UNIVERSITAT DE BARCELONA

Grau de Lingüística

Treball de Fi de Grau

Curs 2018-2019

DETECCIÓ AUTOMÀTICA D'ORACIONS AMB NEGACIÓ

Autora: Marina Bolea Centelles

Tutores: Mariona Taulé i M. Antònia Martí

Barcelona, juny 2019

Resum

Aquest treball, emmarcat en els camps de la lingüística computacional i el Processament del Llenguatge Natural (PLN), explica el procés de desenvolupament d'un classificador que detecta si una oració conté, com a mínim, una estructura negativa. El classificador és un programa escrit en Python que s'ha desenvolupat a partir de tècniques d'aprenentatge automàtic supervisat i de les dades del corpus SFU ReviewSP-NEG, un corpus de ressenyes de productes i serveis.

Paraules clau: negació, aprenentatge automàtic, PLN, lingüística computacional

Abstract

This final project, written within the framework of computational linguistics and Natural Language Processing (NLP), describes the process of developing a classifier that detects whether a sentence contains at least one negative structure. The classifier is a program written in Python that has been built using Machine Learning techniques and uses the data found in the SFU ReviewSP-NEG corpus, a corpus of product and service reviews.

Keywords: negation, machine learning, NLP**,** computational linguistics

Agraïments

A la Mariona i la Toni, per donar-me l'oportunitat de treballar en aquest projecte, per convèncer-me que era capaç de fer-ho i per tot el suport que m'han donat durant la realització del treball.

A la Mónica González, per estar sempre disposada a oferir-me la seva ajuda i per fer-se responsable de la meitat de l'anotació que necessitàvem pel treball.

I, molt especialment, al Javier Beltrán, per haver-me ajudat amb el desenvolupament del programa i per l'entusiasme contagiós que té per l'aprenentatge automàtic.

Índex

1. INTRODUCCIÓ

Actualment, la investigació en el camp del tractament automàtic de la negació és una de les línies de recerca amb més rellevància dins de la lingüística computacional i el Processament del Llenguatge Natural (PLN). Poder tractar la negació automàticament és important perquè la negació sovint canvia la polaritat d'una oració, és a dir, n'inverteix el valor de veritat. Aquest canvi en la polaritat és especialment rellevant en camps com el de la medicina, en què és crucial saber si hi ha negació i quina part de la frase s'està negant per determinar la presència o absència d'una malaltia en el pacient. El canvi en la polaritat també és important en el camp de l'anàlisi de sentiments, en què cal destriar les opinions positives de les negatives. La presència de negació pot fer que la interpretació d'una frase passi de favorable a desfavorable, i viceversa.

Les expressions negatives tenen tres components bàsics: els marcadors de negació, l'abast (la part de l'oració que està sent negada) i el focus (aquell element, dins l'abast, que és directament afectat per la negació). En l'àmbit del PLN es poden realitzar diferents tasques relacionades amb la negació. Alguns exemples són detectar si una oració conté estructures negatives o no, identificar els marcadors de negació, l'abast i el focus o avaluar l'efecte que té la negació en la polaritat d'una oració.

La llengua més estudiada i per la qual s'han desenvolupat més recursos des del PLN és l'anglès, i el tractament de la negació no és una excepció dins del camp. Tenint en compte que l'espanyol no només és una de les llengües més parlades al món, sinó també la tercera llengua amb més presència a internet, actualment es considera prioritari disposar d'eines que permetin tractar aquesta llengua automàticament. No obstant això, els recursos de PLN que existeixen per aquesta llengua són escassos, i el nombre de grups de recerca dedicats al tractament computacional de la negació en espanyol és reduït. Per tal de combatre aquesta escassetat de recursos, l'any 2017 es va crear el grup NEGES, que agrupa els investigadors que es dediquen al tractament automàtic de la negació en espanyol. Aquest grup organitza anualment el *workshop* NEGES, en què es proposen una sèrie de tasques l'objectiu de les quals és el desenvolupament de recursos per tractar automàticament la negació en espanyol.

El CLiC (Centre de Llenguatge i Computació) de la Universitat de Barcelona participa a una de les tasques del *workshop* NEGES 2019[1], organitzat dins el marc de la Conferència Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN).

[1] Tota la informació sobre aquest *workshop* es pot trobar a http://www.sepln.org/workshops/neges2019/.

En concret, participa a la Subtasca A, que consisteix en la detecció automàtica dels marcadors de negació. He tingut l'oportunitat de treballar amb els investigadors del grup per resoldre una tasca relacionada amb la detecció dels marcadors de negació: detectar si una frase conté, com a mínim, una estructura negativa. Aquesta és una tasca més simple que la proposada al *workshop* NEGES 2019, però que necessita uns recursos i utilitza uns procediments similars per tal d'arribar a una solució.

Així doncs, en aquest treball presento el procés de desenvolupament d'un classificador que detecta si una frase conté alguna estructura negativa. Aquest classificador s'ha desenvolupat utilitzant tècniques d'aprenentatge automàtic supervisat.

A l'apartat 2 d'aquest treball es presenten els seus objectius, així com les dificultats que pot tenir una tasca aparentment senzilla. A l'apartat 3 es presenta la hipòtesi amb què es treballa. Els antecedents existents en el camp de la detecció automàtica de la negació es presenten a l'apartat 4. L'apartat 5 consisteix en la presentació de la metodologia emprada per desenvolupar el classificador. Aquesta metodologia consta de vàries fases: en primer lloc, cal formular la tasca (5.1). Després comença el procés d'aprenentatge automàtic (5.2), en què s'obtenen les dades necessàries per resoldre la tasca (5.2.1) i es construeix el classificador (5.2.2) a partir d'un algorisme d'aprenentatge automàtic (5.2.2.1) i d'un conjunt de trets (5.2.2.2). Finalment, s'avalua el classificador (5.2.3). L'apartat 6 és una anàlisi dels resultats obtinguts després de l'avaluació del classificador. Per acabar, a l'apartat 7 presento les conclusions sobre la realització del treball.

2. OBJECTIUS

L'objectiu general del treball de recerca ha estat aprendre com funciona tot el procés de construcció d'un programa que resolgui un problema lingüístic i que estigui construït a partir de tècniques d'aprenentatge automàtic supervisat, des de la formulació de la tasca fins l'avaluació final. L'objectiu no és només familiaritzar-se amb la metodologia que permeti desenvolupar un classificador, sinó també amb les eines necessàries per fer-ho, podent així aprofundir en les competències de programació en el llenguatge Python.

L'objectiu concret d'aquest treball ha estat el desenvolupament d'un classificador d'oracions que detecti si una oració conté, com a mínim, una estructura negativa. Entenem per estructura negativa una unitat sintàctica que expressa el valor semàntic de negació. La dificultat d'una tasca que pot semblar senzilla radica en el fet que la presència d'una partícula negativa ('no', 'jamás', 'nadie', 'nada, 'sin', etc.) no implica que una unitat sintàctica tingui aquest valor

semàntic. Els marcadors de negació poden aparèixer en oracions comparatives (1), tenir un valor retòric (2) o no indicar negació si pertanyen a una categoria gramatical concreta (en l'exemple, la categoria nominal) (3b) o apareixen en uns contextos específics (4b).

1. No me gusta tanto como lo otro.
2. El coche lo compré para viajar, ¿no?
3. (a) No sucede **nada**.
   (b) El hotel está localizado en el medio de la **nada**.

'Nada' pot ser un pronom que expressa negació, equivalent a 'ninguna cosa' (3a), però també un nom sense aquest valor semàntic (3b).

4. (a) **En toda mi vida** he hecho una reserva con tanta antelación.
   (b) **En toda mi vida** de estudiante trabajé duro.

L'expressió 'en toda mi vida', i altres expressions com 'en la vida' o 'en su vida', pot tenir un significat literal (com a 4b) o expressar negació (com a 4a).

El classificador que hem desenvolupat és un programa escrit en el llenguatge de programació Python. Aquest llenguatge, creat l'any 1991, és un dels més utilitzats actualment, i té una presència preponderant als camps de la Intel·ligència Artificial (IA) i el PLN. El programa pren com a *input* una oració en espanyol i retorna com a *output* un dels valors d'una variable binària: 1 si l'oració conté alguna estructura de negació o 0 si no en conté cap.

El classificador s'ha desenvolupat utilitzant tècniques d'aprenentatge automàtic. L'aprenentatge automàtic és una branca de la IA que permet que els ordinadors aprenguin pel seu compte, és a dir, que els coneixements que obtinguin vagin més enllà de la programació de què estan dotats. L'aprenentatge automàtic pot ser no supervisat, si simplement es proporciona a l'ordinador un conjunt de dades sense cap mena d'etiqueta o informació suplementària, o supervisat, quan les dades amb què s'entrena el sistema apareixen acompanyades de les etiquetes o els valors que, un cop entrenat el programa, el sistema haurà d'assignar a nous conjunts de dades pel seu compte. Els sistemes d'aprenentatge automàtic supervisat són molt útils per construir sistemes de regressió[2] o classificació, ja que permeten avaluar fàcilment la qualitat dels resultats, i per aquest motiu hem optat per aquesta aproximació a l'hora de construir el nostre classificador.

[2] La regressió consisteix en la predicció d'un valor numèric continu (per exemple, el preu d'una casa o l'esperança de vida) a partir de les dades de l'*input*.

3. HIPÒTESI

La hipòtesi d'aquest treball és que es pot detectar automàticament amb un alt grau d'encert si una oració conté una estructura amb valor de negació aplicant mètodes d'aprenentatge automàtic.

4. ANTECEDENTS

El tractament automàtic de la negació és un àmbit de recerca important dins la lingüística computacional i el PLN, especialment pel que fa als dominis clínic i de l'anàlisi de sentiments. És precisament des d'aquests dominis d'on han sorgit les propostes més pioneres pel que fa a la detecció de la negació i dels elements que hi estan relacionats, com el seu abast o els marcadors de negació. Tot i que la major part d'aquestes propostes tracten la negació a la llengua anglesa, actualment hi ha un nombre creixent de treballs que tracten la negació a la llengua espanyola.

Per tal de poder estudiar i tractar la negació, existeixen diferents corpus que inclouen l'anotació de la negació en espanyol. Podem citar, per exemple, l'UAM Spanish Treebank (Moreno *et al.*, 2003), un corpus de frases de diaris espanyols anotades sintàcticament, ampliat amb l'anotació de la negació, el SFU ReviewSP-NEG (Jiménez-Zafra *et al.*, 2018), un corpus de ressenyes sobre productes i serveis que hem utilitzat en aquest treball, o el corpus NewsCom (Taulé *et al.*, pendent de publicació), un corpus de comentaris de notícies publicades en diaris digitals en què, per primera vegada, s'anota el focus de la negació en espanyol.

Dins el domini clínic, existeixen recursos com els corpus UHU-HUVR (Cruz *et al.*, 2017) i l'IULA Spanish Clinical Record Corpus (IULA-SCRC) (Marimon *et al.*, 2017), dos corpus d'informes clínics anotats amb negació.

Els corpus citats segueixen uns criteris diferents pel que fa a l'anotació de la negació. A Martí i Taulé (2018) es comparen les diferents aproximacions a l'anotació de cada un d'aquests corpus (excepte NewsCom, que encara no s'havia publicat) i es proposa una sèrie de recomanacions per anotar la negació en el futur.

El domini clínic va ser el pioner pel que fa al desenvolupament de sistemes de detecció automàtica de la negació. Cal destacar, entre les propostes sorgides d'aquest domini, l'algorisme NegEx (Chapman *et al.*, 2002), que es va desenvolupar per tractar documents en anglès. Es tracta d'un algorisme d'expressions regulars que té com a objectiu determinar la presència o absència d'una malaltia en informes clínics de pacients.

Treballar la negació (o qualsevol altre fenomen lingüístic) resulta més senzill quan es fa en un domini restringit, i especialment en el domini mèdic, en què hi ha poca ambigüitat lèxica i les negacions es limiten a uns tipus semàntics concrets, com les malalties, els medicaments o les proves mèdiques, que solen ser sintagmes nominals i no frases o verbs (Chapman *et al.*, 2002). D'aquesta manera, NegEx utilitza un algorisme simple i no necessita metodologies sofisticades per tal d'obtenir resultats satisfactoris en el domini pel qual va ser dissenyat, però no té una bona generalització a altres dominis.

NegEx s'ha adaptat a altres llengües, com el suec, el francès, i també l'espanyol (Chapman *et al.*, 2013).

Entre les diferents adaptacions de NegEx a l'espanyol, podem destacar l'aproximació de Costumero *et al.* (2014), en què es van traduir els marcadors de negació identificats a NegEx i es van afegir, també, sinònims i termes extrets de l'anotació manual de textos clínics en espanyol.

Pel que fa al camp de l'anàlisi de sentiments, en l'àmbit de la llengua anglesa podem mencionar el treball de Cruz *et al.* (2015), en què, a partir del corpus SFU Review (Taboada, 2008), un corpus de ressenyes de productes, es va desenvolupar un classificador de SVM (Support Vector Machine)[3] que detecta les expressions negatives i el seu abast.

Pel que fa a la llengua espanyola, podem destacar les propostes sorgides del *workshop* NEGES 2018, associat a la conferència anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Una de les tasques d'aquest *workshop* va ser la detecció de marcadors de negació. Per resoldre aquesta tasca, els organitzadors van proporcionar als participants el corpus SFU ReviewSP-NEG com a corpus d'aprenentatge.

Els diferents grups de recerca van optar per diferents aproximacions. Els investigadors de la UNED, per exemple, van utilitzar tècniques de Deep Learning[4] (Fabregat *et al.*, 2018); el grup de recerca de la UPC va optar per una aproximació d'aprenentatge automàtic basada en l'estadística: els Conditional Random Fields (CRF) (Loharja *et al.*, 2018). Tots dos grups van obtenir bons resultats, i al *workshop* de l'any 2019 s'ha proposat la mateixa tasca, per tal de poder avançar en aquesta línia de recerca.

[3] SVM és un model d'aprenentatge automàtic supervisat que es basa en la geometria.
[4] El Deep Learning és un camp dins de l'aprenentatge automàtic que utilitza algorismes inspirats en l'estructura i el funcionament del cervell.

5. METODOLOGIA

La construcció del classificador que presentem en aquest treball ha seguit els passos que Hladka i Holub (2015) proposen per tal de resoldre la tasca de construcció d'un classificador mitjançant tècniques d'aprenentatge automàtic (vegeu Figura 1). Aquest procés comença amb la formulació de la tasca, que dona pas al procés d'aprenentatge automàtic, en què s'obtenen les dades i es construeix i s'avalua el classificador. El resultat d'aquest procés d'aprenentatge automàtic és el classificador final.

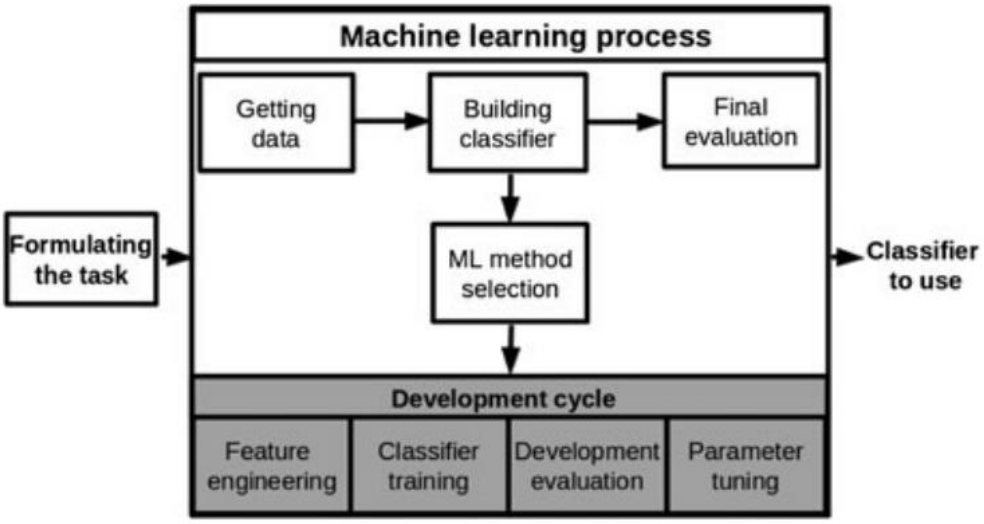

Fig. 1. El procés de construcció d'un classificador (Hladka i Holub, 2015)

5.1. Formulació de la tasca

La tasca que s'aborda en aquest treball és la construcció d'un classificador que detecti frases amb negació. Aquest classificador, com ja s'ha mencionat, és un programa escrit en el llenguatge de programació Python, concretament, en Python3, la tercera versió del llenguatge i la més utilitzada actualment.

L'*input* que espera el classificador és una frase en espanyol, anotada amb lemes i categories gramaticals.

L'*output* del classificador és un dels valors de la variable binària NEG = {0,1}. L'*output* és 1 si la frase conté alguna estructura negativa i 0 si no en conté cap.

5.2. Procés d'aprenentatge automàtic

El procés d'aprenentatge automàtic consta de tres frases: en primer lloc, cal obtenir les dades que s'utilitzaran per tal d'entrenar i d'avaluar el classificador. En segon lloc, es construeix el classificador a partir d'un algorisme d'aprenentatge automàtic i d'un conjunt de trets. Finalment, s'avalua el classificador.

5.2.1. Obtenció de dades

Les dades que s'han utilitzat per entrenar i avaluar el corpus són les proporcionades als participants de la tasca NEGES 2019, una tasca els resultats de la qual es presentaran a l'IberLEF (Iberian Languages Evaluation Forum), un taller que es celebrarà en el marc del Congrés de la SEPLN al setembre de 2019.

El conjunt de dades proporcionat és un subconjunt del corpus SFU Review SP-NEG. Els textos d'aquest corpus són una col·lecció de ressenyes sobre cotxes, hotels, rentadores, llibres, telèfons mòbils, música, ordinadors i pel·lícules extrets de la web Ciao.es. Tot i que l'anotació d'aquest corpus inclou l'àmbit i l'event[5] de la negació, així com informació sobre la polaritat de l'oració i l'efecte que hi té la negació, per elaborar el classificador només hem utilitzat l'anotació proporcionada als participants de NEGES 2019: el lema, la categoria gramatical i els marcadors de negació.

El corpus s'ha dividit en dos subconjunts: el corpus de desenvolupament i el corpus d'avaluació. El corpus de desenvolupament és el que s'ha utilitzat durant tot el procés de construcció del classificador. El corpus d'avaluació només s'ha utilitzat un cop acabat aquest procés, per tal d'avaluar el classificador final i poder valorar els resultats de la seva actuació.

El corpus de desenvolupament, per la seva part, també s'ha dividit en dos subconjunts: el corpus d'entrenament i el corpus d'avaluació de desenvolupament. El corpus d'entrenament és el que s'utilitza per construir el classificador. L'algorisme d'aprenentatge automàtic, a partir de les dades del corpus d'entrenament, aprèn la relació que existeix entre els trets presents a les frases del corpus i el valor corresponent a cada frase (1 si hi ha negació, 0 si no hi ha negació). Amb la informació que ha après sobre aquesta relació, pot assignar valors a les frases que rebi com a *input*.

[5] L'event és l'element afectat directament per la negació. La diferència entre l'event i el focus és que la identificació del primer es basa en la sintaxi i per identificar el segon també es tenen en compte la semàntica i la pragmàtica.

El corpus d'avaluació de desenvolupament és el que s'utilitza per avaluar les diferents versions del classificador, podent així comprovar quin algorisme i quins trets són els que donen millors resultats. Per realitzar aquesta avaluació, es proporcionen al classificador les frases d'aquest corpus com a *input*, i es comparen els valors que els ha assignat el classificador amb la classificació correcta, la present a l'anotació del corpus.

El corpus d'entrenament consta de 264 ressenyes, 33 per cada un dels vuit dominis. El total de frases del corpus d'entrenament és de 7863, de les quals 3862 contenen estructures negatives i 4001 no en contenen cap.

El corpus d'avaluació de desenvolupament consta de 56 ressenyes, 7 per cada domini. De les 1314 frases del corpus d'avaluació de desenvolupament, 423 són frases amb negació i 891 són frases sense cap estructura negativa.

El corpus d'avaluació consta de 80 ressenyes, 10 per cada domini, i té 2203 frases, de les quals 627 són frases amb negació i 1576 són frases sense negació. El corpus d'avaluació va ser proporcionat als participants de la tasca NEGES 2019 sense les etiquetes corresponents a la presència o absència de negació a les frases. Per tant, abans d'utilitzar aquest corpus, vam haver d'anotar-lo, i ho vam fer seguint les directrius presents a la guia d'anotació del corpus SFU ReviewSP-NEG.

A la Figura 2 es presenten el nombre de ressenyes i el nombre d'oracions de cada subconjunt del corpus, així com el nombre d'oracions amb negació i sense.

Corpus	**Ressenyes**	**Oracions**	**Amb negació**	**Sense negació**
Entrenament	264	7863	3862	4001
Avaluació de desenvolupament	56	1314	423	891
Avaluació	80	2203	627	1576

Fig. 2. Dades de cada un dels subconjunts del corpus

Les dades proporcionades als participants de la tasca NEGES 2019 estan en el format CoNLL (Jiménez-Zafra *et al.*, 2019). En aquest format, cada línia correspon a una paraula i cada element de l'anotació apareix en una columna diferent.

```
hoteles_no_1_1  1       1       Me      me      pp1cs000        personal        ***
hoteles_no_1_1  1       2       he      haber   vaip1s0 auxiliary       ***
hoteles_no_1_1  1       3       alojado alojar  vmp00sm main    ***
hoteles_no_1_1  1       4       en      en      sps00   preposition     ***
hoteles_no_1_1  1       5       dos     2       z       -       ***
hoteles_no_1_1  1       6       ocasiones       ocasión ncfp000 common  ***
hoteles_no_1_1  1       7       en      en      sps00   preposition     ***
hoteles_no_1_1  1       8       este    este    dd0ms0  demonstrative   ***
hoteles_no_1_1  1       9       hotel   hotel   ncms000 common  ***
```

Fig. 3. Un exemple de les dades de la tasca NEGES 2019, en el format CoNLL.

A les dades del corpus en format CoNLL, com es pot veure a la Figura 3, la primera columna correspon al nom del fitxer en què apareix la frase (cada fitxer correspon a una ressenya), la segona a la posició que ocupa la frase dins el fitxer (primera frase, segona frase, etc.), la tercera a la posició que ocupa la paraula dins la frase, la quarta a la paraula, la cinquena al lema, la sisena a la categoria gramatical i la setena al tipus de categoria gramatical (auxiliar, preposició, demostratiu, etc.).

Si la frase no conté cap estructura negativa, el contingut de la vuitena i última columna és "***" (aquest és el cas de la frase de la Figura 3). A les frases amb estructures negatives, s'anoten els marcadors de negació, tal i com passa a la frase de la Figura 4.

```
hoteles_no_2_6  9       1       Aun     aun     np00000 proper  -       -       -       -       -       -
hoteles_no_2_6  9       2       estoy   estar   vaip1s0 auxiliary       -       -       -       -       -       -
hoteles_no_2_6  9       3       esperando       esperar vmg0000 main    -       -       -       -       -       -
hoteles_no_2_6  9       4       que     que     cs      subordinating   -       -       -       -       -       -
hoteles_no_2_6  9       5       me      me      pp1cs000        personal        -       -       -       -       -       -
hoteles_no_2_6  9       6       carguen cargar  vmsp3p0 main    -       -       -       -       -       -
hoteles_no_2_6  9       7       los     el      da0mp0  article -       -       -       -       -       -
hoteles_no_2_6  9       8       puntos  punto   ncmp000 common  -       -       -       -       -       -
hoteles_no_2_6  9       9       en      en      sps00   preposition     -       -       -       -       -       -
hoteles_no_2_6  9       10      mi      mi      dp1css  possessive      -       -       -       -       -       -
hoteles_no_2_6  9       11      tarjeta tarjeta ncfs000 common  -       -       -       -       -       -
hoteles_no_2_6  9       12      más     más     rg      -       -       -       -       -       -
hoteles_no_2_6  9       13      ,       ,       fc      -       -       -       -       -       -
hoteles_no_2_6  9       14      no      no      rn      negative        no      -       -       -       -       -
hoteles_no_2_6  9       15      sé      saber   vmip1s0 main    -       -       -       -       -       -
hoteles_no_2_6  9       16      dónde   dónde   pt000000        interrogative   -       -       -       -       -       -
hoteles_no_2_6  9       17      tienen  tener   vmip3p0 main    -       -       -       -       -       -
hoteles_no_2_6  9       18      la      el      da0fs0  article -       -       -       -       -       -
hoteles_no_2_6  9       19      cabeza  cabeza  ncfs000 common  -       -       -       -       -       -
hoteles_no_2_6  9       20      pero    pero    cc      coordinating    -       -       -       -       -       -
hoteles_no_2_6  9       21      no      no      rn      negative        -       -       -       no      -       -
hoteles_no_2_6  9       22      la      lo      pp3fsa00        personal        -       -       -       -       -       -
hoteles_no_2_6  9       23      tienen  tener   vmip3p0 main    -       -       -       -       -       -
hoteles_no_2_6  9       24      donde   donde   pr000000        relative        -       -       -       -       -       -
hoteles_no_2_6  9       25      deberían        deber   vmic3p0 main    -       -       -       -       -       -
hoteles_no_2_6  9       26      .       .       fp      -       -       -       -       -       -       -
```

Fig. 4. Un exemple de les dades de la tasca NEGES 2019, corresponent a una frase amb dues estructures negatives.

Per tal d'elaborar el classificador, necessitem les dades en un altre format: una matriu en què les línies siguin les frases i cada columna indiqui la presència o absència d'un tret a la frase (vegeu Figura 5).

	Tret a	Tret b	Tret c	Tret d	Tret e	Tret f	Tret g	Tret h	Tret i
Frase 1	0	1	1	1	0	0	1	0	1
Frase 2	1	0	0	0	0	1	1	0	1
Frase 3	1	1	1	0	0	0	0	1	0

Fig. 5. Matriu [frase x trets]

Per tant, un cop hem seleccionat les dades, aquestes han estat preprocessades per tal de donar-los el format que necessitem per construir el classificador.

5.2.2. Construcció del classificador

El classificador que hem desenvolupat és un programa que utilitza un model d'aprenentatge automàtic i una sèrie de trets per tal de detectar la presència d'estructures negatives a una oració. Així doncs, per tal de construir aquest classificador, hi ha dues decisions que s'han de prendre: en primer lloc, cal determinar quin model d'aprenentatge automàtic s'utilitzarà. Per tal de prendre aquesta decisió, hem utilitzat els corpus d'entrenament i d'avaluació de desenvolupament per valorar el rendiment de diferents models.

Un cop triat el model d'aprenentatge automàtic que s'utilitzarà, cal decidir quins seran els trets amb què s'entrenarà. Durant el cicle de desenvolupament del classificador, se seleccionen i s'avaluen diferents trets per, finalment, definir els trets amb què s'entrenarà el classificador final.

5.2.2.1.Selecció del mètode d'aprenentatge automàtic

Per construir el classificador, hem provat diversos algorismes d'aprenentatge automàtic, per tal de poder triar el que dona millors resultats. Hi ha una gran varietat d'algorismes que permeten resoldre diferents problemes utilitzant l'aprenentatge automàtic. Aquests algorismes estan construïts utilitzant eines de camps com l'estadística, l'àlgebra, la geometria, la teoria de la probabilitat o la teoria de la informació.

Per triar un algorisme d'aprenentatge automàtic, cal tenir en compte les característiques tant de la tasca que es vol resoldre com de les dades de què es disposa. En el nostre cas, volem desenvolupar un classificador i tenim un corpus anotat. És a dir, tenim una sèrie de dades acompanyades dels valors de la variable NEG = {0,1} que els corresponen. El fet de disposar de dades etiquetades ens ha permès utilitzar algorismes d'aprenentatge automàtic supervisat.

Els diferents algorismes que hem provat, a més de les dades del corpus i les seves etiquetes, necessiten que se'ls defineixin uns trets per aprendre. Els algorismes han d'aprendre la relació

que hi ha entre aquests trets i el valor de la variable NEG = {0,1} associat a les diferents frases. Així, quan se'ls proporcionin noves dades per classificar, els assignaran un valor d'aquesta variable basant-se en els trets que observin a les dades i en les relacions que han establert entre aquests trets i els diferents valors de la variable.

Per tal d'avaluar els diferents algorismes, hem optat per uns trets que sovint s'utilitzen per desenvolupar la primera versió d'un model d'aprenentatge automàtic: una *bag of words* o bossa de paraules, en què s'utilitzen com a trets totes les paraules (o, com en el nostre cas, tots els lemes) presents a les oracions del corpus d'entrenament. Utilitzant aquesta aproximació, cada lema és un tret. Així, a la frase 'No vio nada', el valor dels trets 'no', 'ver' i 'nada' és 1 (indicant la seva presència) i el valor de tots els altres trets és 0, ja que no hi ha cap altre lema present a la frase (vegeu Figura 6).

	'no'	'ver'	'nada'	'recomendar'	'este'	'hotel'
'No vio nada'	1	1	1	0	0	0
'No recomendaría este hotel'	1	0	0	1	1	1
'Es un hotel correcto'	0	0	0	0	0	1

Fig. 6. Exemple simplificat de l'ús de lemes com a trets.

Cal mencionar que, a l'hora d'aplicar els diferents algorismes, no ha fet falta construir-los des de zero. A Python hi ha una sèrie de llibreries que inclouen funcions per resoldre diversos problemes. Per poder realitzar el preprocessament dels textos, i passar de les dades en format CoNLL a la matriu de dades per trets, hem utilitzat la llibreria *pandas*, que s'utilitza habitualment per manipular i analitzar dades.

Per tal d'extreure els trets de les dades, hem utilitzat la llibreria *NumPy*, la llibreria més popular per treballar amb vectors i matrius. I, per aplicar els diferents algorismes que hem utilitzat, hem treballat amb la llibreria *scikit-learn*, que inclou moltes funcions útils per a l'aprenentatge automàtic.

Un dels algorismes que hem utilitzat és l'**SVM (Support Vector Machine)**, un model paramètric (és a dir, que es construeix assumint a priori la forma de la funció que relacionarà les dades de l'*input* i el resultat de l'*output*) basat en la geometria. És un model que sovint s'utilitza per realitzar classificacions binàries. L'objectiu de l'SVM és trobar una línia (o, si es

treballa amb més de dues dimensions, un hiperplà[6]) que separi els dos conjunts de dades (en el nostre cas, les frases amb negació de les frases sense negació) amb el màxim marge possible entre aquesta línia i les dades dels conjunts (vegeu Figura 7).

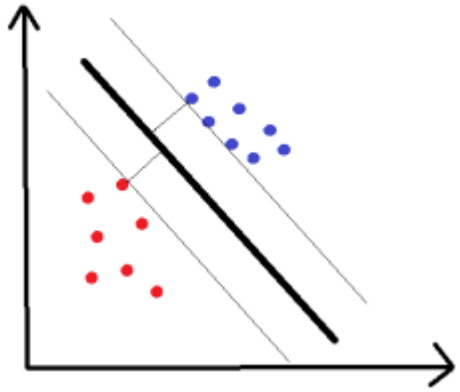

Fig. 7. Una línia que separa dos conjunts de dades amb el màxim marge possible (Maini i Sabri, 2017)

En el cas del classificador que hem construït, una línia no serveix per separar els dos conjunts de frases, ja que cada lema és un tret, i cada tret és una dimensió en l'espai. Com que treballem amb més de dues dimensions, el resultat de l'SVM serà un hiperplà que separarà les frases amb negació de les frases sense negació, maximitzant la distància entre l'hiperplà i els dos conjunts de dades.

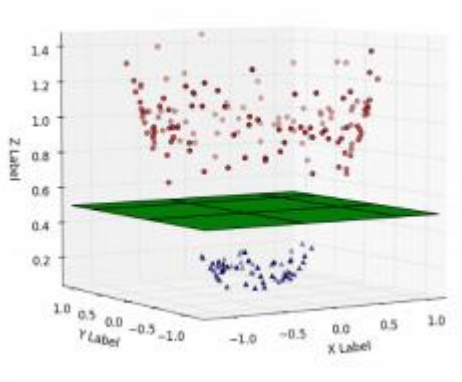

Fig. 8. Un hiperplà que separa dos conjunts de dades en un espai tridimensional amb el màxim marge possible (Maini i Sabri, 2017)

Un altre algorisme que hem provat és el classificador **Naive Bayes**, que sorgeix de la teoria de la probabilitat. Aquest model es basa en el teorema de Bayes (Figura 9) i assumeix independència entre els diferents trets.

[6] Un hiperplà divideix l'espai i té una dimensió menys que aquest (si l'espai té tres dimensions, l'hiperplà que el divideix en té dues; si l'espai en té quatre, l'hiperplà en té tres, etc.).

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Fig. 9. El teorema de Bayes

En el nostre cas, A és un dels dos valors de la variable NEG = {0,1} i B és el conjunt dels trets que utilitzem per entrenar l'algorisme. En el cas de les frases amb negació:

- P (A | B) és la probabilitat que NEG = 1 donats els trets B.
- P (B | A) és la probabilitat que els trets B apareguin en una frase amb el valor NEG = 1.
- P (A) és la probabilitat que una frase tingui el valor NEG = 1.
- P (B) és la probabilitat que els trets B apareguin a una frase.

El model calcula la probabilitat P (A | B) tant per NEG = 1 com per NEG = 0, i selecciona com a *output* el valor amb una probabilitat més alta.

Per tal de construir el classificador, també hem provat d'utilitzar una xarxa neuronal artificial, concretament un ***Multilayer perceptron*** **o perceptró multicapa (MLP)**, il·lustrat a la Figura 10. Les xarxes neuronals artificials són mètodes computacionals que s'inspiren en les xarxes neuronals biològiques que configuren els cervells dels animals. El funcionament de les xarxes neuronals artificials es basa en l'existència d'una sèrie de capes ocultes (*hidden layers*) entre l'*input* i l'*output*. Cada node d'una capa oculta és una funció dels nodes de la capa anterior: els nodes de la primera capa oculta són funcions dels nodes de l'*input*, els de la segona capa oculta són funcions dels de la primera capa oculta, etc. i l'*output* és una funció dels nodes de l'última capa oculta.

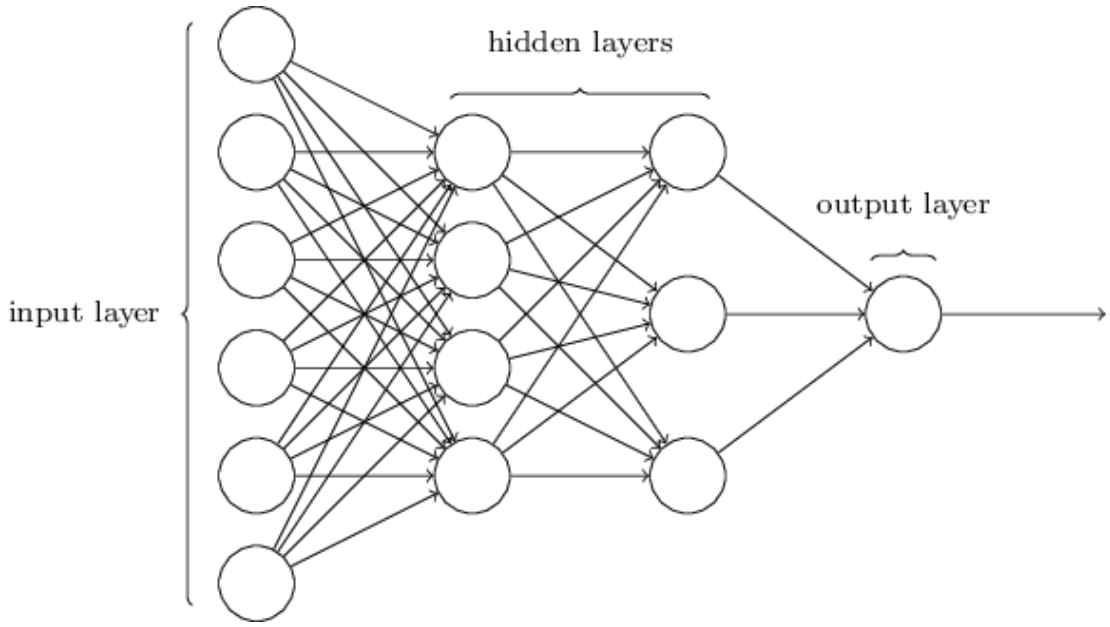

Fig. 10 Un MLP (Cassani, 2017)

També hem provat un algorisme basat en la teoria de la informació: l'**arbre de decisió**. Aquest algorisme no-paramètric (és a dir, sense l'estructura especificada a priori) es basa en l'existència d'una sèrie de nodes que corresponen a condicions que les dades de l'*input* poden complir o no. Les condicions es defineixen a partir dels trets que hem triat. Partint del node arrel, les dades passen per diferents nodes intermedis fins arribar a un node final, que és la classificació que reben (vegeu Figura 11).

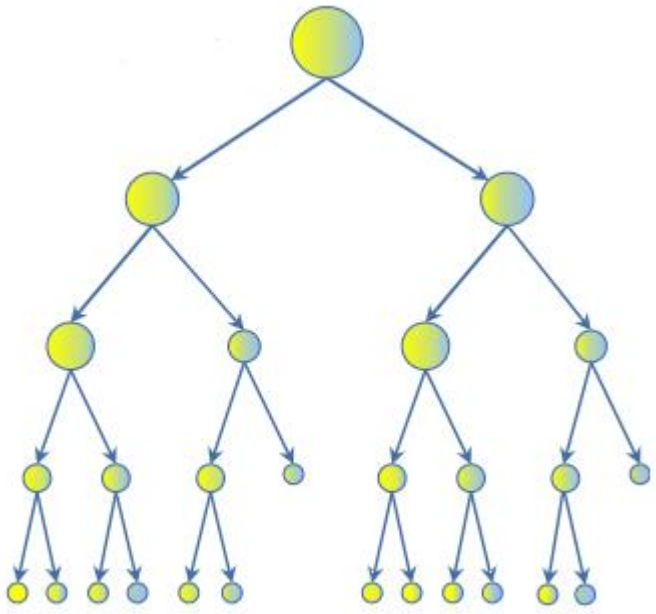

Fig. 11. Un arbre de decisió

Un dels problemes principals dels arbres de decisió és el seu elevat risc d'*overfitting* o sobreajustament. En estadística, es parla d'*overfitting* quan un model funciona molt bé per les dades amb què ha estat entrenat però la seva generalització no és bona: els resultats empitjoren notablement quan treballa amb dades que no ha vist abans.

Per minimitzar aquest risc, s'utilitzen els **random forest**, que agreguen els resultats de varis arbres de decisió entrenats amb diferents paràmetres i seleccionen la classificació que hi apareix amb més freqüència. El *random forest* és un altre dels algorismes que hem provat per resoldre la tasca de la classificació.

Per últim, hem utilitzat un model estadístic: la **regressió logística**. Aquest mètode paramètric és una modificació de la regressió lineal (Figura 12), una funció que a partir de les dades de l'*input* genera un valor numèric continu com a *output*.

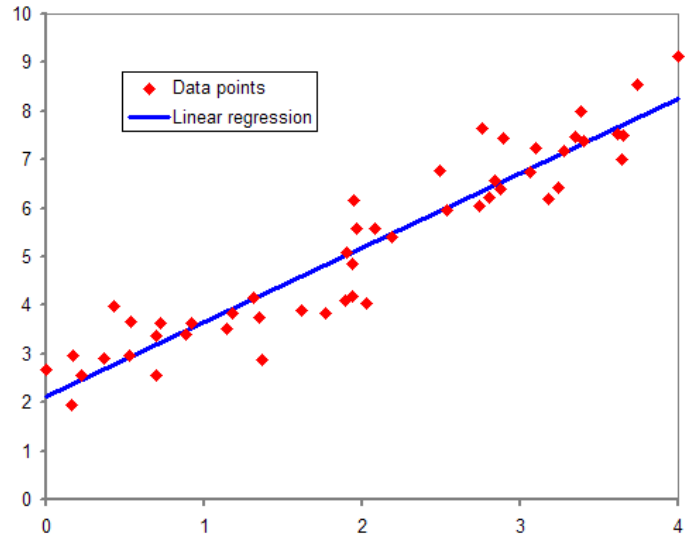

Fig. 12. Un exemple de regressió lineal

La regressió lineal és molt útil per resoldre problemes de regressió, però no soluciona el problema de la classificació. Per resoldre aquesta tasca, ens interessa que l'*output* del model sigui la probabilitat que, donades les dades de l'*input*, el valor de NEG sigui 0 o 1. La probabilitat és un nombre entre 0 i 1, però l'*output* de la regressió lineal pot ser un nombre menor que 0 o major que 1.

Per solucionar aquest problema, la regressió logística insereix la funció de la regressió lineal dins la funció sigmoide (Figura 13), una funció que només pren valors entre 0 i 1.

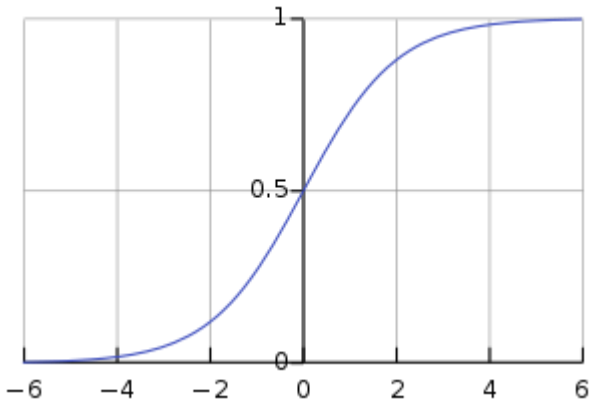

Fig. 13. La funció sigmoide.

Per decidir quin algorisme utilitzaríem per desenvolupar el classificador, vam calcular-ne l'exactitud, utilitzant el corpus d'avaluació de desenvolupament i dividint el nombre d'exemples classificats correctament pel nombre total d'exemples del corpus. A la Figura 14 es donen els resultats dels diferents mètodes utilitzats.

Algorisme	**Exactitud**
Regressió logística	97%
Arbre de decisió	97%
Random forest	95%
MLP (xarxa neuronal)	93%
Naive Bayes	85%
SVM	70%

Fig. 14. L'exactitud de cadascun dels algorismes, amb els lemes com a trets.

La regressió logística i l'arbre de decisió són els dos algorismes que van donar millors resultats, amb una exactitud del 97%. Per desenvolupar el classificador, hem utilitzat aquests dos models amb diferents trets, per poder quedar-nos amb la millor combinació.

5.2.2.2. Selecció dels trets

Un cop decidit quin és el model d'aprenentatge automàtic amb què es treballarà, cal decidir quins trets es faran servir per entrenar aquest model. Quan parlem de trets, fem referència a les variables que l'algorisme tindrà en compte i que li permetran assignar a l'*input* la classificació que li correspon.

La primera versió del classificador, entrenada amb els trets més bàsics, és el que s'anomena *baseline*. En el nostre cas, els trets de la *baseline* són tots els lemes presents al corpus

d'entrenament. L'objectiu del procés de selecció de trets és trobar uns trets que millorin el rendiment de la *baseline*. En el nostre cas, l'objectiu és trobar uns trets que ens permetin arribar a una exactitud de més del 97%, que és el valor que hem obtingut a la *baseline*, tant amb l'arbre de decisió com amb la regressió logística.

El procés de selecció de trets segueix uns passos que es repeteixen per cadascun dels conjunts de trets que s'utilitzen: en primer lloc, es defineixen aquests trets i s'extreuen del corpus; en segon lloc, i utilitzant el corpus d'aprenentatge i els trets escollits, s'entrena l'algorisme. Per últim, s'avalua el rendiment del model amb el corpus d'avaluació de desenvolupament.

Com que els resultats que donava la nostra *baseline* ja eren molt bons, la majoria dels trets que hem provat han empitjorat el seu rendiment. Els trets que no milloren els resultats de la *baseline* són:

- Els lemes més freqüents.
- El nombre de majúscules, de signes de puntuació o de signes numèrics.
- Els lemes i les categories gramaticals, tractats com a trets independents entre si.

L'ús dels lemes i les categories gramaticals com a trets independents empitjora els resultats de la *baseline*. En canvi, sí que s'observa una millora en els resultats quan s'utilitzen com a trets els parells [lema, categoria gramatical] presents al corpus. Per exemple, ['tarjeta', 'ncfs000'] o ['no', 'rn'], que relacionen el lema 'tarjeta' amb la categoria gramatical de nom comú femení singular i el lema 'no' amb la categoria gramatical d'adverbi negatiu. Aquests resultats són encara millors si, a més d'aquests parells de lema i categoria gramatical, s'afegeix com a tret la longitud de la frase.

L'arbre de decisió és l'algorisme amb què s'obté un valor d'exactitud més alt (98%) utilitzant aquests trets. Aquest model només classifica incorrectament 24 frases de les 1314 del corpus d'avaluació de desenvolupament.

Així doncs, el classificador final és un arbre de decisió que utilitza com a trets els parells [lema, categoria gramatical] i la longitud de la frase.

5.2.3. Avaluació final

Un cop establert quin seria el classificador final, l'últim pas del procés d'aprenentatge automàtic ha estat avaluar els resultats d'aquest classificador utilitzant el corpus d'avaluació.

El resultat que hem obtingut ha estat una exactitud del 96%. De les 2203 frases del corpus d'avaluació, 87 han estat classificades erròniament, entre les quals hi ha 45 falsos positius (frases que no contenen negació però que han estat classificades com a frases amb negació) i 42 falsos negatius (frases que contenen estructures negatives que el classificador no ha sabut detectar).

Si comparem aquests resultats amb els obtinguts utilitzant el corpus d'avaluació de desenvolupament, en què l'exactitud és del 98%, observem una baixada de dos punts percentuals. Aquesta baixada no és gens estranya. L'algorisme d'aprenentatge automàtic i el conjunt de trets que configuren el classificador final han estat escollits basant-nos en els resultats que donaven utilitzant el corpus d'avaluació de desenvolupament. Això implica que el classificador tingui un cert biaix a favor de les dades del corpus d'avaluació de desenvolupament.

Pel que fa als falsos positius i als falsos negatius, hi ha una sèrie d'estructures que freqüentment es troben en frases amb una classificació errònia. Pel que fa als falsos positius, cal destacar les estructures comparatives que contenen una partícula negativa però que no tenen valor semàntic de negació. En aquests casos, el classificador sovint confon l'estructura comparativa amb una estructura negativa (5).

5. Si eres una persona sencilla y no tan exigente como yo es el computador perfecto para tu hogar y estilo de vida.

Les preguntes que contenen una partícula negativa que té un valor retòric també són, de manera freqüent, classificades erròniament com a frases que contenen una estructura negativa (6).

6. ¿No es desesperanza lo que ahora mismo siento?

Pel que fa als falsos negatius, l'error que apareix de manera més freqüent està relacionat amb les estructures contrastives, en què el classificador no considera com a negativa l'estructura que, de fet, nega el primer element del contrast (7).

7. Para llegar a la soledad no deseada, sino impuesta, pocos atajos tan directos como el dolor.

També és bastant freqüent que el classificador no detecti la negació a les estructures en què apareix el lema 'ninguno' (8).

8. Ninguna tarjeta de memoria para este teléfono.

6. ANÀLISI DELS RESULTATS

La hipòtesi d'aquest treball és que es pot detectar automàticament i amb un alt grau d'encert si una oració conté una estructura amb valor de negació aplicant mètodes d'aprenentatge automàtic. Aquesta hipòtesi ha estat verificada: hem pogut desenvolupar un classificador que detecta les frases que conten estructures negatives amb un 96% d'exactitud.

Pel que fa als errors que comet el classificador, la majoria no són aleatoris, sinó que corresponen a estructures recurrents: el contrast, la comparació, les preguntes que utilitzen partícules negatives amb valor retòric i lemes concrets com 'ninguno'. D'aquesta manera, si es volgués millorar l'exactitud del classificador, s'hauria d'intentar resoldre la classificació d'aquestes estructures concretes, ja sigui proporcionant més exemples per entrenar el classificador, o bé utilitzant regles.

També seria interessant avaluar el classificador utilitzant altres corpus, per veure si els resultats que dona són equivalents als resultats obtinguts utilitzant el SFU ReviewSP-NEG, especialment pel que fa a corpus que pertanyin a altres dominis i tinguin un contingut molt diferent.

7. CONCLUSIONS

L'objectiu principal del meu treball de recerca ha estat aprendre com funciona tot el procés de construcció d'un programa que resolgui un problema lingüístic i que estigui construït a partir de tècniques d'aprenentatge automàtic supervisat. El desenvolupament d'un classificador que detecta les frases que contenen estructures negatives m'ha permès complir aquest objectiu.

No només he pogut aprendre quins són els passos que cal seguir per construir un programa d'aquestes característiques, sinó que també he pogut familiaritzar-me amb diferents models d'aprenentatge automàtic, que crec que poden oferir moltes possibilitats pel que fa a la recerca en el camp de la lingüística.

A més, he pogut entendre com funciona la construcció d'un programa complex en el llenguatge Python. No és una tasca fàcil, i vull agrair, un cop més, al Javier Beltrán haver-me ajudat a realitzar-la.

Python i les seves llibreries, que inclouen una gran diversitat de funcions, permeten aplicar algorismes com els que hem utilitzat per construir el classificador sense haver d'adquirir els coneixements matemàtics o informàtics que serien necessaris per entendre el seu funcionament

en profunditat. Per aquest motiu, trobo que val molt la pena seguir aprenent a programar en aquest llenguatge.

Aquest treball no hagués estat possible sense l'existència de corpus anotats amb informació lingüística. Tenint en compte les possibilitats que ofereix l'aprenentatge automàtic supervisat, considero que val molt la pena continuar desenvolupant aquests recursos, que poden ajudar a resoldre automàticament tasques lingüístiques complexes.

Per últim, més enllà d'aprendre el procés de desenvolupament del classificador, he pogut entendre com funciona la recerca en els camps de la lingüística computacional i el PLN, i aquest treball m'ha mostrat la importància que té el treball en equip quan es tracta de realitzar projectes de recerca.

8. BIBLIOGRAFIA

Altuna, B., Cruz Díaz, N. P., i Parra Calderón, C. L. (2017). Negation Detection in Spanish: Past, Present and Future. A *Taller de NEGación en ESpañol, NEGES-2017 (*p. 7-12), SEPLN-2017, Múrcia.

Cassani, R. (2017). Multilayer perceptron example. Consultat el 9 de maig a https://github.com/rcassani/mlp-example

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., i Buchanan, B. G. (2002). A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *Journal of Biomedical Informatics, 34,* 301–310.

Chapman, W. W., Hilert, D., Velupillai, S., Kvist, M., Skeppstedt, M., Chapman, B. E., ... i Deleger, L. (2013). Extending the NegEx lexicon for multiple languages. *Studies in health technology and informatics, 192*, 677.

Costumero, R., López, F., Gonzalo-Martín, C., Millan, M., i Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. A *International Conference on Brain Informatics and Health* (p. 366-375). Cham: Springer.

Cruz, N., Morante, R., Maña López, M. J., Vázquez, J. M., i Parra Calderón, C. L. (2017). Annotating negation in Spanish clinical texts. A *Proceedings of the workshop computational semantics beyond events and roles* (p. 53-58).

Cruz, N. P., Taboada, M., i Mitkov, R. (2015). A Machine-Learning Approach to Negation and Speculation Detection for Sentiment Analysis. *Journal of the Association for Information Science and Technology, 67(9),* 2118-2136.

Fabregat, H., Martinez-Romo, J., i Araujo, L. (2018). Deep Learning approach for Negation Cues Detection in Spanish. A *Proceedings of NEGES 2018: Workshop on Negation in Spanish* (p. 43-48), SEPLN-2018, Sevilla.

Guzzi, E., Martí, M. A., Nofre, M., i Taulé, M. (2018). *Guidelines for the annotation of negation in Spanish*.

Hladka, B., i Holub, M. (2015). A Gentle Introduction to Machine Learning for Natural Language Processing: How to Start in 16 Practical Steps. *Language and Linguistics Compass, 9(2)*, 55-76.

Loharja, H., Padró, L., i Turmo, J. (2018). Negation Cues Detection Using CRF on Spanish Product Review Texts. A *Proceedings of NEGES 2018: Workshop on Negation in Spanish* (p. 49-54), SEPLN-2018, Sevilla.

Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Cruz-Díaz, N., Morante, R. (2019). NEGES 2019 task - Negation in Spanish - Subtask A: Negation cues detection. *Workshop NEGES 2019*, SEPLN-2019, Bilbao.

Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Alfonso, Ureña-López, L. A., i Martí, M. A. (2018). SFU Review SP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation, 52(2)*, 533-569.

Maini, V. i Sabri, S. (2017). Machine Learning for Humans. *En línia:* https://medium.com/machine-learning-for-humans

Marimon, M., Vivaldi, J., i Bel, N. (2017). Annotation of negation in the IULA Spanish clinical record corpus. A *Proceedings of the workshop computational semantics beyond events and roles* (p. 43-52).

Martí, M. A., Taulé, M. (2018). Análisis Comparativo de los Sistemas de Anotación de la Negación en Español. A *Proceedings of NEGES 2018: Workshop on Negation in Spanish* (p. 23-28), SEPLN-2018, Sevilla.

Moreno, A., López, S., Sánchez, F., i Grishman, R. (2003). Developing a syntactic annotation scheme and tools for a Spanish treebank. A *Treebanks* (p. 149-163). Dordrecht: Springer.

NEGES. (2019). NEGES 2019 task. Consultat el 7 de juny a http://www.sepln.org/workshops/neges2019/.

Taboada, M. (2008). The SFU Review Corpus. Consultat el 7 de juny a http://www.sfu.ca/~mtaboada/SFU_Review_Corpus.html.

Taulé, M., Nofre, M., González, M., i Martí, M. A. (pendent de publicació). Focus of negation: its identification in Spanish.

UNIVERSITAT DE BARCELONA

Coordinació d'Estudis
Facultat de Filologia

Gran Via
de les Corts Catalanes, 585
08007 Barcelona

Tel. +34 934 035 594
fil-coord@ub.edu
www.ub.edu

Declaració d'autoria

Amb aquest escrit declaro que sóc l'autor/autora original d'aquest treball i que no he emprat per a la seva elaboració cap altra font, incloses fonts d'Internet i altres mitjans electrònics, a part de les indicades. En el treball he assenyalat com a tals totes les citacions, literals o de contingut, que procedeixen d'altres obres. Tinc coneixement que d'altra manera, i segons el que s'indica a l'article 18, del capítol 5 de les Normes reguladores de l'avaluació i de la qualificació dels aprenentatges de la UB, l'avaluació comporta la qualificació de "Suspens".

Barcelona, a 14 de juny de 2019

Signatura: