir d' <i>opinion words</i> .....	14
2.3. Detecció de la polaritat .....	17
2.3.1. Dificultats en la detecció de la polaritat amb criteris preestablerts .....	17
2.3.2. Detecció de la polaritat amb <i>machine learning</i> .....	18
2.4. Detecció de l' <i>opinion holder</i> i del moment .....	20
<b>3. Casos d'ús de l'anàlisi de sentiments</b> .....	22
3.1. Coses a millorar a través de les opinions .....	22
3.1.1. La metodologia .....	22
3.1.2. Recopilació del corpus d'anàlisi .....	23
3.1.3. Creació de les col·leccions .....	23
3.1.4. Identificació del <i>target</i> .....	23
3.1.5. Identificació dels aspectes del <i>target</i> .....	24
3.1.6. Aspectes vinculats a <i>opinion words</i> de polaritat negativa .....	27
3.1.7. Informe .....	28
3.2. Opinions sobre un programa de televisió .....	29
3.2.1. Obtenció de les piulades .....	30
3.2.2. Exploració .....	30
3.2.3. Preprocessat .....	33
3.2.4. Anàlisi de les dades .....	34
3.2.5. Informe .....	40

3.3.	Detecció d'opinions falses .....	40
3.3.1.	Tipificació del problema .....	41
3.3.2.	Com cal abordar el problema? .....	41
3.3.3.	Anàlisi textual .....	43
3.3.4.	Classificador d'opinions falses amb <i>machine learning</i> .....	47
<b>Resum</b> .....		<b>48</b>

## Introducció

En aquest mòdul ens ocuparem de com extreure els sentiments i les opinions a partir dels textos.

L'anàlisi dels sentiments i les opinions s'aplica a gairebé tots els negocis i activitats socials. Les opinions dels altres influeixen en el nostre comportament i la nostra percepció del món. Potser aquesta afirmació sembla una mica exagerada, però, quan reservem un hotel, no mirem abans les seves avaluacions? No llegim les crítiques d'una pel·lícula abans de veure-la? No ens agrada escoltar tertúlies polítiques per a sentir-nos reafirmats en les nostres conviccions quan els tertulians les defensen o les ataquen?

Les productores cinematogràfiques, els partits polítics, les empreses de serveis, etc., inverteixen molts diners a treure partit d'aquesta màxima: les nostres creences i percepcions de la realitat i, sobretot, les nostres decisions estan condicionades per com els altres veuen el món. I ja sabem el profit econòmic que es pot treure quan un sap com influir en les decisions. Però no solament això, si un Govern i les elits econòmiques són capaces de crear un estat d'ànim emocional positiu, els beneficis econòmics i socials són evidents. Valgui com a exemple un estudi de la Universitat d'Indiana el 2010 que demostrava que l'estat d'ànim i els sentiments de la població expressats a Twitter eren predictors de les pujades i baixades de la borsa en un dia concret.

L'anàlisi de sentiments no és una cosa nova. Sempre s'ha fet, però abans de l'era Web 2.0 es feia amb enquestes i reunions de grup. Les coses no eren com ara, que tothom té ganes d'escriure per al públic les seves opinions de manera desinteressada en els blogs, a les pàgines web d'hotels o esdeveniments, a Facebook, a Twitter, a les notícies dels diaris digitals, etc. Com que les enquestes i les reunions de grup són cares de fer, s'ha recorregut al processament del llenguatge natural per a analitzar, amb una despesa molt baixa, un ingent nombre d'opinions publicades cada dia.

Ara bé, el llenguatge natural té unes característiques que fan que la seva anàlisi automàtica sigui complexa. D'algunes d'aquestes característiques ja n'hem parlat en el mòdul anterior, l'ambigüitat, per exemple. No és el mateix dir *una brutal agressió* que dir *aquesta sèrie és brutal* o *aquest plat t'ha quedat brutal*. En la primera opinió, es destaca una cosa negativa, en la segona opinió, l'adjectiu *brutal* pot interpretar-se en sentit positiu i negatiu i en la tercera opinió l'adjectiu només s'interpreta en sentit positiu.

En aquest mòdul explicarem els elements d'una opinió i com aquests es poden detectar i analitzar amb processament del llenguatge natural. Després aplicarem l'anàlisi de sentiments per a abordar una de les qüestions més importants

### Bibliografia recomanada

Johan Bollen; Huina Mao; Xiao-Jung Zeng (2011). «Twitter mood predicts the stock market». *Journal of Computational Science* (vol. 2, núm. 1, pàg. 1-8).

a les quals s'ha d'enfrontar qualsevol negoci, persona o col·lectiu que depèn molt de la reputació en línia. Ens referim a la intervenció d'individus la missió dels quals consisteix a influir en la reputació d'un col·lectiu o d'un negoci.

Per a explicar i exemplificar els conceptes teòrics s'utilitzaran exemples d'opinions trobats en la bibliografia sobre anàlisi de sentiments. Una vegada coneguts els principals elements teòrics, tractarem els següents hipotètics casos d'ús.

1) Els propietaris d'un hotel, els hostes del qual són majoritàriament russos i espanyols, volen saber quins serveis i característiques de l'hotel han de cuidar-se especialment segons la nacionalitat de l'hoste.

2) Una cadena de televisió vol analitzar les opinions sobre un programa dedicat al col·lectiu que es dedica a la prostitució voluntàriament.

3) Un cercador d'hotels, conscient del dany econòmic que pot suposar per al seu negoci la publicació d'opinions falses que perjudiquen la reputació de les cadenes hoteleres, vol aplicar un detector d'opinions falses abans de publicar-les.

Imaginem que aquests projectes són encàrrecs a l'empresa Sense and Suitability (S&S), amb els mateixos protagonistes del mòdul anterior. Anne i Henry treballen en el projecte del programa de televisió, i Joseph i Beth treballen en els altres dos projectes.

## 1. Elements d'una opinió

L'objecte d'estudi d'aquest mòdul són les opinions que expressen sentiments negatius o positius. El primer repte és reconèixer automàticament una opinió. Per a això, cal fer una descripció estructurada del problema.

Les opinions tenen dos components fonamentals:

- 1) l'entitat sobre la qual s'expressa una opinió
- 2) el sentiment expressat en aquesta opinió

L'opinió també pot centrar-se a expressar un sentiment sobre un o més aspectes de l'entitat.

### 1.1. El *target* o *topic*

L'entitat sobre la qual s'expressa una opinió es coneix amb la denominació anglesa *target* o *topic*. Nosaltres utilitzarem la denominació *target*.

Si l'opinió se centra en un sol *target*, analitzem opinions regulars, però si la centrem en més d'un *target*, direm que analitzem opinions comparatives. En aquest bloc ens centrem en les opinions regulars, les quals denominarem simplement opinions.

### 1.2. Els aspectes

Els aspectes són la part o propietats del *target* sobre la quals s'emet una opinió. Per exemple, hi ha opinions que més que opinar sobre un model de telèfon mòbil en el seu conjunt se centren en la pantalla, el teclat o les seves funcions, com la fotografia, l'assistent, etc.

### 1.3. La polaritat

Els sentiments expressats en l'opinió es coneixen com a orientacions del sentiment o de l'opinió. També es coneixen com a polaritat, que és el terme que utilitzarem aquí. Direm que una opinió té polaritat positiva si el sentiment que transmet és positiu, i polaritat negativa si el sentiment transmès és negatiu.

### 1.4. L'*opinion holder* i el moment

El *target* i la polaritat són elements fonamentals, però hi ha altres components que també ho són. Ens referim a qui expressa l'opinió, conegut amb la denominació anglesa *opinion holder*, i al moment en el qual s'expressa.

#### Bibliografia

B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». A: C. Aggarwal; C. Zhai (ed.). *Mining Text Data*. Boston, DT.: Springer.

B. Liu (2010). «Sentiment Analysis and Subjectivity» (cap. 26). A: N. Indurkha; F. J. Damerau (eds.). *Handbook of Natural Language Processing* (2a. ed.). CRC Press.

## 1.5. Exemple dels elements d'una opinió

Vegem un exemple d'opinió on es poden trobar els elements comentats:

Figura 1. Exemple d'opinió

I bought an **iPhone** a few days ago. **It** was such a **nice phone**. The **touch screen** was **really cool**. The **voice quality** was **clear** too. Although the **battery life** was not **long**, that is **ok** for **me**. However, **my mother** was mad with me as I did not tell her before I bought **it**. **She** also thought the **phone** was **too expensive**, and wanted me to return **it** to the shop. . .

Font: B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». A: C. Aggarwal; C. Zhai (ed.). *Mining Text Data*. Boston, MA.: Springer.

### Bibliografia

Aquest exemple s'ha extret de B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». A: C. Aggarwal; C. Zhai (eds). *Mining Text Data*. Boston, MA.: Springer.

En vermell hem marcat el *target*. Cal tenir en compte que la seva denominació no és sempre la mateixa. El *target* es pot expressar amb la seva marca (*iPhone*), pel tipus d'objecte que és (*phone*) o amb un pronom.

En violeta estan marcats els aspectes de l'entitat. La pantalla tàctil, la bateria i per extensió la seva durada són aspectes de l'entitat, que és l'*iPhone*. La qualitat de so de la veu també és un aspecte de l'*smartphone* sobre el qual s'opina. Captar la relació que no es diu de manera explícita entre l'entitat i els seus aspectes, i detectar quins d'aquests aspectes s'estan avaluant, també és un objectiu molt complex del processament del llenguatge i de la intel·ligència artificial.

En marró hem marcat els elements que indiquen la polaritat. El que és difícil és seleccionar, entre els elements que fan referència a polaritat, aquells que es refereixen al *target*. És a dir, triar *nice*, *really cool*, *ok*, *long* o *too expensive*, però no triar *mad* perquè es refereix a la mare, que no és el *target*.

Finalment, en verd estan marcats els dos *opinion holders*: l'escriptor de l'opinió i la seva mare, que opina que el telèfon és massa car.



## 2. Detecció automàtica dels elements d'una opinió

En aquest apartat presentarem els mètodes que permeten detectar els elements de les opinions. Començarem pel *target* i els seus aspectes. Continuarem amb la polaritat i finalment ens ocuparem de l'*opinion holder* i el temps.

### 2.1. Detecció del *target*

Mirem l'exemple d'opinió que hem presentat abans.

Figura 2. Exemple d'opinió

I bought an **iPhone** a few days ago. **It** was such a nice **phone**. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought **it**. She also thought the **phone** was too expensive, and wanted me to return **it** to the shop...

Font: B. Liu; L. Zhang (2012). «A Survey of Opinion Mining and Sentiment Analysis». A: C. Aggarwal; C. Zhai (ed.). *Mining Text Data*. Boston, DT.: Springer.

En vermell estan marcades les referències al *target* de l'opinió, és a dir, la cosa sobre la qual s'opina. Aparentment, les referències són evidents, però des del punt de vista del processament del llenguatge natural, trobar aquestes referències i resoldre que totes apunten a la mateixa entitat no és tan senzill ni de bon tros. Les referències al mateix objecte amb denominacions diferents és un camp d'estudi molt actiu en el processament del llenguatge natural.

En principi, l'opinió se centra entorn d'un producte. Aquest producte té un nom concret, però fixem-nos que també es pot fer referència al producte utilitzant un hiperònim –això és, *phone*, que és la classe o tipus al qual pertany l'entitat. D'altra banda, observem que també es fa referència al mateix objecte amb dues paraules sinònimes (*phone* i *telephone*).

Si ja és complicat reconèixer que totes aquestes denominacions diferents es refereixen a la mateixa cosa, cal afegir el reconeixement que el pronom *it* a «*I bought it*» fa referència al telèfon i no a la pantalla. Resoldre correctament a quina entitat fa referència un pronom es denomina resolució de l'anàfora.

D'altra banda, *phone* és també la denominació del verb *trucar*, així com una unitat de so en el camp de la fonètica. Resoldre que el sentit de *phone* és el mateix que el de *telephone* és feina del desambiguador del sentit d'una paraula, també conegut com a WSD (*Word Sense Disambiguation*).

En aplicacions de l'anàlisi de sentiments el *target* és un nou model de producte, la marca d'una empresa, un polític, un partit polític, un col·lectiu, etc. És a dir, en molts projectes el *target* ja està definit pel client. Per tant, en aquests casos,

el repte consisteix a reconèixer les diferents denominacions d'aquest *target* en les opinions. Vegem, per exemple, com trobem les denominacions del *target* en l'opinió d'exemple.

### **2.1.1. Detecció de les diferents denominacions del *target* amb una ontologia**

Una estratègia possible és utilitzar una ontologia com la de Wordnet. L'estratègia consisteix primer a establir el sentit o *synset* del *target*, i després identificar les denominacions que són o bé sinònimes o bé hiperònimes o hipònimes. Per exemple, tenint com a *target iPhone*, una referència a aquest *target* podria ser el seu hiperònim, com *phone* en l'opinió d'exemple.

Ara bé, noms de productes i marques que s'han popularitzat de tal manera que més que una marca són un concepte, com *iPhone*, no són a Wordnet. No obstant això, com ja expliquem en el mòdul anterior, es pot consultar l'ontologia YAGO a través de la base de dades de la DBpedia i veure quin *synset* de Wordnet és el seu hiperònim. En el *notebook* PLA-2 1.1 es comprova que *telephone* és hiperònim d'*iPhone*. D'aquesta manera, en l'opinió anterior podem establir que els termes *iPhone* i *phone* es refereixen al *target*, sempre que s'hagi comprovat anteriorment que *phone* i *telephone* són en el mateix *synset*.

Cal tenir en compte que DBpedia s'organitza com un graf de coneixement, conegut com a *Knowledge Graph* o KG, i que el formalisme de representació de les relacions existents en aquest graf és el RDF, emprat per al web semàntic. Les relacions consisteixen en un conjunt de tripletes RDF. Cada triplet RDF té els següents elements: un subjecte, un predicat i també una llista d'objectes. En el *notebook* PLA-2 1.1 es pot veure que, en la nostra consulta, el subjecte és la URL que apunta a la pàgina de la DBpedia que descriu *iPhone*, i el predicat és *type*, això és, el tipus (hiperònim) al qual pertany l'entitat. En la resposta a la consulta l'objecte és una llista amb les URL que apunten a les relacions amb el seu hiperònim segons diferents ontologies. Triem les URL corresponents a l'ontologia YAGO.

### **2.1.2. Detecció de les diferents denominacions del *target* amb *word embeddings***

En principi, les diferents denominacions a la mateixa entitat apareixen en contextos molt similars. Per tant, amb un càlcul dels termes més propers al *target* amb el mètode Word2Vec, es poden detectar les referències sinònimes al *target*. El *target* sol ser un sintagma nominal.

En el *notebook* PLA-2, 1.1 es mostra com s'obtenen diferents denominacions del *target Samsung*. Després d'haver creat un model d'opinions sobre telèfons mòbils a Amazon amb Word2Vec, obtenim els termes més semblats al *target* que són sintagmes nominals.

És curiós, però, veure que amb el mètode Word2Vec també s'obtenen termes relacionats amb el *target* per una relació de metonímia, això és, són denominacions de la part pel tot. Com es veu a continuació, la marca *Samsung* s'associa als seus productes. Per tant, l'opinió de la marca es projecta a través de l'opinió dels seus productes (p. ex.: *samsung products, galaxy, touchwiz, touch wiz*). Ara bé, hem d'anar amb compte perquè aquest mètode també revela una tendència en les opinions: referenciar el *target* contrastant els seus productes amb els de la competència (*lg, htc*).

Figura 3. Termes similars a *Samsung* segons un model d'opinions amb Word2Vec

```
['samsung products', 'htc', 'lg', 'galaxy', 'touch wiz', 'touchwiz', 'samsungs']
```

### 2.1.3. Detecció de la referència anafòrica al *target*

Una referència anafòrica és la relació d'identitat que s'estableix entre un sintagma nominal amb un nom o un pronom i una entitat que ha aparegut anteriorment en el discurs, que es diu antecedent. En l'exemple que hem presentat, la relació anafòrica és la relació que s'estableix entre el pronom *it* i *iPhone* a *I bought it* i a *return it*.

La identificació de l'antecedent d'una referència anafòrica és tasca dels analitzadors o *parsers* sintàctics i semàntics de l'oració. Una vegada que es construeix l'arbre sintàctic, es tria com a antecedent de l'expressió anafòrica aquell sintagma nominal que ha aparegut anteriorment en el discurs que millor s'adequa a les restriccions semàntiques a les quals està sotmesa l'expressió anafòrica. Per exemple, per a trobar l'antecedent d'*it* a *I bought it*, el *parser* ha de trobar un antecedent que tingui trets semàntics de les coses *comprables*. Per exemple, que sigui un artefacte o, si és un ésser animat, que no sigui una persona, etc. És ben cert que hi ha *parsers* que resolen l'antecedent sense etiquetatge semàntic aplicant les restriccions purament sintàctiques entre l'expressió anafòrica i l'antecedent establertes per teories gramaticals com el model de *Government and Binding* de Noam Chomsky.

#### Bibliografia recomanada

Per a més informació sobre l'obtenció de denominacions del *target Samsung*, consulteu el *notebook* PLA-2, 1.1.

#### Bibliografia

N. Chomsky (1981). *Lectures on Government and Binding: The Pisa Lectures*. Mouton de Gruyter.

També hi ha *parsers* que resolen l'antecedent una vegada que aprenen aquesta tasca amb mètodes de *machine learning*. És el cas de *neural-coref*, una extensió de spaCy que aprèn la resolució de referències anafòriques amb una xarxa neuronal. Una vegada apresada la tasca és capaç de, donada una frase *F* que conté una expressió anafòrica, i donat un context *C* que conté l'antecedent, generar una versió de *F* en la qual l'expressió anafòrica és substituïda per l'antecedent. A la figura 4 i en el *notebook* PLA-2 1.1 presentem el codi en Python que genera una versió correcta de l'antecedent d'It a *I bought an iPhone a few days ago. It was such a nice phone*. També genera una versió correcta de l'antecedent del pronom *her* a *However, my mother was mad with m/as I did not tell her before I bought it*. No obstant això, a *bought it*, el pronom queda per resoldre.

La resolució anafòrica encara és lluny de ser un tema resolt i les eines que l'apliquen encara no són gaire fiables. Com ja hem comentat abans, la resolució anafòrica automàtica és una de les tasques més complexes del processament del llenguatge natural, per la complexitat cognitiva i les limitacions que encara existeixen en la intel·ligència artificial per a modelar les interpretacions del que alguns anomenen el «sentit comú». Un sentit comú que és el resultat de les experiències individuals, el coneixement del món, de l'entorn, etc.

Figura 4. Codi de *neural-coref* per a generar la versió d'una opinió amb les expressions anafòriques resoltes.

```
import neuralcoref
import spacy
import torch

nlp = spacy.load('en_coref_sm')
# alternative: import en_coref_sm nlp = en_coref_sm.load()
# bigger corpus: en_coref_lg

doc = nlp(u'I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. However, my mother was mad with me as I did not tell her before I bought it. Although the battery life was not long, that is ok for me')

doc._coref_resolved

'I bought an iPhone a few days ago. an iPhone was such a nice phone. The touch screen was really cool. The voice quality was clear too. However, my mother was mad with me as I did not tell my mother before I bought The touch screen . Although the battery life was not long, that is ok for me'
```

## 2.2. Detecció dels aspectes

Ens ocupem seguidament dels aspectes del *target*. Això és, les propietats i característiques de l'objecte de l'opinió.

### Bibliografia recomanada

K. Clark; C. D. Manning (2016). «Deep Reinforcement Learning for Mention-Ranking Coreference Models», EMNLP.  
<https://github.com/huggingface/neuralcoref>  
<https://bit.ly/2xt4zdb>

### 2.2.1. Detecció a partir de les *infoboxes*

Els aspectes d'un producte o d'una entitat es poden trobar en *les infoboxes* de Wikipedia. Per exemple, a la *infobox* de l'entrada *iPhone* que veiem a la figura 5, apareix una relació dels aspectes d'aquest producte (desenvolupador, fabricant, càmera, teclat, bateria, etc.).

Figura 5. Entrada d'iPhone a la Wikipedia anglesa amb la seva *infobox* en el marge dret


**iPhone** (/ˈaʊfoʊn/ *EYE-fohn*) is a line of **smartphones** designed and marketed by **Apple Inc.** They run Apple's **iOS** mobile operating system. The **first-generation iPhone** was released on June 29, 2007, and there have been multiple new hardware iterations with new iOS releases since.

The **user interface** is built around the device's **multi-touch** screen, including a **virtual keyboard**. The iPhone has **Wi-Fi** and can connect to **cellular networks**. An iPhone can **shoot video** (though this was not a standard feature until the **iPhone 3GS**), **take photos**, **play music**, send and receive **email**, **browse the web**, send and receive **text messages**, follow **GPS navigation**, record notes, perform mathematical calculations, and receive **visual voicemail**. Other functionality, such as video games, reference works, and social networking, can be enabled by downloading **mobile apps**. As of January 2017, Apple's **App Store** contained more than 2.2 million applications available for the iPhone.

Apple has released eleven **generations** of iPhone models, each accompanied by one of the eleven major releases of the **iOS** operating system. The original **first-generation iPhone** was a **GSM** phone and established design precedents, such as a button placement that has persisted throughout all releases and a screen size maintained for the next four iterations. The **iPhone 3G** added **3G** network support, and was followed by the **3GS** with improved hardware, the **4** with a metal chassis, higher display resolution and front-facing camera, and the **4S** with improved hardware and the voice assistant **Siri**. The **iPhone 5** featured a taller, 4-inch display and Apple's newly introduced **Lightning connector**. In 2013, Apple released the **5S** with improved hardware and a **fingerprint reader**, and the lower-cost **5C**, a version of the 5 with colored plastic casings instead of metal. They were followed by the larger **iPhone 6**, with models featuring 4.7 and 5.5-inch displays. The **iPhone 6S** was introduced the following year, which featured hardware upgrades and support for **pressure-sensitive touch inputs**, as well as the **SE**—which featured hardware from the 6S but the smaller form factor of the 5S. In 2016, Apple unveiled the **iPhone 7 and 7 Plus**, which add water resistance, improved system and graphics performance, a new rear dual-camera setup on the Plus model, and new color options, while removing the 3.5 mm headphone jack found on previous models. The **iPhone 8 and 8 Plus** were released in 2017, adding a glass back and an improved screen and camera. The **iPhone X** was released alongside the 8 and 8 Plus, with its highlights being a near bezel-less design, an improved camera and a new facial recognition system, named **Face ID**, but having no home button, and therefore, no **Touch ID**.

The original iPhone was described as "revolutionary" and a "game-changer" for the mobile phone industry. Newer iterations have also garnered praise, and the iPhone's success has been credited with helping to make Apple one of the world's most valuable publicly traded companies.

**iPhone**



The front face of iPhone X

<b>Developer</b>	Apple Inc.
<b>Manufacturer</b>	Foxconn, Pegatron (contract manufacturers)
<b>Type</b>	Smartphone
<b>Release date</b>	June 29, 2007 <span>[show]</span>
<b>Units sold</b>	1.2 billion <sup>[1]</sup>

Font: Wikipedia

Per tant, amb la consulta a la *Knowledge Graph* de la DBpedia es poden obtenir les propietats de l'entitat. En *el notebook* PLA-2 1.2 s'explica com. Les propietats que hem obtingut són les següents:

Figura 6. Propietats d'iPhone a la seva *infobox*

```
[ 'unitsold', 'developer', 'reason', 'sound', 'date', 'dimensions', 'power', 'caption', 'display', 'related', 'storage', 'memory', 'soc', 'graphics', 'releasedate', 'connectivity', 'service' ]
```

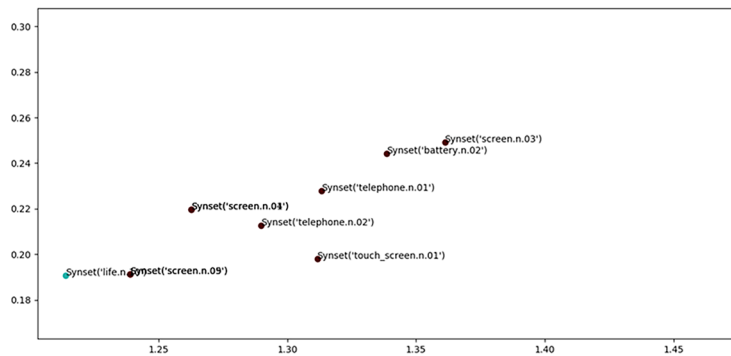
Si volem identificar els aspectes del *target* de l'opinió que ens serveix d'exemple, això és, *iPhone*, podem agrupar els sentits dels termes que són més propers al seu hiperònim, que és *telephone*.

En *el notebook* PLA-2 1.2 podem veure com es fa aquesta agrupació. Es construeix una matriu a les files de la qual hi ha els possibles sentits segons Wordnet dels termes de l'opinió. Per exemple, a les files hi ha els sentits que pot tenir la paraula polisèmica *battery*. Un d'aquests sentits és *battery.n.02*, que es defineix com a *a device that produeixes electricity; may have several primary or secondary cells arranged in parallel or series*, i també *battery.n.01*, que es defineix com a *group of guns or missile launchers operated together at one place*.

Les columnes de la matriu són els sentits que poden tenir els termes de les propietats a l'*infobox* d'iPhone. Paraules que també són polisèmiques, com *power* (p. ex.: *the rate of doing work; measured in watts* y *(of a government or*

*government official*) *holding an office means being in power*), entre d'altres. Cada cel·la conté la distància semàntica entre el sentit d'un terme i el sentit d'una propietat. Amb un mètode d'agrupació o *clustering* (K-means) podem veure que els sentits més propers a *telephone.n.01* (el sentit de l'hiperònim del *target*) són *battery.n.02*, *telephone.n.02* (el sentit de *phone*) i *touch\_screen.n.01*, com es veu a la figura 7:

Figura 7. Sentits més propers al sentit *telephone.n.01* en l'opinió d'exemple



El sentit *battery.n.02* (*a device that produceixes electricity; may have several primary or secondary cells arranged in parallel or series*) és compatible com a aspecte del *target*.

Cal tenir en compte, però, que el terme no és *battery* sinó *battery life*. No obstant això, ni en Wordnet ni en la DBpedia no s'ha trobat un sentit de Wordnet per a *battery life*.

### 2.2.2. Detecció a partir d'*opinion words*

La idoneïtat del mètode anterior depèn de l'ontologia que utilitzem. Ja hem vist que *battery\_life* no s'ha detectat com un aspecte perquè no hem trobat el seu *synset* a WordNet. D'altra banda, amb aquesta metodologia, aspectes com la *voice quality* són difícils d'associar a un telèfon perquè la distància ontològica entre un artefacte, una entitat física, com és un telèfon, i una entitat abstracta, com és la veu, i més la seva qualitat, és considerable. No obstant això, quan llegim l'opinió tots dos conceptes ens semblen propers.

Entre els sistemes d'anàlisi de sentiments anomenats *aspect-based opinion mining* (ABOM) *systems*, especialitzats a trobar la valoració dels aspectes de l'entitat, hi ha els que detecten els aspectes a partir d'elements del discurs anomenats *opinion words*. L'*opinion word* és una paraula que té una càrrega negativa o positiva en l'opinió. Quan l'*opinion word* qualifica positivament o negativament un terme que no està allunyat en el discurs, aquest terme és l'aspecte.

En el cas de l'opinió que estem analitzant, l'adjectiu *clear* qualifica positivament l'aspecte *voice quality* i *long* qualifica positivament *battery life*. Ara bé, per la mateixa raó, *phone* hauria de prendre's com un aspecte perquè l'adjectiu *nice* el qualifica positivament, i *mother* també s'hauria de prendre com un aspecte perquè l'adjectiu *mad* qualifica aquest terme negativament. Efectivament, l'adjectiu serveix per a localitzar un candidat a ser aspecte. No obstant això, hi ha més probabilitats que en les opinions sobre productes l'adjectiu *cool* s'apliqui a un aspecte del producte que l'adjectiu *mad*. La idea és agrupar les *opinion words* amb els candidats que són més coherents. *Mother* no hauria de prendre's com a aspecte d'iPhone perquè una persona no pot ser l'aspecte d'un objecte i en canvi *touch screen* sí.

A més de les *opinion words* cal afegir les *opinion phrases*, com *costar un ull de la cara* o *va ser un malson*. No obstant això, per a facilitar l'exposició, anomenarem *opinion words* tant les paraules com les expressions. Hi ha dos tipus d'*opinion words*: els bàsics i els comparatius. Uns adjectius com *bé* o *dolent* serien *opinion words* bàsics i els comparatius serien *millor* i *pitjor*.

### **Diccionaris d'*opinion words***

Per a algunes llengües és fàcil identificar les *opinion words* consultant diccionaris. Es distingeixen tres tipus de diccionaris. El primer tipus és el diccionari de polaritat, el segon tipus és el diccionari d'intensitat i el tercer tipus és el diccionari de context.

Els diccionaris de polaritat són llistes d'*opinion words* classificades de manera binària (positiu-negatiu). El diccionari *General Inquirer* és el més antic d'aquest tipus. També és conegut i utilitzat el *Polarity Lexicon* de Hu i Liu del 2004, elaborat a partir de ressenyes de productes. Aquest diccionari recull expressions de sentiment en xarxes socials, però igual que el *General Inquirer* no recull emoticones ni acrònims. També cal esmentar el lexicó del LIWC, validat per psicòlegs, sociòlegs i lingüistes.

En els diccionaris d'intensitat les *opinion words* estan associades a uns nivells d'intensitat de la polaritat. Un dels més utilitzats és *SentiWordnet*. Es tracta d'un diccionari derivat de Wordnet. Cada *synset* de Wordnet té tres puntuacions numèriques. Una indica el grau de positivitat, l'altra indica el grau de negativitat i l'última el grau de neutralitat del *synset*. La suma de les tres puntuacions dona 1.

També es troben en aquesta categoria el diccionari *AFINN* i *SenticNet*. Termes com *ira*, *adoració* o *respecte* tenen un valor de polaritat en una escala contínua entre el -1 i el +1, segons la seva polaritat negativa o positiva. A *SenticNet* les *opinion words* estan descrites com a conceptes relacionats en forma de xarxa amb informació de sentit comú.

També comptem amb el diccionari *ML-SentiCon*, que recull *opinion words* del català, l'espanyol, el gallec i el basc. Cada paraula té un valor que va del -1 al +1, segons la seva polaritat, i un valor d'ambigüïtat que és el resultat de combinar els diferents valors de polaritat de cadascun dels sentits que pot tenir una paraula.

## Aplicació del mètode

El mètode s'aplica en el *notebook* PLA-2 1.2. Consisteix, primer, a tenir un model d'opinions. El model que utilitzem en el *notebook* és el d'opinions sobre *smartphones* realitzat amb Word2Vec.

Per tant, donada una opinió *O*, cada frase d'*O* que té alguna *opinion word*<sup>1</sup> s'analitza sintàcticament. Els sintagmes nominals d'*O* es consideren candidats a ser aspectes. Amb els candidats i les *opinion words* es procedeix a modelar *O*. Els candidats i les *opinion words* són el vocabulari, i el vectoritzador crea, per a cada terme del vocabulari, un vector amb les similituds d'aquest terme respecte als altres termes.

Recordem que els termes del vocabulari són vectors en el model d'opinions de *smartphones* realitzat amb Word2Vec. Per tant, la forma de calcular la similitud de termes del vocabulari serà amb una mètrica de distància entre vectors com la *cosine similarity*

Una vegada creats els vectors de similitud, s'agrupen amb un mètode de *clustering* en tres classes. La figura 8 mostra el resultat. Com podem veure, la CLASS\_0 és la que agrupa més coherentment les *opinion words* amb els aspectes del *target*.

Figura 8. *Clustering* d'aspectes i *opinion words* en una opinió

```
CLASS_0 ['day', 'phone', 'battery', 'life', 'phone', 'battery life']
CLASS_1 ['bought', 'mother', 'tell', 'bought', 'thought', 'return', 'shop', 'telephone', 'mad']
CLASS_2 ['iphone', 'touch', 'screen', 'voice', 'quality', 'ok', 'touchscreen', 'voice quality', 'nice', 'cool', 'clear']
```

Hi ha uns detalls que hem de comentar. En primer lloc, hem normalitzat el terme *touch screen* amb la forma *touchscreen*, que és la més habitual en el model d'opinions de *smartphones*. D'altra banda, veiem que *voice* i *clear* s'han agrupat coherentment i que, en el model d'opinions sobre *smartphones*, *nice* i *cool* estan més associats a *touchscreen* que, en canvi, amb *phone*, tal com està a l'opinió. Finalment, veiem que hi ha termes com *thought* i *tell* que s'han desambiguat malament. S'ha pres el sentit nominal d'aquestes paraules. Així i tot, podem veure que el mètode ens permet identificar aspectes de l'*smartphone*, tals com el seu *touchscreen* i la qualitat de veu.

### Vegeu també

Consulteu l'apartat 2.1.2.

<sup>(1)</sup>Les *opinion words* són les del diccionari *AFINN*.

### Bibliografia recomanada

Per a aprofundir en el tema de la *cosine similarity* i altres mètriques basades en espais vectorials, és recomanable llegir l'apartat 8.5.1 del llibre de Manning i Schütze (2000) i l'apartat 6.4 del llibre de Jurafsky i Martin (2018).



Cal reconèixer que *battery* no s'ha associat amb l'adjectiu que l'acompanya en l'opinió, que és *long*. La raó és la següent: *long* no és en el diccionari d'*opinion words*, ni tampoc *expensive*, com es deu haver observat. Cal ser conscients que les llistes d'*opinion words* no recullen paraules que, segons el context en el qual s'usen, tenen o no tenen una polaritat. Quan parlem de la vida d'una bateria, *long* té una polaritat positiva, però fora d'aquest context no ha de tenir pas una polaritat. Així mateix, *expensive*, que no apareix en cap CLASS, evoca un aspecte de l'entitat que no s'explicita en l'opinió, que és el preu. No obstant això, per a un classificador automàtic d'opinions negatives i positives, que ha estat entrenat amb un corpus d'entrenament etiquetat segons la seva polaritat, adjectius com *long* o *expensive* poden ser *features* importants per a classificar l'opinió.

### 2.3. Detecció de la polaritat

La detecció de la polaritat, això és, el sentit positiu, negatiu o neutral d'una opinió, és l'aspecte més complex de l'anàlisi de sentiments. És complexa la seva detecció en el nivell de paraula, com ja hem apuntat a l'apartat anterior, i ho és també quant a l'opinió en el seu conjunt.

#### 2.3.1. Dificultats en la detecció de la polaritat amb criteris preestablerts

En el nivell de paraula, es pot considerar que els adjectius i els adverbis carreguen la polaritat de les opinions amb diferent intensitat, intensitat que està declarada en els diccionaris d'*opinion words*. En aquests diccionaris, la càrrega negativa o positiva d'una paraula està preestablerta pel seu autor. Ara bé, alguns adjectius no són *opinion words per se*. Ho són en les opinions de determinats dominis i contextos i el domini i el context determinen també la seva polaritat (per exemple, el cas de *long*, ja comentat a l'apartat anterior). Així doncs, quan un adjectiu o adverbi no està reconegut com una *opinion word* en un diccionari, se'n calcula la tendència positiva o negativa segons la probabilitat de coocurrència amb *opinion words* d'una polaritat determinada. Per exemple, un adjectiu *A* té una tendència negativa si sol coocórrer amb *opinion words* de polaritat negativa. La tendència es detecta amb mètriques com el *PMI*.

Els símbols que no plantegen cap dubte sobre la seva polaritat són les emoticones. L'emoticona :) és universalment interpretada com un símbol de polaritat positiva i :( com un símbol de polaritat negativa. Ara bé, són símbols que no se sotmeten a les regles gramaticals, com les paraules, i per això tenen un ús menys previsible i el seu abast –és un símbol que marca la polaritat de tota l'opinió o solament un aspecte del *target* de l'opinió?– és difícil de determinar.

Altres elements que pertanyen a una gramàtica de l'expressió d'opinions, que és una gramàtica més líquida i mutant, són recursos tipogràfics com les paraules en majúscula, o la repetició de lletres i signes de puntuació, que indiquen sorpresa, perplexitat, etc. (p. ex.: «aquesta pel·lícula és HORRIBLE!!!!!!!!!!

de debò?????»), i també l'ús d'expressions sentides en algun mitjà de comunicació o anunci publicitari de gran impacte (p. ex.: «QUÈ DIUS ARAAA????», «S'HA MATAT EN PEP»). No hi ha cap regla gramatical que estableixi que cal escriure «ÉS HORRIBLE», ni el nombre de vegades que s'ha de repetir el signe d'exclamació o la lletra A. Tampoc hi ha un diccionari que reculli expressions sentides en els mitjans de comunicació que momentàniament es fan populars.

Una altra manifestació de l'ús «creatiu» del llenguatge, sobretot a les xarxes socials, i que afecta la normalització de les dades, és l'ús d'acrònims com a sinònim d'una paraula. N'és un exemple DD per referir-se a *drets*.

Quant a la polaritat en el nivell de l'opinió en el seu conjunt –això és, saber per exemple si l'opinió d'una pel·lícula és positiva o negativa–, la seva detecció tampoc és senzilla. No es pot aplicar la regla següent: en una opinió positiva, solament apareixen *opinion words* positives i en una opinió negativa solament apareixen *opinion words* negatives. L'opinió que hem analitzat n'és un exemple. La impressió global del comentari sobre l'iPhone és positiva, però hi ha *opinion words* positives acompanyades per una negació (*the battery was not long*). Són els anomenats *polarity flippers* perquè canvien la polaritat positiva a negativa i viceversa. A més, malgrat que l'opinió global és positiva, apareixen també *opinion words* negatives, com *expensive*, si s'aplica al preu, i *mad*.

Hi ha propostes de càlcul global de la polaritat, que és el resultat de la suma d'expressions positives i negatives de diferents intensitats, però cal tenir en compte que quan l'opinió és irònica o sarcàstica, un càlcul d'aquest estil pot donar un resultat oposat completament a la intenció de l'opinador (p. ex.: *l'iPhone m'ha costat 900 euros. Baratíssim*). Per desgràcia, el processament del llenguatge natural està encara molt lluny d'aconseguir detectar la ironia i el sarcasme.

### 2.3.2. Detecció de la polaritat amb *machine learning*

Els mètodes de *machine learning* han anat dominant la detecció de la polaritat des del moment en el qual es va veure que era millor no especular sobre la complexitat lingüística de les opinions, i deixar que un classificador automàtic aprenqués a classificar opinions amb comentaris reals etiquetats manualment segons la seva polaritat.

Com ja expliquem en el mòdul anterior, en la classificació amb mètodes de *machine learning* cal fer diversos passos. Per a classificar opinions en fòrums, Twitter, etc., l'execució d'aquests passos pot tenir una complexitat més gran que la que va suposar el projecte de la classificació de titulars. Vegem per què.

## Preprocessament

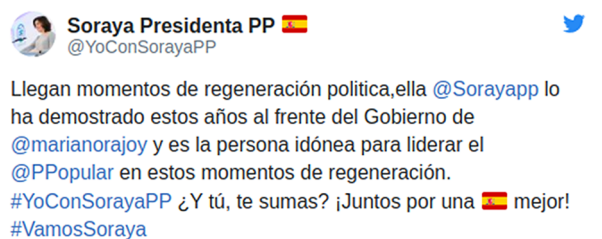
Recordem que per a entrenar un classificador, cal preprocessar les dades amb la finalitat que el sistema aprengui les *features* rellevants. Això significa que cal llevar de les opinions els elements que produeixen soroll i impedeixen que el sistema aprengui bé. En el mòdul anterior, el classificador de titulars aprengia de textos ben escrits. No obstant això, en entorns com Twitter o fòrums d'opinió, la tasca de preprocessament és més complexa. Cal eliminar els *emojis* que no aporten res excepte soroll, cal aplicar correctors que corregeixin errors de teclieg i faltes d'ortografia. També cal eliminar URL i expressions recurrents com RT, a més de distingir comptes d'usuari i eliminar dades personals, etc. En el *notebook* PLA-2, 1.3 veiem alguns exemples de preprocessament.

### Exploració i definició de les *features*

El resultat del preprocessament és un text molt alterat. S'han llevat *emojis*, URL, etc., que eren entre dues paraules i caldrà anar amb compte per a no processar falsos n-grames; és a dir, que no s'uneixin dues o tres paraules que estaven abans separades i que haurien de continuar estant-ho. A més, les etiquetes de vegades es comporten com a paraules en una oració ben formada, però altres vegades etiqueten la piulada sencera. Cal decidir, per tant, si les etiquetes es tracten a part com a etiquetes o es processen com si fossin termes normals.

D'altra banda, a Twitter els *emojis* cada vegada més reemplacen les paraules. És una tendència general que és present fins i tot a les piulades d'institucions, partits polítics, etc., com a la figura 9, on la paraula *Espanya* a *una Espanya millor* és substituïda per la icona de la bandera espanyola.

Figura 9. Exemple d'*emojis* substituint paraules en una piulada



Font: Twitter

Els trets de polaritat que hem comentat (majúscules, repetició de signes d'interrogació, exclamació, etc.) són noves *features* que cal tenir en compte. Ser conscients que la gramàtica de l'expressió d'una opinió és canviant i moltes vegades inesperada suposa fer sempre una exploració de les dades per si hi ha algun tret no previst que pugui millorar els resultats del classificador. Tractarem aquesta qüestió amb més detall en el mòdul següent.

## Tipus de classificadors

Els classificadors més habituals són els anomenats classificadors bayesians. Al *notebook* PLA-2, 1.3 es mostra un classificador que distingeix piulades amb opinions positives i negatives. També s'apliquen classificadors SVM (*support vector machines*) i *logistic regression model*. Aquests classificadors assumeixen que l'ordre dels termes no influeix en la classificació (*bag of words*). No obstant això, hi ha classificadors de xarxes neuronals que sí que tenen en compte l'ordre dels termes. Com en el mètode Word2Vec, els vectors que representen els termes de l'opinió es construeixen tenint en compte el context de paraules ordenades en les quals apareix el terme.

### 2.4. Detecció de l'*opinion holder* i del moment

La detecció de l'*opinion holder* és interessant per a saber els aspectes d'un producte més valorats i criticats segons el tipus de client. Amb un *parser* que reculli els subjectes de verbs d'opinió com *think*, es pot saber el tipus de client que sol expressar una opinió favorable o desfavorable. En l'opinió sobre l'iPhone que hem pres d'exemple tenim el cas d'una mare, representant de les generacions no acostumades a comprar *smartphones*, que valora sobretot el preu. En canvi, el seu fill valora més el so, la càmera, la pantalla, etc.

L'anàlisi de l'*opinion holder* també serveix per a valorar els resultats de decisions tals com obrir un producte a nous països o a noves tipologies de client. També per a millorar els serveis, com els d'un hotel amb clientela majoritària d'un país, que sap els aspectes que la gent d'aquest país valora i critica més. Si tens una clientela majoritàriament anglesa, no n'hi ha prou de tenir les habitacions i el bany nets, o un bufet variat i abundant; cal assegurar-se que les instal·lacions estan adaptades a clients amb mobilitat reduïda, si aquest és un aspecte molt valorat en les opinions sobre hotels per part d'opinadors anglesos. Aquest, precisament, és el tema del primer cas d'ús que veurem.

També és important la detecció de l'*opinion holder* quan s'analitzen polèmiques interessades i campanyes de desprestigi a una marca o a un col·lectiu. La informació sobre el *background* de l'opinador, per exemple el seu partit polític, és útil per a saber la intencionalitat d'aquestes campanyes.

Quan les polèmiques interessades i les campanyes de desprestigi es produeixen a Twitter, la informació del perfil del compte de l'usuari és útil per a obtenir informació sobre l'*opinion holder*, assumint que no sempre en els perfils hi ha informació explícita i veraç. La figura 10 mostra com s'obté la informació del perfil del compte de Donald Trump amb la llibreria Python Tweepy.

La informació sobre els *opinion holders* ha de tractar-se amb totes les precaucions perquè les dades personals han d'estar protegides per la protecció de dades. Les dades del perfil a les xarxes socials són públiques, si així ho decideixen els usuaris, i ells decideixen donar informació real. Quan busquem perfils de

polítics o gent cèlebre i de prestigi, assumim que aquesta informació és veraç. Fora de les xarxes socials, poques vegades els usuaris exposen voluntàriament informació sobre ells mateixos. Un de les dades més fàcils d'obtenir és el país en el qual han publicat una opinió en línia a través de la IP.

Figura 10. *Script* per a obtenir el perfil del compte de Twitter de Donald Trump

```
import tweepy
from tweepy import API
from tweepy import OAuthHandler

consumer_key = XXXXX
consumer_secret = XXXXX
access_key = XXXXX
access_secret = XXXXX

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth)

description_user = api.get_user('@realDonaldTrump').description

print (description_user)
```

El càlcul de la influència de certs opinadors a les xarxes socials també té importància per a predir l'impacte de les seves opinions. Les mètriques que mesuren la influència dels usuaris de xarxes socials queden fora de l'abast d'aquest text.

Quant al temps, aquesta dada està present en totes les publicacions en línia, des de Twitter als fòrums d'opinions de viatges, d'hotels, etc. Amb aquesta dada es pot seguir el *timeline*, els pics de participació i l'estat d'opinió dels participants en una controvèrsia. També es poden fer anàlisis sobre la probabilitat d'escriure un comentari negatiu o positiu segons el dia de la setmana o l'hora del dia. Anàlisis que poden fonamentar fins i tot estudis psicològics dels consumidors i de certs personatges influents del món de la política, la cultura, etc.

### 3. Casos d'ús de l'anàlisi de sentiments

Com ja hem dit en la introducció, exposarem tres casos d'ús. El primer és l'anàlisi d'opinions sobre hotels per a determinar sobre quines coses els hostes són més exigents segons la seva nacionalitat. Aplicarem a aquest cas d'ús la teoria sobre els elements d'una opinió.

Els dos casos restants il·lustren l'anàlisi de sentiments aplicat a opinions que tenen la intencionalitat d'incidir en la reputació d'un producte o del creador d'aquest producte. D'una banda, analitzarem les opinions sobre un programa de televisió i, d'altra banda, tractarem la difusió d'opinions falses.

Com en el mòdul anterior, aquests casos d'ús s'explicaran com si fossin encàrrecs al grup Sense and Suitability (S&S), que va treballar en el predictor de notícies del *New York Times* motivadores de comentaris.

#### 3.1. Coses a millorar a través de les opinions

El grup S&S rep un encàrrec dels directius d'una cadena hotelera: analitzar opinions d'hotels publicades en TripAdvisor. Atès que una anàlisi manual seria massa costós en temps i diners, els propietaris de l'hotel volen que S&S analitzi automàticament les opinions d'hostes russos i espanyols amb la finalitat de saber amb quins serveis i característiques de l'hotel són més exigents. Així sabran quins aspectes han de cuidar especialment i millorar, tenint en compte que la majoria dels seus clients són d'aquestes dues nacionalitats.

Peter reuneix l'equip. Estan tots d'acord que serà un projecte sense gaires complicacions i que es podrà fer en pocs dies. Peter, el coordinador del grup, assigna aquest projecte a Joseph i a Beth.

##### 3.1.1. La metodologia

En la segona reunió amb Peter tots tres discuteixen la metodologia que ha de seguir-se. Prèviament han mirat els tipus de dades que ofereix TripAdvisor de cada opinió i discuteixen sobre quines dades, a més de la nacionalitat de l'*opinion holder*, poden fer l'anàlisi. Aviat arriben a l'acord de prendre la valoració per punts que posa l'opinador com a dada principal. El raonament és el següent: *analitzant les opinions que donen una baixa valoració a un hotel sabrem sobre quines coses els hostes russos i espanyols són més exigents.*

Estableixen, doncs, que el corpus d'anàlisi seran les opinions negatives escrites per opinadors russos i espanyols. Amb aquest criteri decideixen realitzar les tasques següents:

- 1) Recopilar el corpus d'anàlisi.
- 2) Crear dues col·leccions de documents. La primera col·lecció recull les opinions d'*opinion holders* russos. La segona col·lecció recull les opinions d'*opinion holders* espanyols.
- 3) Per a cada col·lecció de documents:
  - identificar el *target* de l'opinió,
  - identificar els aspectes d'aquest *target*,
  - trobar els aspectes més vinculats a *opinion words* de polaritat negativa.

A continuació explicarem com Joseph i Beth van desenvolupant cada pas. D'aquesta manera il·lustrarem l'aplicació de la detecció dels elements d'una opinió, que ha estat el tema de l'apartat 2.

### 3.1.2. Recopilació del corpus d'anàlisi

Per sort, Joseph troba una base de dades pública amb opinions d'hotels de TripAdvisor. Aquesta base de dades té llicència per a ser processada.

La sort no acaba aquí. Cada opinió té associada la nacionalitat de l'opinador i, a més, ha estat etiquetada manualment com una opinió favorable o desfavorable. Això és important perquè Joseph ja es veia passant les opinions per un classificador automàtic de polaritat per a poder recollir, de totes les opinions, les que eren desfavorables. Molt millor tenir el criteri d'un etiquetador humà que el d'un etiquetador automàtic.

#### Vegeu també

Per a més informació sobre el classificador automàtic de polaritat, vegeu l'apartat 2.3.2.

### 3.1.3. Creació de les col·leccions

Aprofitant la dada sobre la nacionalitat de l'*opinion holder*, Joseph crea les col·leccions d'opinions per a cada nacionalitat. La col·lecció d'opinions de clients russos és d'unes 3.900 opinions i la de clients espanyols és d'unes 4.800.

### 3.1.4. Identificació del *target*

Atès que no es volen analitzar les opinions d'un hotel concret, Joseph i Beth decideixen prendre el terme *hotel* com el terme que fa referència al *target* al voltant del qual pivoten els aspectes concrets que fonamenten la mala valoració d'un hotel i que cal detectar.

### 3.1.5. Identificació dels aspectes del *target*

Joseph i Beth discuteixen sobre quina metodologia de les dues que hem explicat a 2.2.1 i a 2.2.2 seria més adequada per a saber els aspectes més criticats pels clients. Joseph, que té una tendència ja coneguda per a solucionar les coses via *machine learning*, diu a Beth que ell sempre és partidari d'embrutar-se *les mans* amb les dades amb les quals treballen abans que recórrer a recursos lèxics predefinits i independents del domini que estan estudiant. Beth, que també sabem que té una predilecció per les taxonomies, les gramàtiques i els recursos que formalitzen el llenguatge, al principi pensava utilitzar Wordnet. No obstant això, canvia d'opinió quan Joseph li diu que tenen un corpus suficient per a fer un model d'opinions de clients russos i espanyols, la qual cosa els permetrà tenir dades interessants bastant aviat. Per tant, decideixen seguir el mètode basat en la presència d'*opinion words* amb un model Word2Vec.

#### Creació dels models

Joseph fa un model de les opinions d'*opinion holders* russos que donen una baixa valoració i fa un altre model del mateix tipus d'opinions amb *opinion holders* espanyols. Els models els fa aplicant el mètode Word2Vec.

#### Elecció dels candidats a aspectes a partir d'un model d'opinions

Joseph i Beth estableixen que els candidats a aspectes del *target* seran els sin-tagmes nominals del model d'opinions que apareixen més de deu vegades i tenen una proximitat amb el terme *hotel* (el *target*) superior a 0,8<sup>2</sup>. A la figura 11 presentem una mostra de candidats a aspecte prenent els termes propers a *hotel* en el model d'opinions rus.

Figura 11. Candidats a aspectes d'un hotel

'make', 'site', 'use', 'trip', 'hotel', 'place', 'window', 'ceiling', 'room', 'cost', 'way', 'day', 'time', 'view', 'noise', 'glass', 'entrance', 'stair', 'space', 'bathroom', 'cafe', 'staff', 'shower', 'problem', 'issue', 'floor', 'smoke', 'corridor', 'option', 'work', 'night', 'station', 'call', 'block', 'towel', 'wall', 'water', 'price', 'breakfast', 'paris', 'stay', 'morning', 'ground floor', 'payment', 'table', 'restaurant', 'dinner', 'pay', 'bar', 'suitcase', 'wifi', 'shampoo', 'personnel', 'reception', 'hour', 'arrival', 'coffee', 'city', 'people', 'bed', 'location', 'guy', 'cleaning', 'change', 'parking', 'walk', 'card', 'design', 'leave', 'hair', 'look', 'area', 'lift', 'food', 'check', 'system', 'furniture', 'bath', 'stuff', 'picture', 'balcony', 'request', 'luggage', 'lady', 'internet', 'gym', 'amount', 'rate', 'charge', 'money', 'family', 'street', 'smell', 'sauna', 'try', 'service', 'facility', 'renovation', 'elevator', 'euro', 'account', 'ventilation', 'carpet', 'face', 'start', 'building', 'train', 'minibar', 'taxi', 'order', 'slipper', 'tv', 'take', 'lobby', 'find', 'door', 'fix', 'help', 'air conditioner', 'toilet', 'offer', 'wash', 'air conditioning', 'speed', 'detail', 'closet', 'experience', 'sink', 'promise', 'double bed', 'connection', 'socket', 'park', 'pool', 'drink', 'menu', 'mention', 'paper', 'receptionist', 'sleep', 'phone', 'desk', 'london', 'reason', 'cause', 'person', 'guest', 'life', 'credit card', 'crowd', 'date', 'spa', 'deposit', 'repair', 'tea', 'pillow', 'client', 'neighbour', 'question', 'bathrobe', 'car', 'lock', 'hall', 'comfort', 'control', 'city center', 'conditioner', 'business', 'cup', 'neighbor', 'case', 'hot water', 'construction', 'return', 'heating', 'air', 'access', 'result', 'dish', 'visit', 'travel', 'laundry', 'mattress', 'tram', 'year

Aquesta és una mostra de més de dos-cents candidats a ser aspectes d'un hotel. Alguns termes, efectivament, són característiques i serveis d'un hotel (*room*, *shower*, *minibar*, *air conditioning*, *sauna*, *receptionist*, etc.), fins i tot podem veure termes relacionats amb aquestes característiques, com *noise* o *smell*. No obstant

#### Bibliografia recomanada

Per aprofundir sobre aquest punt, vegeu *notebook* PLA-1, 3.5.

<sup>(2)</sup>Recordem que el valor de proximitat segons Word2Vec es troba entre el 0 i l'1.



això, Beth i Joseph consideren que és un nombre excessiu, i més tenint en compte que hi ha molts candidats que no es poden interpretar com a aspectes d'un hotel (*trip, lady, family, London, visit, tram, hour, hair*).

Joseph se sorprèn que s'hagin «colat» tants termes que no tenen a veure amb hotels. En la bibliografia que ell ha llegit, quan s'analitzen opinions sobre *smartphones*, i més quan es publiquen a Twitter, no sol haver-hi tants aspectes falsos. No obstant això, a Beth no li sorprèn tant. A ella li encanta viatjar i sol publicar opinions dels hotels on s'ha allotjat. Sap que, quan un es queixa, sol explicar les circumstàncies que l'han portat a una situació frustrant. Per exemple, una vegada va publicar una queixa explicant al principi que va caure un xàfec a la ciutat, que es va mullar tots els cabells i la roba, que no portava paraigua, i que un cotxe a tota velocitat va passar per un toll i la va deixar encara més molla. Quan va arribar a l'hotel va tenir la desagradable experiència de no poder dutxar-se amb aigua calenta ni eixugar-se els cabells perquè l'assegador no funcionava. Per tant, en la seva queixa, Beth va fer referència a una ciutat, un xàfec, els seus cabells, la seva roba, un paraigua, un cotxe i un toll abans de fer esment als aspectes de l'hotel causants de la queixa. Això no sol passar quan un opina sobre un *smartphone* i la restricció de caràcters de Twitter obliga a centrar-se en els aspectes i no permet desenvolupar tot un relat previ a la queixa.

Per tant, qualsevol terme pot ser a prop d'*hotel* en el model Word2Vec, però la proximitat en un model d'opinions sobre hotels no garanteix que aquest terme sigui un dels seus aspectes. Joseph té l'esperança que les *opinion words* seran discriminants dels aspectes. Els aspectes veritables haurien d'estar relacionats amb adjectius com *dirty, expensive, noisy*, etc. Beth li recorda, però, que hi ha *opinion words* com *awful* que es poden aplicar tant a les característiques d'un hotel com a una ciutat i també al viatge en el seu conjunt.

En definitiva, que el tema de detectar els aspectes no és tan fàcil com semblava al principi.

### **Elecció dels candidats a aspectes amb un model de llengua general**

Hem vist que els termes més propers a *hotel* en un dels models d'opinions és una llista de termes heterogenis. Joseph també prova amb *bathroom* i li surten termes com *ceiling* i *pillow*, que, malgrat superar el llindar de proximitat segons el model, no són propers segons el sentit comú.

Beth veu en això una conseqüència del fet que les opinions sobre hotels pertanyen a un domini permeable en el qual entren termes d'altres àmbits de la vida quotidiana. Els coixins no són una característica dels hotels, són en altres

llocs, encara que en les opinions sobre hotels els termes *bathroom* i *pillow* apareguin en contextos semblants. Així doncs, la proximitat segons el model no sempre ho és segons el nostre sentit comú.

Quan Joseph i Beth pensaven que havien arribat a un atzucac, Joseph recorda que Google deixa disponible un model Word2Vec de notícies en anglès. Aquest model es va fer a partir d'un corpus de notícies de Google News, i conté tres mil milions de termes (3,6 GB)<sup>3</sup>. Joseph creu que, en ser un corpus d'ús de la llengua sobre molts temes, aquest recurs els permetrà discriminar els aspectes més relacionats amb el sentit comú.

<sup>(3)</sup>Podeu descarregar-vos el model en aquest enllaç: <https://github.com/mmihaltz/word2vec-googlenews-vectors>

A la taula 1 es pot veure la comparació entre els termes més relacionats amb *hotel* i *bathroom* segons el model d'opinions i els termes més relacionats amb *hotel* i *bathroom* segons el model de Google News. Es pot comprovar que els termes més propers segons el model de Google News són els més reconeixibles pel sentit comú.

Veient això, decideixen prendre com a candidats els sintagmes nominals de les opinions que apareixen més de deu vegades i que són propers al terme *hotel* en el model de Google News. El significatiu és que, amb aquest model, Joseph pot abaixar el llindar de proximitat per a tenir resultats adequats. A la figura 12 es mostren els termes més propers a *hotel* amb un llindar de 0,3. Com es pot veure, el nombre de candidats s'ha reduït considerablement i la majoria fan referència a aspectes d'un hotel (*bar, bathroom, breakfast, restaurant, sauna, receptionist, etc.*).

Figura 12. Candidats a aspectes d'un hotel en les opinions amb el model de Google News

'room', 'bathroom', 'cafe', 'breakfast', 'restaurant', 'bar', 'bed', 'balcony', 'luggage', 'sauna', 'elevator', 'minibar', 'taxi', 'lobby', 'receptionist', 'guest', 'spa', 'hall', 'travel', 'laundry'

Ara bé, cal tenir en compte que el vocabulari del model de Google News no recull termes multiparaula, amb la qual cosa no s'ha pogut calcular la distància dels termes *multiword* que ja havien estat recollits en el vocabulari d'opinions. Beth i Joseph decideixen afegir-los a la llista de candidats. Els termes *multiword* que havien obtingut però que no són en el model de Google News són els següents.

Figura 13. Candidats *multiword* en el model d'opinions que no són en el model de Google News

'ground floor', 'air conditioner', 'air conditioning', 'double bed', 'credit card', 'city center', 'hot water'

La filtració dels candidats a aspecte es farà per la seva relació amb les *opinion words*. ja que aquesta relació ha de calcular-se per a veure quins aspectes són valorats més negativament, decideixen abordar aquesta fase directament amb la llista de candidats obtinguts, excepte *city center*, que descarten d'entrada com a aspecte.

Taula 1. Termes semblats a *hotel* i *bathroom* en els models d'opinions i de Google News

<b>Hotel</b>	
Model opinions	[('order', 0,9998066425323486), ('hour', 0,9997938871383667), ('ready', 0,9997932314872742), ('time', 0,9997906684875488), ('minute', 0,9997877478599548), ('reception', 0,9997860789299011), ('ask', 0,9997844696044922), ('receptionist', 0,9997807145118713), ('anything', 0,9997801780700684), ('paid', 0,9997779726982117)]
Model Google News	[('hotels', 0,7709728479385376), ('Hotel', 0,7388719916343689), ('motel', 0,661712110042572), ('boutique_hotel', 0,660549521446228), ('Holiday_Inn', 0,6501166820526123), ('Radisson', 0,6451879739761353), ('luxury_hotels', 0,6348958015441895), ('Hilton_Hotel', 0,6312208771705627), ('hotel_rooms', 0,628795325756073), ('Comfort_Inn', 0,6243367791175842)]
<b>Bathroom</b>	
Model opinions	[('ceiling', 0,9998364448547363), ('size', 0,9998249411582947), ('quite', 0,9998235106468201), ('carpet', 0,9998181462287903), ('pillow', 0,999807596206665), ('furniture', 0,9998031258583069), ('tiny', 0,9998018741607666), ('door', 0,9997978806495667), ('room', 0,9997972846031189), ('enough', 0,9997962117195129)]
Model Google News	[('restroom', 0,8025195598602295), ('bathrooms', 0,7190953493118286), ('lavatory', 0,6996769905090332), ('washroom', 0,6982483863830566), ('toilet', 0,6814740896224976), ('kitchen', 0,6556941270828247), ('downstairs', 0,6407461762428284), ('shower', 0,6207696199417114), ('bedroom', 0,617497444152832)]

### 3.1.6. Aspectes vinculats a *opinion words* de polaritat negativa

Per a conèixer els aspectes vinculats a *opinion words* de polaritat negativa, primer es prenen com *opinion words* de polaritat negativa les paraules del diccionari *AFINN* que tenen un valor menor que zero. Després, seguint el mètode explicat a 2.2.2, es crea un vocabulari amb els aspectes i les *opinion words* i després un vectoritzador crea, per a cada terme del vocabulari, un vector amb les similituds d'aquest terme respecte als altres termes. Una vegada creats els vectors de similitud de cada terme del vocabulari, aquests s'agrupen per a trobar els aspectes més vinculats a *opinion words* negatius. La figura 14 mostra l'agrupació dels aspectes vinculats a *opinion words* negatius en les opinions de clients russos. Els aspectes estan agrupats en cinc classes.

Figura 14. Agrupacions dels aspectes més vinculats a *opinion words* negatius en opinions de clients russos

CLASS0:	'cafe',	'breakfast',	'bar',	'bed',	'weird',	'pay',	'bad']													
CLASS1:	'room',	'restaurant',	'luggage',	'lobby',	'guest',	'laundry',	'air conditioner',	'uncomfortable',	'problem',	'strange',	'limited',	'empty',	'noisy',	'leave',	'lobby',	'inconvenient',	'poor',	'broken',	'terrible']	
CLASS2:	'minibar',	'travel',	'credit card',	'block',	'weak',	'wrong']														
CLASS3:	'bathroom',	'balcony',	'sauna',	'elevator',	'taxi',	'receptionist',	'spa',	'hall',	'air conditioning',	'double bed',	'awful',	'dirty',	'horrible']							
CLASS4:	'ground floor',	'hot water',	'hard']																	

Es pot veure que els termes més vinculats a *opinion words* negatius fan referència a aspectes relacionats amb la restauració (*breakfast, restaurant, cafe, bar*), els serveis relacionats amb l'*spa* i la sauna, els llits i l'aire condicionat, i també apunten a temes de mobilitat (*luggage, elevator, taxi*).

### 3.1.7. Informe

A partir de la taula comparativa dels clústers obtinguts amb l'anàlisi de les opinions de clients russos i espanyols, Joseph i Beth redacten un informe segons el qual es poden establir els següents aspectes sensibles per als clients d'ambdues nacionalitats:

- serveis de restauració (restaurant, bar, servei de desdijuni, etc.)
- elements de l'habitació (tipus de llit, aire condicionat, bany, etc.)
- serveis de sauna i *spa* (sobretot el tema de l'aigua calenta és important)
- gestió de l'equipatge
- mobilitat dins de l'hotel i transport (per exemple, a l'aeroport)

De les *opinion words* es dedueix que els clients russos i espanyols es fixen en la neteja, la comoditat, el soroll i si els aparells funcionen.

Figura 15. Comparativa de clústers de les opinions de clients russos i espanyols (en vermell els termes comuns i en blau, els aspectes diferents)

Cientes rusos
CLASS0: ['cafe', 'breakfast', 'bar', 'bed', 'weird', 'pay', 'bad'] CLASS1: ['room', 'restaurant', 'luggage', 'lobby', 'guest', 'laundry', 'air conditioner', 'uncomfortable', 'problem', 'strange', 'limited', 'empty', 'noisy', 'leave', 'lobby', 'inconvenient', 'poor', 'broken', 'terrible'] CLASS2: ['minibar', 'travel', 'credit card', 'block', 'weak', 'wrong'] CLASS3: ['bathroom', 'balcony', 'sauna', 'elevator', 'taxi', 'receptionist', 'spa', 'hall', 'air conditioning', 'double bed', 'awful', 'dirty', 'horrible'] CLASS4: ['ground floor', 'hot water', 'hard']
Cientes españolas
['reservation', 'restaurant', 'bar', 'suite', 'bedroom', 'lobby', 'receptionist', 'luggage', 'taxi', 'bad', 'noisy', 'hard', 'uncomfortable', 'poor', 'empty', 'problem', 'limited', 'stop', 'pay', 'lobby', 'leave', 'awful', 'broken', 'terrible', 'wrong', 'horrible', 'strange'] ['worst'] ['difficult', 'negative'] ['conciierge', 'amenity', 'airport', 'waiter', 'ground floor', 'hot water', 'neighbour', 'disappointed', 'cancel'] ['room', 'bed', 'breakfast', 'elevator', 'bathroom', 'guest', 'spa', 'sauna', 'lounge', 'terrace', 'travel', 'air conditioning', 'double bed', 'weird', 'weak', 'dirty', 'lack', 'badly']

Entre els aspectes presentats anteriorment, el minibar, el cafè i el balcó de l'habitació mereixen una especial atenció entre els clients russos. Per la seva banda, el *lounge* i les terrasses mereixen especial atenció per als clients espanyols.

Els aspectes distintius dels clients russos són:

- el tipus de pagament (targeta de crèdit)
- la bugaderia

Per la seva banda, els aspectes distintius dels clients espanyols són:

- la reserva
- els veïns

Cal destacar que, sobretot entre els clients espanyols, és important que no se sentin decebut davant les expectatives generades després de veure les característiques anunciades per l'hotel.

L'informe conclou que, malgrat les diferències, hi ha molts aspectes en comú la supervisió i millora dels quals satisfaran tant els clients russos com els espanyols.

### **3.2. Opinions sobre un programa de televisió**

El dilluns 4 d'abril del 2016 es va emetre a la nit, al Canal 33 de la Televisió de Catalunya, un programa dedicat al tema de la prostitució voluntària. Aquest va ser un dels programes d'una sèrie anomenada *La gent normal*, que s'emetia en aquella cadena setmanalment. Durant aquest programa unes dones que es dedicaven a la prostitució de manera voluntària explicaven el seu dia a dia, les relacions amb els clients, com havien decidit elles mateixes dedicar-se a la prostitució, etc. La finalitat era mostrar que hi havia dones que es dedicaven a la prostitució empoderades, autònomes, a les quals els agradava exercir-la i que reclamaven que es reconeguessin els seus drets com a qualsevol altra persona que es dedica a una altra professió.

#### **Nota**

Aquest cas d'ús està inspirat en un estudi que es va fer sobre aquest programa, dins del projecte PROMEDIAR, liderat per Anna Maria Clua, dels Estudis de Ciències de la Informació i de la Comunicació de la UOC, i que es troba publicat a: Anna Clua Infant; Joaquim Moré (2016). «El debate público sobre prostitución. Estudio de caso de la repercusión de un programa televisivo en la esfera Twitter». A: Ramón Cotarelo García (coord.); Javier Gil (comp.). *Ciberpolítica: Gobierno abierto, redes, deliberación, democracia* (pàg. 291-308). Madrid: INAP.

Televisió de Catalunya encarrega a l'empresa S&S un estudi sobre l'impacte del programa. L'estudi vol comprovar si la temàtica va provocar un debat a Twitter i quin caire va tenir aquest, si favorable o, per contra, negatiu cap a la cadena i cap a les persones que van aparèixer al programa defensant la prostitució voluntària. Peter decideix que aquest estudi el facin Anne i Henry.

Anne i Henry decideixen fer l'estudi amb la planificació següent:

- obtenció de les piulades
- exploració de les piulades per a recopilar dades sobre:
  - tema,
  - aspectes del tema,
  - *opinion holders*,
  - polaritat segons els *opinion holders*,
  - temps en què es van publicar les piulades,
  - preprocessament.
- anàlisi de les dades
- informe

### 3.2.1. Obtenció de les piulades

Televisió de Catalunya va crear el compte @lagentnormal i una etiqueta amb el nom del programa (#lagentnormal). Anne i Henry decideixen recollir les piulades que contenen aquesta etiqueta publicades des del dia en el qual la cadena va començar la promoció del programa, una setmana abans, fins a tres dies després d'emetre'l.

D'entre les diferents llibreries que fan peticions a l'API de Twitter per a obtenir piulades, Anne i Henry trien la llibreria de Python anomenada Tweepy. A la figura 16 es pot veure un exemple de *script* per a fer peticions. És important assenyalar que a l'*script* s'han encriptat les credencials que són necessàries perquè un desenvolupador pugui fer peticions a l'API de Twitter<sup>4</sup>.

<sup>(4)</sup>Per a crear les credencials, consulteu <https://stackoverflow.com/questions/1808855/getting-new-twitter-api-consumer-and-secret-keys>

Anne i Henry decideixen posar *lagentnormal* a la *query* de cerca, sense el símbol d'etiqueta, perquè d'aquesta manera podran tenir també les piulades que fan referència al compte del programa @lagentnormal. Així obtenen un corpus d'anàlisi amb uns 2.200 piulades.

### 3.2.2. Exploració

Anne i Henry decideixen que les dades amb les quals treballaran seran les següents:

- moment de la publicació de la piulada
- compte de l'usuari que publica la piulada o la repiula

- text de la piulada

El moment de la publicació de la piulada és interessant per a fer un seguiment temporal de les reaccions al contingut del programa. El compte de l'usuari és una dada que permet conèixer el *background* de l'opinador i també les interrelacions entre els usuaris que van participar en un debat que està encara per demostrar si es va produir. El text de la piulada és la font d'anàlisi de les opinions que els participants van intercanviar. A la figura 16 es veu com imprimir, precisament, els textos de les piulades obtingudes.

Figura 16. *Script* per a obtenir mil piulades sobre anàlisi de sentiments

```
import tweepy

CONSUMER_KEY = xxxxx
CONSUMER_SECRET = xxxxxxx
ACCESS_KEY = xxxxxxxxx
ACCESS_SECRET = xxxxxxxx

#Autenticación
auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
auth.set_access_token(ACCESS_KEY, ACCESS_SECRET)
api = tweepy.API(auth)

#Búsqueda de twits
query = 'sentiment analysis' # Término que se quiere buscar
max_tweets = 1000 #Número máximo de twits a obtener

#Obtener resultados
searched_tweets = [status for status in tweepy.Cursor(api.search, q=query).items(max_tweets)]
#Lista de resultados

#Imprimir los textos de los twits
for st in searched_tweets:
    print(st.text)
```

Anne i Henry es reparteixen l'exploració de la següent manera: Anne s'ocupa de l'exploració dels textos i Henry s'ocupa dels comptes dels opinadors i del temps de publicació.

## Exploració dels textos

Els textos de les piulades repiulades es repeteixen en els resultats de la cerca de piulades sobre el programa. Henry opina que cal mantenir els textos repetits perquè el comptatge de termes en els missatges repetits serà una bona mètrica per a valorar-ne la rellevància. Creu que com més es repeteix un terme en els missatges, incloent-hi les repiulades, més rellevant és aquest terme en el debat.

Anne proposa processar a part les piulades originals (les que no tenen l'expressió regular *RT*), amb la finalitat de detectar el seu autor i descobrir un patró dels usuaris actius al debat, enfront dels passius, que són els que es limiten a repiular. Henry hi està d'acord.

En els textos dels missatges hi ha URL i *emojis* que Anne i Henry estan d'acord a eliminar. Anne té dubtes sobre si convertir el text a minúscules perquè els canvis majúscula-minúscula són un recurs per a expressar sentiment i emotivitat. Henry proposa unificar el text i posar-lo en minúscules, ja que l'anàlisi està més centrat en els termes que apareixen que en la intensitat amb la qual s'expressen. A més, s'ha adonat que una mateixa etiqueta pot estar en majúscula o minúscula (p. ex.: *#prostitució*, *#Prostitució*).

De totes maneres, es plantegen qüestions noves, relacionades amb la llengua dels missatges. Els missatges estan escrits en català però hi ha també missatges en espanyol. Això significa que cal assegurar-se que les piulades estan en una codificació unicode (UTF-8) i que no s'han perdut les paraules amb vocals accentuades o amb dièresis. Si no és així, cal transformar el text a UTF-8. Ara bé, cal anar amb compte quan s'eliminen els *emojis*, ja que també estan en codificacions diferents, i cal verificar que després de l'eliminació dels *emojis* les paraules conserven els accents i les dièresis.

Una altra qüestió relacionada amb l'ús de dues llengües afecta la normalització de les dades. Dues paraules diferents en els dos idiomes s'explicaran per separat encara que denotin el mateix concepte. La solució seria normalitzar els textos traduint tots els missatges a una sola llengua. No obstant això, Henry, davant el volum de piulades que caldria traduir, creu que la traducció manual no és abordable. S'hauria de recórrer a la traducció automàtica. Un detector automàtic de llengua detectaria la llengua del missatge i la traduiria automàticament a la llengua de destinació, si la llengua d'origen i la de destinació no coincideixen.

Anne creu que si s'adopta aquesta solució la probabilitat d'afegir soroll a les dades és molt elevada. Tenint en compte que moltes piulades tenen faltes d'ortografia i errors de teclieg, el resultat de la traducció automàtica podria ser molt confús. Per exemple, en català *espero que estiguis bé* es tradueix a l'espanyol com *espero que estés bien*, però si l'autor del missatge oblida la titlla (*be*), cosa que sol passar molt sovint, la traducció seria *espero que estés cordero*. Per això, Henry i Anne confien que la proximitat entre les dues llengües permeti identificar les dades rellevants per a l'anàlisi.

No posar accents és una pràctica general en l'escriptura de piulades i opinions en els fòrums, però això afecta la normalització de les dades. És evident que cal normalitzar el terme *prostitució*, independentment que s'hi posi l'accent o no. Per aquesta raó, Henry decideix aplicar un corrector automàtic.

Les etiquetes i els noms de compte també són un tema de debat. Com ja s'ha comentat, hi ha etiquetes que no ho són per al comentari, sinó que són constituents sintàctics que no poden tractar-se a part. Vegem-ne l'exemple següent:

#### Exemple d'etiqueta

#### Enllaç d'interès

Una llibreria Python que serveix per a correcció automàtica és *pyenchant*. <https://github.com/rfk/pyenchant>.



Dones que fan de la #prostitució el seu mitjà de vida perquè així ho volen.

Si es tracta #prostitució a part, es perdria un constituent sintàctic necessari per a entendre l'oració. Anne i Henry decideixen tractar les etiquetes com a termes. De totes maneres, Anne s'adona que hi ha etiquetes que aglutinen paraules que poden ser processades per separat. Es tracta d'etiquetes com, per exemple, #jonosocòmplice. Henry prefereix mantenir l'etiqueta com està i incrementar en 1 la freqüència dels termes aglutinats, això és, s'incrementaria en 1 la freqüència de *còmplice*.

Per la seva banda, els comptes es consideren dades personals la privadesa de les quals ha de salvaguardar-se. Per aquesta raó, es decideix eliminar-les.

### Comptes dels opinadors

Henry aplica l'*script* que obté els perfils d'usuari dels opinadors. Per sort, pot tenir els perfils públics de la gran majoria dels intervinents i, també per sort, la majoria dels perfils tenen dades sobre la professió, les aficions i els moviments socials als quals pertanyen.

En els perfils es poden trobar casos de substitució de paraules per *emojis*. Per exemple, *dones* i *homes* a:

Figura 17

Feminismo defiende q las 🧑🏻 deben tener los mismos DD reales y efectivos que los 🧑🏻

També hi ha etiquetes que són constituents de l'oració i expressions reduïdes.

### Temps de publicació

El moment de la publicació és en el format de Twitter, que és el següent:

Taula 2

Nom del dia	Nom del mes	Dígit del dia	Hora: Minuts: Segons:	+0000	Any

Com que l'objectiu de l'estudi no és tenir un seguiment de les opinions al minut, Henry decideix agafar només el dia i el mes.

#### 3.2.3. Preprocessat

Els textos passen pel corrector automàtic, es canvia el text a minúscules, i els acrònims, els *emojis* i les emoticones se substitueixen per les paraules que representen. Es lleva l'expressió regular RT amb el nom del compte de l'autor de la piulada repiulada, i també es lleven els noms dels altres comptes.

#### Vegeu també

Per a més informació, vegeu l'apartat 2.4.

#### Temps de publicació a Twitter

Fri Apr 08 01:27:13 +0000  
2016

En els llocs on s'han eliminat *tokens* s'hi posa una marca que ho indica. D'aquesta manera evitem que el càlcul de n-grames per a trobar col·locacions es faci amb falses paraules consecutives. La marca s'afegeix a la llista de *stopwords* del català i l'espanyol. També es lleva el símbol del coixinet a l'etiqueta.

Per a l'anàlisi dels perfils s'aplica un anonimitzador que oculta el nom i altres dades personals de l'usuari. L'anonimitzador substitueix les dades sensibles (DNI, comptes bancaris) per uns codis. Amb un *script* de Python o un altre llenguatge de programació, aprofitant la seva llibreria d'expressions regulars, es pot fer la substitució fàcilment. El complicat és quan no es disposa d'una llista de les dades sensibles ni quan les dades són expressions regulars. No obstant això, hi ha eines que ajuden a detectar dades sensibles. Per exemple, les eines que realitzen *named-entity recognition* (*NER recognition*), això és, sistemes que detecten referències a persones, entitats, carrers, etc. *Parsers* com FreeLing tenen incorporat un *NER recognition*<sup>5</sup>. És el sistema que utilitzen Anne i Henry.

### 3.2.4. Anàlisi de les dades

Les dades bàsiques són els termes. Atès que els textos que han d'analitzar-se són multiidioma, Henry decideix buscar els termes aplicant mètodes independents de la llengua. Això suposa no aplicar el mètode de detecció de *patterns* sintàctics, ja que el *parser* ha de configurar-se d'acord amb l'idioma dels textos. La cerca de termes se centra en la cerca d'unigrames<sup>6</sup> i de col·locacions aplicant mètriques com la Pointwise Mutual Information.

La lematització tampoc es pot fer, ja que el lematitzador necessita saber quin PoS té el terme en la llengua del text. Anne veu interessant no lematitzar per una altra raó. *Drets* es lematitzaria com a *dret* però no és el mateix dir *lluïtar pels meus drets* que *lluïtar pel meu dret*.

Anne i Henry han de posar-se d'acord per on començar a analitzar. Henry proposa començar analitzant els textos dels missatges per a trobar els temes i els seus aspectes, però Anne el convenç que és millor començar analitzant els *opinion holders*, això és, saber primer quins usuaris seran més influents en la discussió, conèixer la seva *background*, i analitzar-ne després els missatges per a veure quins temes difonen i com van influir en les opinions dels altres participants.

### Anàlisi dels *opinion holders*

Henry crea un *word cloud* amb els termes més freqüents en els perfils d'usuari:

#### Enllaç d'interès

A <https://github.com/xangis/extra-stopwords> es poden obtenir llistes de *stopwords* extra a les que ja té NLTK, entre aquestes el català.

<sup>(5)</sup>NER recognition <https://talp-upc.gitbooks.io/free-ling-4-0-user-manual/content/modules/ner.html>.

<sup>(6)</sup>Anne i Henry poden filtrar els unigrames que són *stopwords* en espanyol. No disposen d'una llista de *stopwords* del català (p. ex.: *són, què*).

#### Bibliografia recomanada

Per a veure com es fa un *word cloud*, vegeu PLA-1, 3.3.

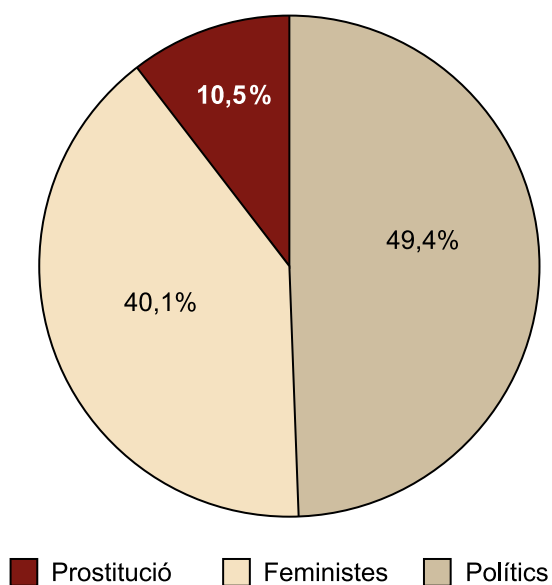


Henry i Anne decideixen agrupar les piulades originals segons tres grups d'usuaris, que són els següents: usuaris de perfil polític, usuaris de perfil feminista i usuaris que es dediquen a la prostitució. Com es pot veure en la gràfica de la figura 19, dominen els usuaris de perfil polític i feminista, mentre que els usuaris que es dediquen a la prostitució són minoria.

### Anàlisi dels temes i els seus aspectes

L'agrupació de les piulades originals segons els usuaris serveix per a respondre a la pregunta següent: els temes i els aspectes dels comentaris són diferents segons el grup al qual pertany el participant?

Figura 19. Distribució de les piulades originals segons els grups d'usuaris



Henry vectoritza les piulades originals de cada grup amb un *tf.idf vectorizer* per a trobar els termes més rellevants a cada grup.

La figura 20 mostra els termes més rellevants del grup dels polítics, ordenats pel seu valor de *tf.idf*.

Figura 20. Termes més rellevants a les piulades del grup dels polítics

Término	TfIdf
tv3amagalatrata	0.401647
prostitució	0.355829
lagentnormal33	0.237219
què	0.213497
parlarà	0.200823
cara	0.160659
día	0.122185
pública	0.122185
veu	0.122185
normalitza	0.120494
trata	0.094888

Fixem-nos que el terme més rellevant en les piulades de les persones vinculades a la política és *tv3amagalatrata*. Aquesta és una etiqueta la comprensió de la qual requereix l'explicació del context en el qual es va emetre el programa.

La televisió pública catalana va emetre el programa el Dia Internacional contra l'Explotació Infantil. Aquesta coincidència va motivar que es relacionés el tema del programa –la prostitució– amb l'explotació sexual. Això explica que aparegui el terme *trata* i que *trata* estigui a l'etiqueta més rellevant. Una etiqueta de denúncia a la televisió perquè, segons els opinadors que el van incloure a les seues piulades, el contingut del programa es podia interpretar com una forma de normalitzar i amagar el drama que és la prostitució en totes les seves manifestacions. Un missatge que no toleraven, sobretot venint d'una televisió pública.

La figura 21 mostra els termes més rellevants del grup feminista, ordenats pel seu valor de *tf.idf*.

Figura 21. Termes més rellevants a les piulades del grup feminista

Término	TfIdf
lagentnormal33	0.545041
què	0.292461
prostitució	0.225993
tv3complicetratamenors	0.225082
tv3complicetractamenors	0.222535
prostitución	0.172818
dones	0.154062
protesta	0.135049
mujeres	0.132937
són	0.119826
infantil	0.106350
trata	0.106350
día	0.102708

Veiem que també es manifesta la protesta i la indignació contra la televisió catalana. Aquesta vegada se l'acusa de ser còmplices del tràfic (en castellà *trata*) de menors. També, i en coherència amb la naturalesa del grup d'opïnadors, es relaciona el tema infantil amb el de les dones.

Finalment, la figura 22 mostra els termes més rellevants del grup de persones relacionades amb la prostitució, ordenats pel seu valor de *tf.idf*.

Figura 22. Termes més rellevants a les piulades del grup de persones relacionades amb la prostitució

Término	TfIdf
lagentnormal33	0.480986
putafeminista	0.296138
putafemista	0.259121
què	0.174904
putas	0.112610
abolicionistas	0.111052
censurar	0.111052
derechos	0.111052
estigma	0.111052

El focus no està en la denúncia al programa sinó en les persones que es denominen feministes abolicionistes, que veuen la prostitució com una forma d'explotació en totes les seves manifestacions i que són favorables a no exposar un plantejament contrari a la televisió pública.

També veiem un terme específic d'aquest grup –*putafeminista*– que es refereix a les dones que es dediquen a la prostitució i que se senten empoderades i lliures de practicar-la, que era precisament el tema del programa. Més que a l'explotació, fan referència a l'estigma social que per a elles representa la visió de la prostitució de les abolicionistes.

### Anàlisi de la polaritat

En els tres grups la polaritat és negativa. El grup dels polítics i de les feministes opinen negativament sobre el programa i sobre la cadena de televisió que l'emet. Fins i tot difonen etiquetes que tenen aparença de consigna de campanya de desprestigi (*tv3amagalatrata*). La polaritat de les piulades del grup de les persones relacionades amb la prostitució també és negativa. Opinen negativament sobre les abolicionistes perquè les censuren i fomenten l'estigma social de les persones que es dediquen a la prostitució.

Anne ha distingit de les piulades de cada grup les *opinion words* positives i negatives. Ha utilitzat com a recurs la traducció al català i al castellà del diccionari de Hiu i Liu. Segons aquest diccionari, la paraula *estigma* té una polaritat negativa, que és conseqüent amb la queixa del grup de les dones que es dediquen a la prostitució.

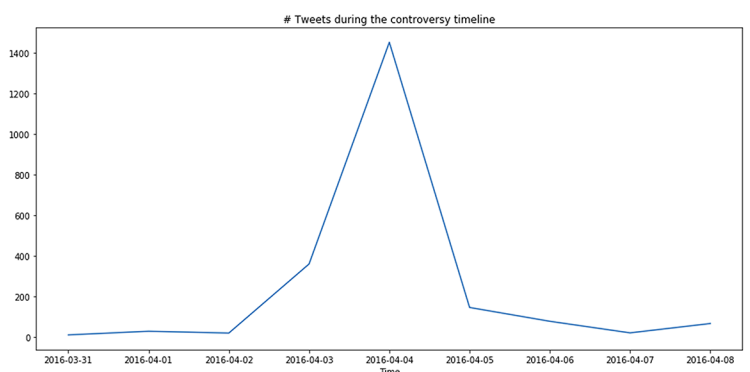
Ara bé, una cosa que crida l'atenció a Anne és que el diccionari no recull paraules de les opinions que ella considera *opinion words*. Per exemple, *humiliacions*, *pedòfiles*, *discriminatòria* i *mercantilització*. D'altra banda, també es dona el cas contrari: *infantil* es pren com una paraula negativa perquè el diccionari recull el sentit d'aquesta paraula equivalent a *irresponsable*, *inconscient*, etc. Un sentit que no té res a veure amb el sentit que té *prostitució infantil*.

Anne també veu una cosa molt curiosa. La paraula *puta* té una polaritat negativa segons el diccionari, però en les opinions de les persones relacionades amb la prostitució aquesta paraula té una càrrega positiva, usada amb orgull per pertànyer a un col·lectiu. La paraula es relaciona amb *empoderades* (amb un *tf.idf* de 0,07 a les piulades de les persones relacionades amb la prostitució, i sense valor en altres grups). En resum, l'anàlisi de la polaritat amb un diccionari d'ús general no sempre s'ajusta del tot al context de les opinions.

## Anàlisi del temps

La següent gràfica mostra el nombre de piulades durant els dies que va durar la controvèrsia. Evidentment, el pic de piulades correspon al dia en què es va emetre el programa. Però el que crida més l'atenció és el fet que la controvèrsia ja s'iniciés un dia abans de la seva emissió.

Figura 23. Piulades publicades durant els dies que va durar la controvèrsia (vegeu *notebook PLA-2*, 1.5)



Henry fa un cop d'ull a les piulades i comprova que el dia abans ja s'escalfaven motors en contra del programa i de la cadena, i també en contra dels qui defensen les postures abolicionistes. Vegem-ne aquí alguns exemples:

### Exemples de piulades iniciadores d'una campanya de desprestigi

Proxenetisme de "la gent normal"...

4 abril Dia contra la prostitució infantil @lagentnormal33 @tv3cat normalitza la prostitució com a professió

@tv3cat @lagentnormal33 la normalització ens involuciona com a societat!  
#tv3amagalatrata @TradxAbolicion

Instamos a @tv3cat a no ser còmplice de la trata de niñas

#Prostitución por voluntad propia. Que el abolicionismo no censure las voces de las trabajadoras sexuales.

qué os parece si mañana le damos caña al HT #Putafeminista junto al HT del programa para comentar?

### 3.2.5. Informe

Henry i Anne redacten un informe en el qual s'expliquen les dades que hem presentat anteriorment. Arriben a conclusions que són d'interès per als estudiosos en ciències polítiques, però no les exposarem aquí. No obstant això, sí que és rellevant la dada que mostren sobre com s'assemblen o s'allunyen les piulades entre els grups. Per a obtenir aquesta dada, creen una col·lecció de tres documents, cadascun amb les piulades d'un grup, els vectoritzen amb un *tf.idf* vectorizer i calculen la *cosine similarity* entre els vectors. El resultat és el següent.

Taula 3

Distància polítics-feministes	Distància polítics-persones relacionades amb la prostitució	Distància feministes-persones relacionades amb la prostitució
0,341	0,170	0,367

Com es veu, els missatges dels polítics s'apropen als missatges del grup feminista, mentre que mantenen una posició més allunyada respecte al grup de les persones relacionades amb la prostitució. No obstant això, la distància entre el grup de persones relacionades amb la prostitució i el grup feminista és similar a la que té aquest últim amb els polítics. Això apunta al fet que el debat es va centrar més en el discurs feminista, la qual cosa es va confirmar en veure que en el total de piulades i repiulades els termes rellevants d'aquest grup eren els més freqüents. D'altra banda, demostra que el fet que es parli més de les idees d'un grup ideològic no significa que se simpatitzi amb ell. Molt sovint, els grups ideològics recorren a la metareferència dels conceptes del grup ideològicament contrari per a reafirmar les seves pròpies posicions.

### 3.3. Detecció d'opinions falses

La pràctica molt estesa de consultar les opinions abans de prendre una decisió ens ha fet molt permeables al criteri dels altres. Conscients de l'efecte que es pot produir quan una crítica negativa s'estén per les xarxes socials, entenent per xarxes socials no solament Twitter i Facebook, sinó també fòrums com TripAdvisor, hi ha opinadors que es dediquen a publicar de forma deliberada



opinions falses amb l'objectiu de danyar la reputació d'una empresa, institució, etc. També existeix l'opinador que publica opinions injustificadament positives amb la finalitat de promocionar i millorar la reputació en línia.

És evident el dany que poden fer les opinions falses. Una campanya orquestrada en contra d'una empresa, un polític, etc., pot arruïnar la seva reputació en molt poc temps. Sobretot quan la reputació està centrada actualment en la reputació en línia. En canvi, si la campanya és positiva, augmentar el grau d'expectatives que aviat es veuran frustrades tampoc és beneficiós.

Detectar les opinions falses és, per tant, una necessitat imperiosa i actualment és una de les aplicacions de l'anàlisi de sentiments amb més demanda.

En aquest subapartat ens ocuparem del següent cas d'ús:

Un cercador d'hotels, conscient del dany econòmic que pot suposar per al seu negoci la publicació d'opinions falses interessades a danyar la reputació de cadenes hoteleres, vol aplicar un detector d'opinions falses abans de publicar-les.

L'empresa S&S és la que rep aquest encàrrec. Peter, el coordinador del grup de lingüística computacional, assigna a Beth i a Joseph aquest projecte.

### 3.3.1. Tipificació del problema

Beth s'ocupa de consultar la bibliografia sobre el tema i observa que, generalment, es distingeixen tres tipus d'opinions falses (Jindal i Liu, 2008).

- 1) Opinions que donen informació no fiable sobre el producte amb la finalitat de promoure o, per contra, perjudicar i danyar la seva reputació.
- 2) Opinions que es focalitzen en l'empresa, la marca, els fabricants o els venedors del producte, però que no expliquen una experiència amb un producte concret. Per exemple, *Odio Microsoft. Mai compraré cap producte que porti aquesta marca.*
- 3) Opinions que indirectament fan esment del producte en forma d'anuncis no explícits i textos que no contenen opinions.

### 3.3.2. Com cal abordar el problema?

Principalment hi ha dos mètodes. El primer és detectar opinions falses trobant trets (*features*) dels textos de les opinions. El segon consisteix a trobar comportaments dels opinadors que surten del patró normal.

El primer mètode reforça sobretot la detecció d'opinions falses del primer tipus, això és, opinions que responen a la intenció d'afavorir o perjudicar una reputació. També reforça les opinions falses del segon tipus. Amb aquest mètode, els dos tipus es tractarien detectant el *target* i els seus aspectes i la seva proximitat amb *opinion words*. A més, s'analitzarien altres trets, com la freqüència

#### Bibliografia recomanada

- N. Jindal; B. Liu. «Opinion Spam and Analysis». *Proceedings of First ACM International Conference on Web Search and Data Mining (WSDM-2008)*, 11-12 de febrer del 2008, Stanford University, Stanford, Califòrnia, USA.
- J. Piskorski; M. Sydow; D. Weiss (2008). «Exploring Linguistic Features for Web Spam Detection: A Preliminary Study». *AIRWeb '08 Proceedings of the 4th international workshop on Adversarial information retrieval on the web*.

de determinades paraules i de categories gramaticals que ja s'han estudiat, com a característics dels missatges brossa. No obstant això, les opinions del tercer tipus, que no són explícites, serien més difícils de detectar.

Per Beth, els pros i contres del primer mètode són els següents:

Taula 4

Pros	Contres
Es poden aplicar els mètodes d'anàlisi de textos que ja s'han aplicat en projectes anteriors (titulars de <i>The New York Times</i> , opinions sobre hotels, controvèrsia a Twitter).	No es té una idea molt clara de quines són les <i>features</i> distintives d'una opinió falsa.
Els elements de l'opinió no haurien de ser difícils de detectar.	Les opinions falses de tipus 3 són més difícils de detectar.
En les opinions de tipus 2, el contrast entre els esments a l'empresa en general però no al producte és abordable.	En les opinions de tipus 1 és difícil de detectar la fiabilitat.

El segon mètode, centrat a detectar comportaments en l'usuari que se surten del normal, podria detectar opinions falses dels tres tipus, si l'esment del *target*, els seus aspectes i els trets lingüístics i estilístics de les seves opinions conformessin un model temporalment dispar respecte al model de l'usuari al llarg d'un temps perllongat. La disparitat seria una conseqüència del comportament inusual de l'opinador. Els pros i contres del segon mètode són els següents:

Taula 5

Pros	Contres
És una oportunitat per a endinsar-se en el món de la lingüística forense.	No tenen experiència en lingüística forense ni a modelar comportaments.
El model de comportament d'un opinador que difon opinions falses pot basar-se en dades quantitatives (nombre d'opinions en un espai de temps, longitud de les opinions, temps de connexió de la mateixa IP), que no requereixen la complexitat de l'anàlisi de text.	Hi ha un treball extra verificant que el model de comportament dels opinadors que difonen opinions falses és aplicable en qualsevol escenari.
L'anàlisi del comportament pot anar acompanyat d'accions concretes que afecten solament el que difon opinions falses (bloquejar-lo, advertir-lo, etc.). De totes maneres, és difícil saber si són opinadors individuals o actuen en grup.	Definir què és el que surt del normal significa també modelitzar el comportament dels opinadors honestos. I cal preguntar-se si els opinadors que segueixen un patró acceptable no són en realitat gent que també opina obeint interessos de tercers.

La decisió sobre quin mètode treballaran depèn sobretot de les dades de les quals disposen per a treballar. Tenen un conjunt d'opinions de polaritat positiva i negativa, ja classificades prèviament com a reals i falses, però sense informació sobre l'opinador. Per tant, desenvoluparan el treball analitzant les dades del text.

Ara bé, hi ha una discrepància entre Beth i Joseph. Beth vol desenvolupar processos d'anàlisi de text, mentre que Joseph considera que, atès que les opinions estan classificades en veritables i falses, es podria aplicar *machine learning* per a entrenar directament un classificador d'opinions falses. Beth diu que, en qualsevol cas, l'anàlisi de text previ és imprescindible per a saber amb quines *features* s'entrenarà el classificador. Joseph pensa que amb unes *features* bàsiques, com la freqüència d'aparició o el *tf.idf* dels *n*-grames, n'hi ha prou per a classificar les opinions. Almenys provar-ho. Al final decideixen que Beth farà l'anàlisi textual, Joseph aplicarà les tècniques de *machine learning* i després contrastaran resultats.

### 3.3.3. Anàlisi textual

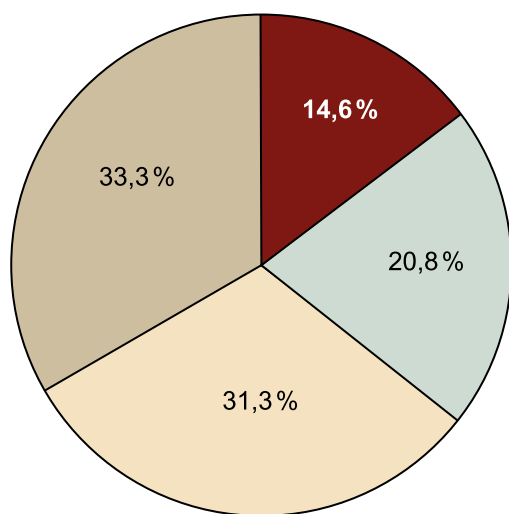
#### Esment a la marca

Beth comença per verificar si és cert que l'esment de la marca, el fabricant o el venedor, sense esmentar un producte concret, és un tret de les opinions falses. En el domini en què estan treballant, el de l'hostaleria, l'equivalent a la marca és el nom de la cadena hotelera. Beth considera com a producte concret els serveis (habitació, bar, wifi, etc.).

Beth pren les opinions veritables i falses (o *fake*) de vint cadenes hoteleres, i les organitza de la següent manera: opinions veritables de polaritat positiva, opinions veritables de polaritat negativa, opinions falses de polaritat positiva i opinions falses de polaritat negativa. La proporció d'aquests tipus d'opinions que contenen el nom de la cadena hotelera es mostra a la figura 24:

Figura 24. Proporció d'opinions veritables i falses (*fake*) que esmenten la cadena hotelera

#### Distribució d'opinions veritables i falses



El resultat és prou significatiu perquè Beth no s'ocupi de distingir les referències al nom de la cadena hotelera de les referències als serveis. Les opinions amb més referències al nom de la cadena estan majoritàriament en les opinions falses. És interessant notar que l'esment a la cadena és més característica de les opinions negatives, siguin aquestes falses o no.

### L'opinion holder

Beth continua analitzant les opinions amb dades textuais. Ara prenent dades sobre l'*opinion holder*. Atès que les opinions negatives falses són les més nocives, Beth es concentra a veure les diferències entre les opinions negatives falses i les negatives veritables.

En aquestes opinions els pronoms de primera persona indiquen el punt de vista de l'opinador. Beth situa aquestes opinions en un espai vectorial amb un *tf.idf vectorizer*. A continuació presentem la comparativa de distribució dels valors de *tf.idf* dels pronoms de primera persona en els dos tipus d'opinions.

Taula 6. Valors de rellevància (*tf.idf*) dels pronoms de primera persona en opinions negatives veritables i falses

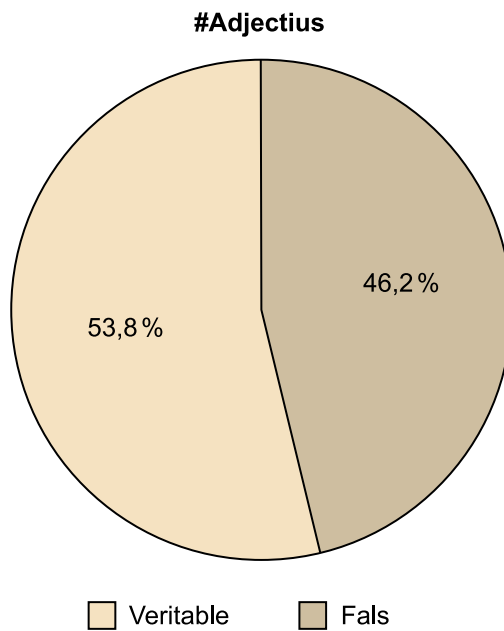
Pronoms	Tf.idf opinions veritables	Tf.idf opinions falses
<i>I</i>	0,570	0,678
<i>em</i>	0,096	0,088
<i>maig</i>	0,181	0,240
<i>myself</i>	0,002	0,004

Segons aquesta taula, veiem que els pronoms *I* i el possessiu *my* són més rellevants en les opinions falses negatives. Beth creu que això es deu al fet que, en les opinions falses, la voluntat expressa d'influir es manifesta en una major presència del «jo» en les opinions.

### Polaritat

Els adjectius són les paraules que tenen més probabilitat d'expressar la polaritat de l'opinió. Beth obté el nombre d'adjectius dels dos tipus d'opinions i s'adona que el nombre d'adjectius és menor en les opinions falses, com es veu a la figura 25:

Figura 25. Proporció d'adjectius en les opinions veritables i falses (*fake*)



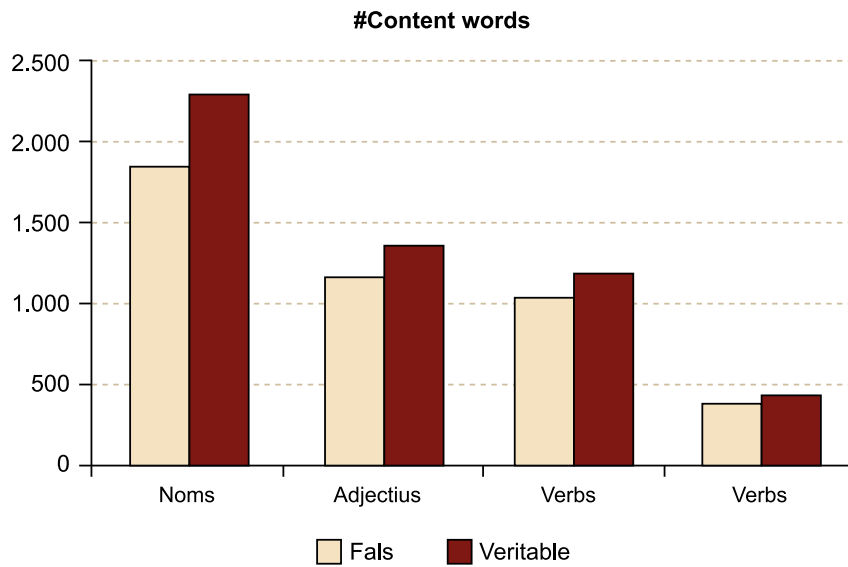
### Una mica d'estilometria

L'estilometria és una disciplina que analitza textos amb la finalitat de donar dades numèriques que són característiques d'un estil o d'un tipus de text. Les dades típiques d'una anàlisi estilomètrica són el nombre d'adjectius o adverbis, per exemple, o bé la freqüència de les paraules, com a dades explicatives d'un tipus de text. L'estilometria s'aplica per a determinar l'autoria d'un text, i en aquest cas Beth creu que es pot aplicar per a caracteritzar un cert estil en les opinions falses negatives.

Beth, després d'analitzar les referències a la marca i la polaritat, té la intuïció que en les opinions falses no hi ha tantes *content words* (paraules amb contingut semàntic) com en les opinions reals, això és, hi ha més paraules que cohesionen el discurs però no aporten contingut. Aquesta intuïció la porta a fer una anàlisi d'estilometria comparant, en els dos tipus d'opinions, el nombre de substantius, verbs, adjectius i adverbis, presos com a *content words*. El resultat es mostra a continuació:

#### Enllaç d'interès

Per a una aplicació de l'estilometria amb una llibreria Python vegeu: <https://github.com/jpotts18/stylo-metry>.

Figura 26. *Content words* en les opinions veritables i falses (*fake*)

Com es pot veure, el nombre *de content words* en les opinions veritables supera el nombre *de content words* en les opinions falses. El nom és la categoria gramatical que supera en un grau major les *fake opinions*.

### Conclusions de l'anàlisi textual

De l'anàlisi textual de les opinions *fake* negatives, feta per Beth, es poden treure les conclusions següents

- 1) En les opinions falses hi ha més esments a la marca (el nom de la cadena hotelera).
- 2) En les opinions falses hi ha més pronoms i determinants possessius de primera persona.
- 3) En les opinions falses hi ha sobretot menys noms.
- 4) En les opinions falses hi ha també menys adjectius i adverbis, la qual cosa té incidència en l'expressió de la polaritat.
- 5) Es pot inferir que en *les fake opinions* hi ha més paraules que no són *content words*, això és, preposicions, articles, etc.

Beth comenta aquests resultats amb Joseph i aquest creu que podrien ser *features* a tenir en compte per al classificador d'opinions falses que ha estat elaborant. De totes maneres, ell ha treballat amb altres *features*, tal com explicarem a l'apartat següent.

### 3.3.4. Classificador d'opinions falses amb *machine learning*

Joseph aprofita les opinions ja etiquetades com a opinions falses i opinions veritables per a entrenar un classificador automàtic d'opinions falses. Com Beth, se centra en la classificació d'opinions negatives falses pel fet de ser les més sensibles.

#### **Establiment de les *features***

Joseph ha anat prenent nota dels comentaris que li ha fet Beth per a establir les *features*. El fet que apareguin més noms i adjectius en les opinions falses li fa considerar el tf (freqüència d'un terme en un document) com *una feature* a considerar.

D'altra banda, Joseph s'adona que quan tracta unigrames, en el rànquing de termes amb major tf.idf apareixen pràcticament els mateixos termes (*room, hotel, stay, service, desk*, etc.) en els dos tipus d'opinions, fora d'algunes excepcions en les opinions veritables, com *bathroom*. Per aquesta raó, vol afegir com a *feature* la longitud dels n-grames. Decideix considerar n-grames de fins a quatre *tokens* per a veure si l'ordre dels n-grames és una *feature* important a l'hora de classificar les opinions.

La freqüència de no *content words* en les opinions falses serà una *feature* a considerar en el cas que les *features* basades en n-grames no fossin prou significatives per a entrenar el classificador.

#### **Mètode de classificació**

Joseph decideix aplicar el mètode bayesià pel fet de ser un mètode que la literatura reconeix que funciona bé per a tasques d'anàlisi de sentiments. Vol provar si també és així per a classificar opinions falses i veritables. El resultat i la seva avaluació seran tractats en el següent mòdul, dedicat a l'avaluació del mètode de classificació de tipus d'opinions.

## Resum

Aquest mòdul s'ha dedicat als mètodes per a extreure els sentiments i les opinions a partir de textos. En primer lloc, hem presentat els elements d'una opinió i hem ofert mètodes de detecció d'aquests elements de forma automàtica.

Hem explicat mètodes de detecció del tema, dels aspectes i de la polaritat aplicant els procediments tractats en el mòdul anterior. Hem aplicat mètodes basats en ontologies i també en diccionaris. Davant l'evidència que els aspectes i la polaritat depenen en gran manera del domini temàtic de l'opinió i també del seu context, a més d'altres factors encara no resolts, com la intenció de l'opinador per ser irònic, hem aprofundit també en la detecció dels elements de l'opinió a partir d'un model del domini i amb mètodes de *machine learning*, classificant la polaritat de les opinions directament de les dades, sense assumpcions predeterminades.

També hem tractat l'anàlisi de l'opinió des del punt de vista de l'*opinion holder*, aprofitant eines d'anàlisi de les xarxes socials i prenent en consideració el grau d'influència dels usuaris a la xarxa social.

A continuació hem presentat tres casos d'ús. El primer cas ha estat l'anàlisi de les opinions dels clients segons la seva nacionalitat, amb la finalitat de millorar els serveis d'un hotel. El segon cas s'ha centrat en una controvèrsia a Twitter sobre el tema d'un programa emès per una cadena de televisió pública. Aquest cas ha servit per a exemplificar els mètodes de detecció dels elements d'una opinió i també com a cas d'ús il·lustratiu de com s'apliquen els mètodes explicats en el mòdul anterior. S'ha vist que els elements d'una opinió són rellevants perquè permeten descobrir com els participants dominen el discurs i difonen el missatge que volen en una xarxa social. El cas d'ús també exemplifica la importància i l'esforç necessari per a desenvolupar el preprocessament de les opinions. Una tasca costosa però absolutament necessària per a obtenir dades fiables.

Finalment, el tercer cas d'ús ha estat la detecció d'opinions falses. A més de ser un tema que preocupa i molt a tots els que depenen d'una reputació en línia, aquest cas ens ha servit per a introduir aspectes lexicomètrics. També ha resultat útil per a veure la necessitat de flexibilitzar el nostre criteri per a trobar les *features* amb les quals entrenar un classificador. Si estàvem acostumats al fet que les *features* fossin a les *content words*, en casos com la detecció d'opinions falses les paraules que no són *content words* també són importants, a més de considerar com a possibles *features* dades que no són tan lingüístiques com de molt baix nivell (presència de majúscules, o la longitud del n-gram, per exemple).



De totes maneres, la flexibilització del criteri ha d'acompanyar-se d'un mètode de classificació que també sigui flexible. Tractarem aquesta qüestió en el proper mòdul, quan exemplifiquem la millora o l'empitjorament dels resultats d'un classificador amb l'addició de noves *features*.

