Clasificación translingüe para la detección de depresión en usuarios de Twitter

por

Denys Laritza Coello Guilarte

Tesis sometida como requerimiento parcial para obtener el grado de

Maestra en Ciencias Computacionales

Instituto Nacional de Astrofísica, Óptica y Electrónica

Agosto, 2019

Tonantzintla, Puebla

Supervisores:

Dr. Luis Villaseñor Pineda

Dr. Manuel Montes y Gómez

Coordinación de Ciencias Computacionales

INAOE

Dra. Rosa María Ortega Mendoza

Universidad Politécnica de Tulancingo

Índice general

Índice de figuras

Índice de tablas

Agradecimientos

Esta investigación fue realizada gracias al apoyo otorgado por el Consejo Nacional de Ciencia y Tecnología (CONACYT),a través de la beca No.869498.

Agradezco inmensamente a mis asesores: Dr. Luis Villaseñor Pineda, Dr. Manuel Montes y Gómez y la Dra. Rosa María Ortega Mendoza por todo el conocimiento aportado, su disposición, constancia y compromiso con la investigación. Permitiendo así que fuera un proceso exitoso y muy instructivo. Gracias por todas sus críticas y comentarios que siempre fueron enfocadas al mejoramiento de la investigación.

Agradezco al INAOE, a sus trabajadores y a todos los profesores que nos trasmitieron sus conocimientos para alcanzar la meta final.

Agradezco a toda mi familia, mi esposo y mi hijo que me acompañaron y ayudaron durante todo el proceso. A mis padres que aunque lejos nunca dejaron de colaborar y estar al pendiente de mis avances. De manera general a todos aquellos que de una forma u otra colaboraron con la investigación.

Resumen

La depresión es un desorden mental con fuerte impacto en la vida social y económica de las personas. Los síntomas que presentan las personas que la padecen están relacionados con su comportamiento diario incluyendo la forma en la que se expresan. En los últimos años, las redes sociales han sido un medio a través del cual las personas comparten sus sentimientos y estados de ánimo. Esto ha propiciado el desarrollo de varias investigaciones, las cuales han explorado el análisis del contenido de las redes sociales para identificar usuarios con depresión. Todas ellas siguiendo una estrategia de aprendizaje supervisado soportado en la disponibilidad de los datos etiquetados para el entrenamiento. Desafortunadamente, el proceso de recolectar y etiquetar datos para el entrenamiento es muy complejo y costoso. Motivados por este problema, y basados en la idea de que a pesar de las diferencias entre lenguajes las personas que padecen de depresión comparten y expresan información similar, en este trabajo, se propone un enfoque translingüe basado en la idea de que los datos etiquetados ya existentes en un idioma específico, pueden ser aprovechados para detectar depresión en otros idiomas. El método propuesto está basado en un proceso de alineación a nivel de palabra. Particularmente se proponen dos representaciones, las cuales permiten capturar la correspondencia entre ambos idiomas. Para evaluar el enfoque propuesto, fueron utilizados Tweets en inglés y español como los datos fuente y objetivo respectivamente. Después de una primera etapa de clasificación, se propone una segunda etapa de re-etiquetado con el objetivo de mejorar las etiquetas de clase de cada documento. Este proceso de re-etiquetado usa las etiquetas

asignadas por el primer clasificador, las somete a un proceso de refinación basado en diccionarios de palabras para generar un nuevo conjunto de entrenamiento, y posteriormente construir un nuevo clasificador. Los resultados obtenidos superan la solución basada en la traducción automática de textos, confirmando la utilidad del enfoque propuesto. Mostrando además el beneficio de los diccionarios en la etapa de re-etiquetado y su influencia en el mejoramiento de las etiquetas de los datos.

Abstract

Depression is a mental disorder with strong social and economic implications. The symptoms of this disorder are closely related to the conduct and the way of expressing of the people who suffer it. In recent years, social media had become in a popular media where people share their feelings. Recently several works have explored the analysis of social media content to identify and track depressed users following a supervised learning strategy supported on the availability of labeled training data. Unfortunately, acquiring such data is very complex and costly. To handle this problem and based on the idea that despite of their cultural diversity, people with depression tend to share similar information and to express in an analogous way, this investigation proposes a crosslingual approach based on the idea that data already labeled in a specific language can be leveraged to classify depression in other languages. The proposed method is based on a word-level alignment process. Particularly, we propose two representations for the alignment; whose capture correspondences between languages. For evaluating the proposed approach, we faced the detection of depression by employing English and Spanish tweets as the source and target data respectively. After a first attempt at classification, a second stage of re-labeling is proposed. It uses the labels thrown by the first classifier, next submits them to a refining process based on word dictionaries to generate a new training set and subsequently build a new classifier. The results outperformed solutions based on automatic translation of texts, confirming the usefulness of the proposed approach. Also it shows the usefulness of the dictionaries in the re-labeling stage and how this

can influence to improve the label of the data.

Capítulo 1

Introducción

1.1. Problemas de salud mental

La salud mental es un componente fundamental para la capacidad de cada individuo de pensar, interactuar con la sociedad, ganar un sustento económico y de manera general, tener una buena calidad de vida. La OMS [Asamblea Mundial, 2013] define la salud mental como un estado de completo bienestar físico, mental y social, por lo que su protección es un tema que involucra tanto a personas, comunidades como a las sociedades de todo el mundo. Muchos son los factores que se asocian a una mala salud mental, entre ellos: entornos de trabajos estresantes, exclusión social, indicadores de pobreza y riesgo de violencia, aunque también tiene asociados factores biológicos como es el caso de la genética o desequilibrios bioquímicos cerebrales [Yu, 2018]. Entre los trastornos mentales más frecuentes con influencia directa en la calidad de la salud mental de las personas, encontramos la depresión, el trastorno afectivo bipolar, la esquizofrenia y otras psicosis. Todos ellos aumentando cada año así como su consecuencia social y económica. Anualmente, estos trastornos tienen costos millonarios [Yu et al., 2016] y las pérdidas se calculan en el período de 2011-2030 en 16 mil millones de dólares [Saxena et al., 2013]. El ámbito laboral también

tiene notables impactos debido a las ausencias y por transitividad gran disminución en el rendimiento de las personas.

Específicamente, la depresión es un trastorno mental frecuente donde las personas que lo padecen presentan signos de desinterés, deterioro de la memoria, dificultad de pensamiento, disminución de la capacidad de concentración e insomnio o hipersomnio [Spitzer et al., 2013]. Estos síntomas causan un malestar clínico y un deterioro significativo en el ámbito social, laboral y otras esferas de la vida cotidiana, que en casos severos puede conllevar al suicidio; según datos de la OMS, 800,000 personas se suicidan cada año y es la segunda causa de muerte en el grupo etario de 15 a 29 años. Estos datos también reflejan que este trastorno afecta cerca de 300 millones de personas en todo el mundo, representando así un 4,3 de la carga mundial de morbilidad [Organization, 2001] donde más de la mitad no son tratados ya sea por falta de recursos o de personal sanitario calificado. Según la OMS en países de bajos recursos se estima que entre un 76 % y 85 % de las personas no son atendidas y que sólo el 36 % de personas que viven en estos países están amparados por una legislación en materia de salud mental. Ante esta situación la OMS emitió un plan de acción para el período 2013-2020 donde, entre sus objetivos cita poner en práctica estrategias de promoción y prevención en el campo de la salud mental.

La detección clínica de la depresión realizada por psicólogos es basada en herramientas que consisten en un grupo de preguntas cuyas respuestas van enfocadas a monitorear el comportamiento de los pacientes. Los síntomas que presentan las personas deprimidas están dados por la unión de varios factores respecto al comportamiento [Spitzer et al., 2013], entre ellos se pueden mencionar las palabras o frases que usan y la forma en la que lo hacen. Las personas que padecen de depresión, en la mayoría de los casos, son identificadas por observaciones propias de cambios en su conducta. Sin embargo, una parte de la población como es el caso de niños, adolescentes y ancianos presentan dificultad para evaluarse con los métodos ya existentes [Savard, 2004].

Recientemente, las redes sociales han sido un medio a través del cual personas deprimidas comparten sus ideas, emociones y estados de ánimos. Esto ha generado una gran cantidad de información que, con el acceso cada vez más extendido a las nuevas tecnologías, ha sido posible su recopilación y por ende la observación del comportamiento de los usuarios en diversas plataformas. Diferentes investigaciones haciendo uso de técnicas de clasificación de textos, específicamente guiados por un enfoque de clasificación supervisada han estudiado la detección de este trastorno mental.

Los resultados hasta el momento reportados por estos estudios son congruentes con la teoría clínica existente. Por ejemplo [Tsugawa et al., 2013, De Choudhury et al., 2013] revelan que las personas deprimidas tienden a expresar mayormente sentimientos negativos y que sus textos muestran un aumento significativo en el uso de pronombres, específicamente, en primera persona. También [Reece et al., 2017] concluyen que es notable el uso de palabras referidas tanto a los síntomas como al tratamiento y [Nguyen et al., 2014] evidencian, que a pesar de los diferentes estados emocionales que presentan las personas deprimidas, los cuales varían a lo largo del día, tienden a publicar sus emociones incluso cuando presentan signos de mal humor. Lo anterior demuestra que es posible a partir de un análisis de las palabras y textos que escriben las personas deprimidas distinguirlas de las personas sanas tomando ventaja de la notable diferencia respecto al lenguaje que existe entre personas sanas y mentalmente enfermas.

1.2. Problemática

En los últimos años la detección de depresión ha sido tratada como un problema de clasificación supervisada de textos utilizando grandes cantidades de datos para este proceso. Muchos de ellos han sido proporcionados por las redes sociales, por ejemplo de Twitter. Varias investigaciones haciendo uso de estos datos revelan

modelos que permiten detectar personas con depresión a partir del contenido de los textos que escriben. Todos ellos limitados al idioma de los datos que se analizan (en su mayoría inglés y uno en japonés [Tsugawa et al., 2015]), provocando que no puedan ser aplicados a otros idiomas a pesar que comparten características a la hora de realizar el análisis.

La recolección y el etiquetado de datos es un proceso costoso y complejo que requiere tiempo y recursos para lograr un proceso exitoso, sobre todo cuando se trata de grandes cantidades de datos. Para la obtención de los datos en estudios relacionados con depresión, se han usado dos enfoques básicos. El primero consiste en aplicar los cuestionarios para detectar depresión a un grupo determinado de personas, y para las que su diagnóstico sea positivo, bajo su consentimiento, se adquiere su historial de Twitter. El segundo consiste en localizar usuarios que en alguna de sus publicaciones hayan declarado haber sido diagnosticados clínicamente con depresión y de igual manera se obtienen sus historiales de Twitter. Una desventaja del primer enfoque es que a pesar de que hay una mayor certeza que la persona pueda padecer la enfermedad, se necesita un gran número de recursos, ya sea desde los pacientes que deseen participar hasta el personal médico que los evalúe.

Ante esta situación, lo más recomendable sería utilizar los datos ya etiquetados y aprovechar el conocimiento que se puede obtener de ellos. A pesar de que se han realizado un gran número de trabajos en el área, aún no se reporta alguna solución que aproveche el conocimiento adquirido en un idioma para ser utilizado en idiomas con pocos o ningún recurso orientado a esta tarea. A este tipo de problemas donde se utilizan recursos de un idioma fuente para luego asignar clases a documentos en un idioma objetivo, se le conoce como clasificación translingüe. La clasificación translingüe no ha sido objetivo en las investigaciones que involucran la tarea de detección de depresión a partir del texto, sin embargo ha sido aplicada exitosamente en otras tareas como es el caso de análisis de sentimientos [Abdalla and Hirst, 2017, Al-Shabi et al., 2017].

El principal desafío de la clasificación translingüe es la brecha que existe entre los idiomas fuente y objetivo. Una estrategia simple de la clasificación translingüe sería hacer la traducción usando traductores automáticos para transformar los datos y llevarlos al mismo idioma. La desventaja aquí es que en las redes sociales específicamente en Twitter por la brevedad de los textos que se escriben, no existe suficiente contexto para realizar una correcta traducción; abundan las faltas de ortografía y errores gramaticales así como el uso excesivo de abreviaturas convencionales y no convencionales. De manera general prevalece un lenguaje coloquial. A continuación se proponen los objetivos de la presente investigación basados en la problemática anteriormente expuesta.

1.3. Objetivos

1.3.1. Objetivo general

Proponer un enfoque translingüe para la detección de depresión en textos generados por usuarios de Twitter para el idioma Español, a partir de instancias etiquetadas en idioma Inglés, que mejore los resultados obtenidos por una traducción automática como mecanismo de alineación entre idiomas.

1.3.2. Objetivos específicos

- Analizar diferentes características en el lenguaje inglés para identificar rasgos relevantes en la tarea de detección de depresión.
- Proponer un método de alineación basado en recursos psicolingüísticos.
- Proponer una representación basada en la alineación a partir de representaciones distribuidas de las palabras.

- Proponer un método para el refinamiento del etiquetado usando diccionarios específicos del dominio de depresión.

- Analizar los resultados alcanzados usando como referencia un método basado en traducción automática.

1.4. Contribuciones

La presente investigación propone un nuevo enfoque para la detección de depresión en usuarios de Twitter basada en clasificación translingüe, el cual asigna clases a documentos escritos en español usando datos de entrenamiento en un idioma diferente (Inglés). A continuación se detallan las principales contribuciones del presente trabajo:

- Se presenta como primera investigación en el estudio de la depresión desde una perspectiva translingüe a partir de textos.

- Se creó un corpus en Español que recoge los historiales de usuarios de Twitter los cuales representan a dos grupos de usuarios: depresivos y no depresivos.

- Se propone una representación basada en una alineación a nivel de palabra usando word embeddings.

- Se propone un método para el baseline basado en una traducción automática, el cual es usado para evaluar los resultados del enfoque propuesto.

- Se propone un método que permite, a través de un proceso de depuración, mejorar las etiquetas de clase asignadas a los documentos escritos en español.

1.5. Estructura de la tesis

El resto del documento se estructura de la siguiente forma. El capítulo 2 aborda los conceptos básicos que serán necesarios para describir el método propuesto. El capítulo 3 refleja los diferentes trabajos de investigación que se han reportado relacionados con la temática de la tesis. Por un lado se encuentran los que tratan el problema de detección de depresión como un problema de clasificación de textos. Por otro lado los que tratan la clasificación de textos desde una perspectiva translingüe. El capítulo 4 describe y formaliza el método propuesto y en el capítulo 5 se exponen los experimentos derivados del método, así como su configuración. Se exponen además los resultados y una discusión de los mismos. Finalmente, son presentadas las conclusiones del trabajo.

Capítulo 2

Marco Teórico

En el presente capítulo se hace un acercamiento a los principales conceptos relacionados con la clasificación de textos monolingüe y translingüe usados durante el desarrollo de esta investigación; así como conceptos relacionados con *word embeddings* y los algoritmos para su obtención. Se describen además los clasificadores usados en la fase de experimentos.

2.1. Clasificación de textos

Con el aumento progresivo de la información que se genera en el mundo, la búsqueda de información específica se vuelve más rigurosa. Para darle solución a esto se han creado mecanismos donde la información es agrupada por temas comunes. Cuando nos referimos a documentos, la tarea consiste en agrupar documentos por clases; una clase describe no sólo el tema sino otras características que permitan identificarlo [Baeza-Yates et al., 1999].

La *Clasificación de textos* se define como el proceso de asignar clases a un documento [Sebastiani, 2002]. Esta, ofrece una mayor organización permitiendo así

un mejor entendimiento e interpretación de los datos. Para realizar la clasificación de textos han sido desarrollados varios algoritmos que permiten de manera automática, dado un conjunto de ejemplos creados por especialistas, predecir la clase de nuevos documentos que no han sido vistos con anterioridad.

2.1.1. Representaciones de los textos

Por lo general, para que un clasificador pueda realizar sus funciones lo que recibe como entrada es cada documento representado por un vector de características. A continuación se describen algunas de las representaciones comúnmente usadas para la clasificación de textos.

Bolsa de palabras

La bolsa de palabras (BoW) es una de las representaciones más usadas en el análisis automático de textos. Los documentos son representados por un conjunto de palabras sin importar el orden en el que aparezcan. Dado un vocabulario V, que es el conjunto de cada palabra w_i de todos los documentos, y un conjunto de documentos $D = \{d_1, d_2, ..., d_n\}$. Cada documento dj puede ser representado por un vector v cuya dimensión sería $|V|$. Cada elemento del vector v tiene al peso de W_i en dj (ver Ejemplo 2.1).

Ejemplo 2.1 Se tiene el conjunto de documentos D=$\{d_1$: *Las decepciones cierran el corazón, d_2: El silencio duele en el corazón* $\}$ de donde se obtiene el vocabulario V=$\{$*decepciones, cierran, corazón, silencio, duele* $\}$. La tabla 2.1 muestra la representación de estos documentos mediante BOW.

Tabla 2.1: Ejemplo de una representación con BoW

	decepciones	cierran	corazón	silencio	duele
d_1	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}
d_2	W_{21}	W_{22}	W_{23}	W_{24}	W_{25}

De manera que W_{ij} representa el peso (los tipos de pesado se explican en la sección 2.1.2) de la palabra w_i en el documento d_j. Si la palabra no se encuentra en el documento entonces $W_{ij} = 0$.

N-gramas de palabras

El uso de los $n - gramas$ captura la secuencia de las palabras. Estos son secuencias de n palabras consecutivas, donde n determina el tamaño de la secuencia, mientras mayor sea el valor de n más específica será la representación que se obtenga de él. Tomando el ejemplo anterior, se pueden extraer los siguientes n-gramas:

Tabla 2.2: Ejemplo de n-gramas de palabras con n=1, n=2 ,n=3

n=1	*decepciones*,*cierran*,*corazón*,*silencio*,*duele*
n=2	*decepciones-cierran*,*cierran-corazón*,*silencio-duele*,*duele-corazón*
n=3	*decepciones-cierran-corazón*,*cierran-corazón-silencio*,*corazón-silencio-duele*

La manera en que un documento puede ser representado por los $n - gramas$ es similar al explicado en la representación anterior. Considerando el conjunto de documentos del ejemplo anterior, la tabla 2.3 muestra el ejemplo de la representación para los bi-gramas (n=2). Donde W_{ij} es el peso del bi-grama j en el documento i.

Tabla 2.3: Ejemplo de una representación con n-gramas de palabras.

	decepciones-cierran	cierran-corazón	silencio-duele	duele-corazón
d_1	W_{11}	W_{12}	W_{13}	W_{14}
d_2	W_{21}	W_{22}	W_{23}	W_{24}

Representación basada en LIWC

Linguistic Inquiry and Word Count (LIWC) es un recurso lingüístico diseñado por investigadores de la rama de la psicología que permite el análisis de componentes emocionales, cognitivos y estructurales presentes en los textos [Pennebaker et al., 2007]. Este recurso tiene disponible diccionarios en varios idiomas. Cada diccionario está formado por un grupo de categorías que a su vez contienen palabras en dependencia de lo que reflejen, incluye además categorías sintácticas. Algunas categorías con sus respectivas palabras presentes en el recurso son: *Hogar*{*jardín, patio, piscina* }, *Familia* {*fomenta, hermano, individuo* }, *Emociones positivas*{*acaricia, afable, atractivo*} y *Emociones negativas*{*avaricia, bajeza, calumnia*}, entre otras. Cada palabra puede estar asociada a más de una categoría. Una representación podría lograrse para un documento a partir de las categorías de LIWC y tener en cuenta las palabras que están dentro de ellas y que también pertenecen al documento.

2.1.2. Pesado de los atributos

Para construir el vector que represente a un documento, cada elemento del espacio de características aporta un peso al mismo. No todos los términos en un documento son útiles para describirlo, por lo que caracterizar el término a partir de su peso en el documento ayuda a resaltar su importancia [Baeza-Yates et al., 1999]. Ese peso es obtenido de diferentes maneras y depende del problema o la situación que se quiera analizar. A continuación se describen tres de los más usados.

Booleano

La forma de pesado más simple es la booleana o binaria; donde el peso correspondiente a una palabra en un documento es 1 si la palabra está presente en el documento y 0 de lo contrario. Este pesado aporta información sobre la relación que existe entre un documento y una clase de acuerdo a los atributos que contiene.

Frecuencia del término en el documento

Este pesado denotado como tf depende de la frecuencia de ocurrencia de la palabra en el documento. Anteriores estudios han demostrado que el uso de este pesado es útil para encontrar palabras que están altamente relacionadas con el dominio [Nadeem, 2016]. En la ecuación 2.1 se resume:

$$tf_{ij} = f_{ij} \tag{2.1}$$

donde la frecuencia f_{ij} del término w_i en el documento d_j es la cantidad de veces que el término ocurre en el documento. Si la frecuencia f_{ij} es normalizada por la longitud del documento la ecuación 2.1 es transformada a:

$$tf_{ij} = \frac{f_{ij}}{|d_j|} \tag{2.2}$$

donde $|d_j|$ es la longitud del documento con respecto al número de términos.

TF-IDF

Otra forma muy usada de calcular el peso correspondiente a una palabra en un documento es la combinación de la frecuencia del término y la frecuencia inversa del documento en la colección de documentos. Este tipo de pesado es beneficioso si se desea diferenciar los textos por temas o contenidos. La ecuación 2.3 lo formaliza

$$tf - idf_{ij} = (1 + log f_{ij}) \times log(N/n_i) \tag{2.3}$$

Donde f_{ij} es la frecuencia del término en el documento, N es el número de documentos en la colección y n_i la cantidad de documentos en los que aparece el término.

2.2. Algoritmos de clasificación

Varios algoritmos se han propuesto y han sido utilizados para entrenar un clasificador cuando ya se ha logrado una representación. En esta sección se explican los algoritmos *Naïve Bayes* y *Support Vector Machine* pues serán los utilizados en los experimentos que se explicarán en capítulos posteriores.

2.2.1. Clasificador Naïve Bayes

El clasificador Naïve Bayes es un algoritmo probabilístico que asume que todos los atributos son independientes dado el atributo clase. Sea $A_d = \{a_1, a_2, .., a_d\}$ el conjunto de atributos y $C_n = \{c_1, c_2, .., c_n\}$ el conjunto de clases, la tarea del clasificador es predecir la clase más probable de la siguiente manera:

$$arg_C[max[P(C_i|a_1, a_2, .., a_d)]] \tag{2.4}$$

Este clasificador está basado en la regla de Bayes:

$$P(C_i|a_1, a_2, .., a_d) = \frac{P(C_i) * P(a_1, a_2, ..., a_n|C_i)}{P(a_1, a_2, ..., a_n)} \tag{2.5}$$

Como el clasificador asume que todos los atributos son independientes dada la clase, la ecuación anterior quedaría como sigue:

$$P(C_i|a_1, a_2, .., a_d) = \frac{P(C_i) * P(a_1|C_i) * P(a_2|C_i) * .. * P(a_d|C_i)}{P(a_1, a_2, .., a_d)} \tag{2.6}$$

La probabilidad de la clase $P(C_i)$ se calcula:

$$P(C_i) = \frac{N_i}{N} \tag{2.7}$$

donde N_i es la cantidad de instancias (documentos) de la clase C_i y N el total de instancias de todo el corpus. La probabilidad de cada atributo $P(a_j|C_i)$ se calcula como sigue:

$$P(a_j|C_i) = \frac{N_{aj}}{N_i} \tag{2.8}$$

donde N_{aj} es el número de veces que ocurre el atributo a_j en la clase C_i y N_i es el número de instancias de la clase C_i. Si N_{aj} no se encuentra en ningún documento de la clase $P(a_j|C_i)$=0.

2.2.2. Clasificador Máquina de Soporte Vectorial

Una máquina de vectores de soporte (*Support vector machine*) es un algoritmo de clasificación supervisada basado en la teoría de aprendizaje estadístico que puede ser utilizado para clasificación binaria, clasificación multiclase y regresión. La idea principal es encontrar un hiperplano que equidiste de los ejemplos más cercanos de cada clase para lograr de esta forma un *margen máximo* a cada lado del hiperplano (ver figura 2.1). Para elegir el hiperplano se escogen sólo los ejemplos de entrenamiento que están en la frontera de este margen, a los que se nombra *vectores de soporte*.

Suponiendo que tenemos un conjunto S de puntos etiquetados para el entrenamiento (x_i, y_i) donde $x_i \in R^N$ son las instancias etiquetadas y $y_i \in \{-1, 1\}$ que son las dos clases posibles. Por lo general, las instancias son representadas en un espacio de características de mayor dimensión Z y el vector correspondiente en el espacio de características sería $z = \delta(x)$. Se desea encontrar el hiperplano de modo que:

$$w.z + b = 0 \tag{2.9}$$

Definido por el par (w, b), donde $w \in Z$ y $b \in R$.

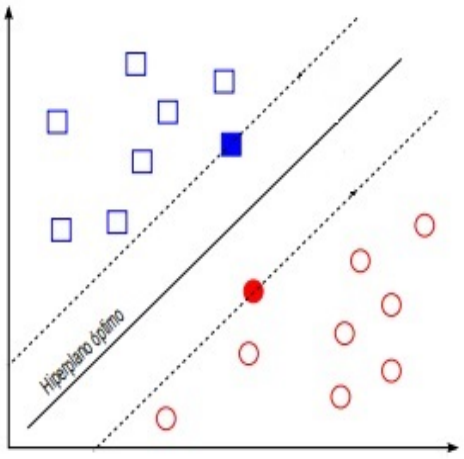

Figura 2.1: Hiperplano de separación óptimo y su margen asociado

2.3. Evaluación del desempeño

Para evaluar el rendimiento de un clasificador se han propuesto el uso de varias métricas de evaluación como es el caso de la exactitud, el recuerdo, la precisión y la medida F.

2.3.1. Precisión

La precisión se formula calculando la fracción de todos los documentos asignados correctamente a una clase por el clasificador con todos los asignados a la clase.

$$P_{Ci} = n_a/n_t \tag{2.10}$$

donde n_a es la cantidad de predicciones correctas para la clase C_i y n_t el total de predicciones para la clase C_i.

2.3.2. Recuerdo

El recuerdo R_{Ci} es definido como la fracción entre los documentos que fueron asignados a la clase por el clasificador y el total de documentos pertenecientes a la clase.

$$R_{Ci} = \frac{n_{ar}}{n_{tr}} \tag{2.11}$$

donde n_{ar} es el número de documentos que el clasificador asignó a una clase y n_{tr} el número total de instancias que pertenecían a esa clase

2.3.3. Medida F1

La medida F es una métrica de evaluación que combina la precisión y el recuerdo resaltando la importancia relativa de ambos para cada clase. La ecuación 2.12 muestra la forma en la que se calcula:

$$F_1(C_i) = \frac{2P_{Ci} * R_{Ci}}{P_{Ci} + R_{Ci}} \tag{2.12}$$

$F_1(C_i)$ es el valor de la medida F, P es la precisión, R el recuerdo, C_i la clase en cuestión.

2.4. Word embeddings

Word embeddings surge como una técnica para modelar el lenguaje que transforma el vocabulario de un corpus en un vector de representación para cada palabra en una baja dimensión [Wohlgenannt et al., 2016]. Usa una red neuronal artificial para generar el vector de representación. No solo permite una representación útil de las palabras sino que es muy eficiente para entrenar grandes corpus con grandes contextos [Rumelhart and McClelland, 1986]. Los diferentes tipos de contextos producen diferentes similitudes entre palabras. Ha sido exitoso en operaciones para

encontrar analogías entre palabras [Mikolov et al., 2013a] y explorar los contextos que son obtenidos en el proceso de aprendizaje. De manera que palabras relacionadas semánticamente tienen vectores cercanos en un espacio de representación. La figura 2.2 muestra un ejemplo de la representación en un espacio de dos dimensiones de varias palabras en idiomas diferentes y que tienen una relación semántica.

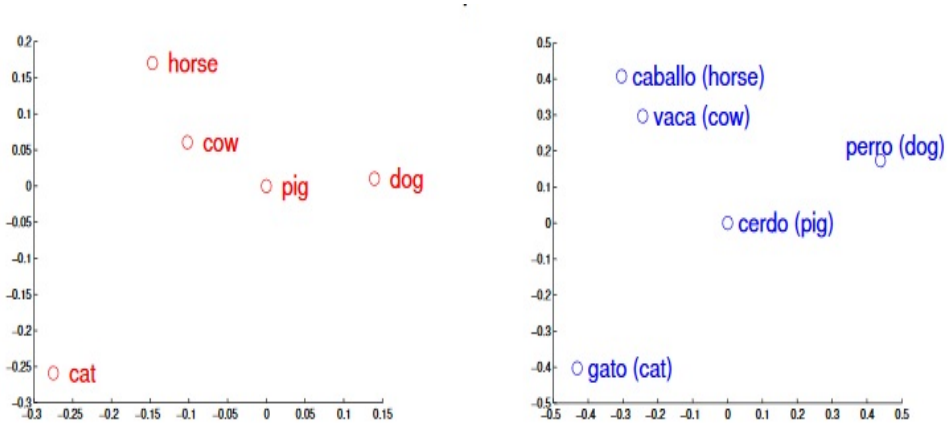

Figura 2.2: Representación vectorial en un espacio de 2 dimensiones de palabras en inglés y su correspondiente palabra en español.Fuente: [Mikolov et al., 2013b]

2.4.1. Algoritmos para su obtención

Word2Vec

Word2Vec es una herramienta creada por [Mikolov et al., 2013c]. Word2Vec aplica una red neuronal entrenada para el contexto lingüístico de las palabras o frases. Su entrada es básicamente un corpus largo, y las salidas son los embeddings que son vectores de representación para las palabras del corpus. Por lo general, la dimensionalidad del vector varía de 50 a 300. Cuando dos vectores están cercanos en el espacio significa que los contextos de sus respectivas palabras son similares. Existen dos modelos de arquitectura para que el algoritmo pueda crear los vectores. Una de ellas es CBOW (Bolsa de palabras continuas) y skip-gramas continuos. Con

CBOW el modelo predice la palabra actual pero usando una ventana de palabras que están alrededor. Con skip-grama el modelo predice la ventana de palabras alrededor usando la palabra actual.

Glove

Similar a Word2Vec, Glove [Pennington et al., 2014] es un algoritmo para obtener representaciones de palabras a través de vectores. Este no es un modelo predictivo sino más bien un modelo basado en conteo, por lo que contabiliza la frecuencia con la que coexisten las palabras en un corpus dado, usando una estadísticas de co-ocurrencia de palabra a palabra . El objetivo del entrenamiento en GloVe es aprender vectores de palabras de modo que su producto punto sea igual al logaritmo de la probabilidad de co-ocurrencia de las palabras. La intuición principal detrás del modelo es la simple observación de que las proporciones de probabilidades de coincidencia palabra-palabra tienen el potencial de codificar alguna forma de significado.

2.4.2. Medidas de similitud entre embeddings

Para cuantificar el grado de asociación entre dos vectores se han propuesto varias medidas de similitud o el cálculo de distancias. Cuando se trabaja con textos y específicamente con vectores de palabras muchos trabajos han recomendado el uso de similitud del coseno. A continuación se formaliza.

Similitud coseno

La similitud coseno calcula el coseno del ángulo entre los vectores, de manera que mientras más cercano a 1 esté el resultado, más asociación hay entre los vectores

y por ende las palabras que lo representan. A continuación se formaliza:

$$cos(v_1, v_2) = \frac{v_1 * v_2}{||v_1|| \times ||v_2||} = \frac{\sum_{i=1}^{d} v_{1i} * v_{2i}}{\sqrt{\sum_{i=1}^{d} {v_{1i}}^2} * \sqrt{\sum_{i=1}^{d} {v_{2i}}^2}} \tag{2.13}$$

donde v_1 y v_2 corresponde a los vectores, v_{1i} y v_{2i} corresponde al $i-ésimo$ elemento de los vectores v_1 y v_2 respectivamente.

2.5. Clasificación de textos monolingüe

La clasificación automática de textos es un tipo de aprendizaje que puede estar guiada a una perspectiva monolingüe, donde los documentos involucrados en el proceso pertenecen al mismo idioma. Puede formalizarse como sigue: dado un conjunto de documentos D y un conjunto de clases C, un clasificador de textos sería una función $f : D \times C$ de modo que a cada documento dj, tal que $dj \in D$, es asignada una clase cp donde $cp \in C$, entonces tendríamos el par (dj, cp).

La clasificación automática de textos es realizada de dos formas: supervisada y no supervisada. En el caso de la supervisada, el objetivo es aprender, a partir de un conjunto de datos de entrenamiento, un mapeo instancias (documentos) - etiquetas (clases). Los datos de entrenamiento están formados por un conjunto de documentos donde se especifica la clase a la que pertenecen de acuerdo a un especialista. Estos datos son usados para que el clasificador aprenda y de acuerdo a ese conocimiento pueda entonces predecir las clases de otros documentos que son proporcionados y que no han sido vistos antes por el clasificador. En el caso de la no supervisada, no se requieren datos etiquetados manualmente para su funcionamiento. Para asignar las clases a las instancias, se necesita criterios que ayuden a descubrir patrones para determinar cuando estamos en presencia de una clase o de otra.

2.6. Clasificación de textos translingüe

La clasificación de textos translingüe constituye una oportunidad en la que se puede aprovechar la información de datos previamente etiquetados. Conceptualmente, la clasificación translingüe es la tarea de asignar clases a documentos escritos en un idioma destino a partir de recursos de un idioma de origen [Gliozzo and Strapparava, 2006]. De esta manera, el clasificador es entrenado con datos del idioma fuente para luego probar en la clasificación de documentos que están en un idioma objetivo. De manera que $D_s = \{d_1^s, d_2^s, ..., d_n^s\}$ es el conjunto de documentos del idioma fuente que es utilizado para el entrenamiento y $D_t = \{d_1^t, d_2^t, ...d_m^t\}$ es el conjunto de documentos que están en el idioma objetivo que es utilizado para la prueba. La idea es entrenar un clasificador Φ_s en D_s para asignar clases a cada elemento de D_t. La figura 2.3 resume los pasos generales de este tipo de clasificación.

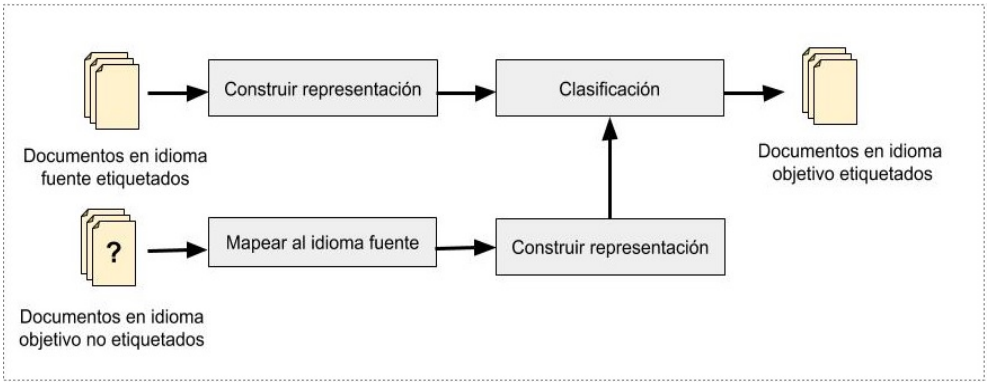

Figura 2.3: Representación del enfoque general de la clasificación translingüe.

Enfoques de la clasificación translingüe

Un estrategia básica para la clasificación translingüe es el uso de traductores automáticos para transformar los datos y mapearlos al mismo idioma. Esta estrategia está soportada por dos enfoques, el primero consiste en traducir el conjunto de prueba D_t y llevarlo al idioma fuente, de manera que el clasificador que está entrenado en

el idioma fuente sea aplicado en los documentos traducidos (ver figura 2.4). En el segundo enfoque el conjunto de entrenamiento D_s es traducido y llevado al idioma objetivo, de modo que el clasificador es entrenado en el idioma objetivo (ver figura 2.5)

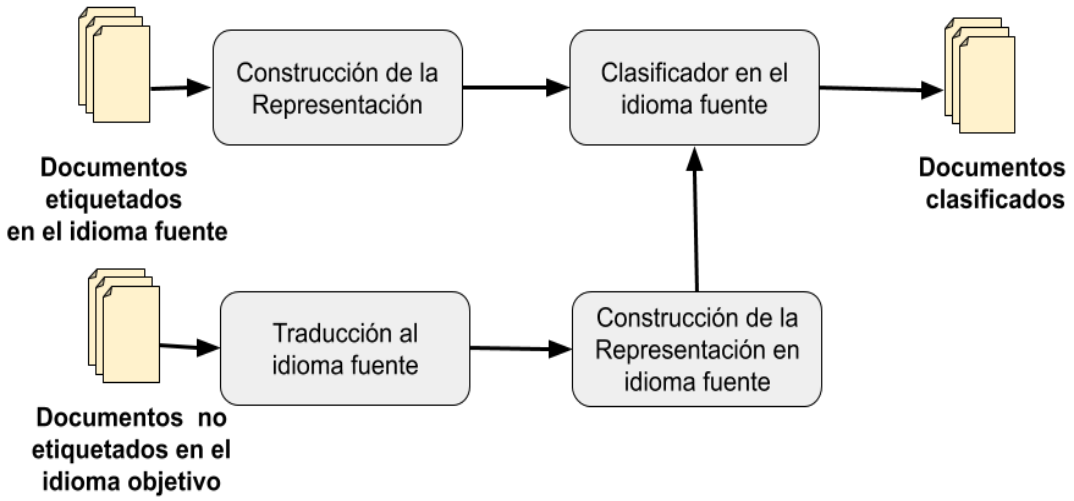

Figura 2.4: Esquema de clasificación translingüe con traducción de documentos del idioma objetivo al idioma fuente

De manera general ambos enfoques tienen sus desventajas, pues el éxito en la traducción depende mucho del origen de los datos, fundamentalmente los errores que pueden contener y si tienen suficiente contexto para hacer una correcta traducción. Este proceso es costoso con respecto al tiempo cuando se tienen muchos textos. Lo más útil sería traducir el conjunto, ya sea fuente u objetivo, que contenga menos datos.

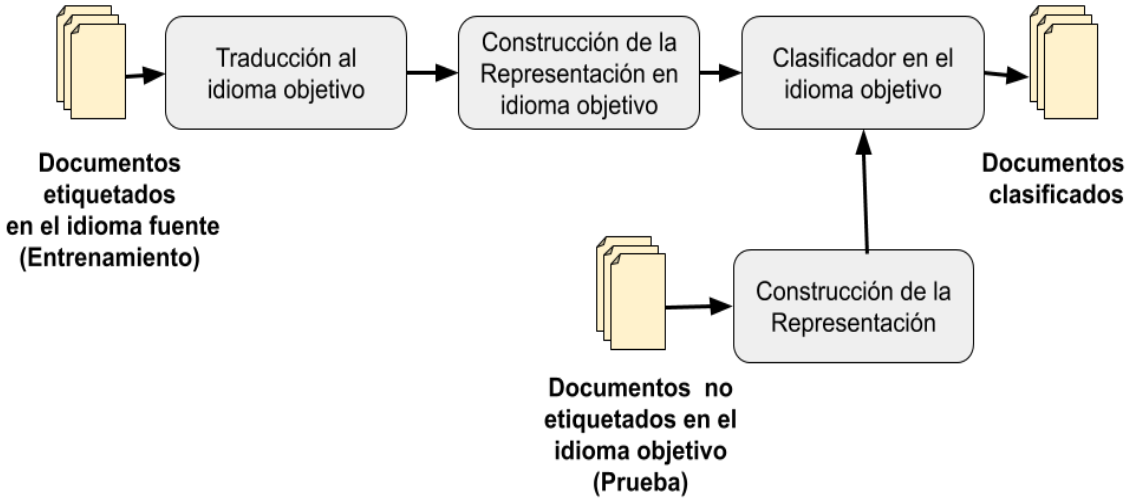

Figura 2.5: Esquema de clasificación translingüe con traducción de documentos del idioma fuente al idioma objetivo

Capítulo 3

Trabajo Relacionado

En el presente capítulo se abordan los trabajos relacionados con la detección de depresión que han sido tratados como un problema de clasificación de textos, todos ellos guiados desde una perspectiva monolingüe. También se exponen trabajos que han utilizado como solución un enfoque de clasificación translingüe en la tarea de análisis de emociones.

3.1. Detección de depresión desde la perspectiva de clasificación de textos

La depresión es una enfermedad mental que afecta el equilibrio emocional de las personas. Su detección está dada fundamentalmente por diferentes patrones de comportamiento de los individuos que la padecen [Spitzer et al., 2013]. En los últimos años varias investigaciones han sido guiadas a la detección de este trastorno mental a través del aprendizaje automático, analizándolo como un problema de clasificación de textos.

3.1.1. Enfoques para la construcción de corpus

El proceso de clasificación de textos consiste en extraer varias características de un conjunto de datos previamente etiquetados y a partir de estos aprender modelos que permiten distinguir entre varias clases. La adquisición de los datos es un paso importante para la clasificación; sin embargo, es un proceso complejo, costoso y que consume tiempo. Específicamente en la tarea de detección de depresión las redes sociales han sido medios que han permitido esta adquisición [Guntuku et al., 2017]. A continuación se describen los diferentes enfoques que se han utilizado para la construcción de las colecciones de datos utilizados en esta tarea.

Enfoque basado en cuestionarios

La ayuda de profesionales de la psicología y sus criterios respecto a los síntomas han sido de gran utilidad para la construcción inicial de los datos [Coppersmith et al., 2015]. La detección hecha por psicólogos es realizada a través de cuestionarios. Los cuestionarios para la detección de enfermedades mentales tiene una gran validez y credibilidad entre los profesionales de la psicología; considerándose el segundo mecanismo más efectivo seguido de las entrevistas clínicas [Guntuku et al., 2017]. Éstos consisten en un conjunto de preguntas relacionadas con el comportamiento diario de las personas; cada pregunta tiene como respuesta varias opciones de las cuales el paciente debe seleccionar una. Concluido el cuestionario se cuantifican las respuestas y teniendo en cuenta ciertos umbrales se decide si la persona padece o no la enfermedad. Varias investigaciones apoyadas en esta forma de detección han adquirido los ejemplos que representan a la clase deprimida [De Choudhury et al., 2013, Tsugawa et al., 2013, Tsugawa et al., 2015, Reece et al., 2017]. Una vez definida la presencia de la enfermedad en ciertos usuarios obtienen sus historiales en Twitter. Este enfoque tiene como desventaja la gran cantidad de recursos y tiempo necesarios para lograr el objetivo final.

Enfoque basado en la auto declaración del diagnóstico clínico de la enfermedad.

Muchos usuarios revelan su cotidianidad en las redes sociales [Shen et al., 2017], algunos en busca de ayuda exponen el diagnóstico de sus enfermedades, lo que ha sido de gran utilidad para los investigadores. Uno de los enfoques más utilizados para definir la presencia o no de la depresión en un usuario es el hecho de que ellos mismos auto declaren haber sido diagnosticados clínicamente con la enfermedad. Varios trabajos reportan haber utilizado expresiones específicas como *I'm diagnosed with depression*, *I have been diagnosed with depression*, *I was diagnosed with depression* [Coppersmith et al., 2014,Shen et al., 2017,Nadeem, 2016]. Una vez localizados estos usuarios es descargado el historial de sus publicaciones. Usando este mismo enfoque [Losada and Crestani, 2016] construyeron un corpus obteniendo la información de redes sociales como Reddit el cual fue posteriormente utilizado en e-Risk para la detección temprana de este trastorno. Este enfoque aunque está sujeto a incluir usuarios que realmente no padezcan la enfermedad, es una forma más manejable para los investigadores.

Enfoque basado en opiniones de foros en la web

Las redes sociales se extienden también a los sitios donde se crean comunidades de usuarios que comparten intereses y se realizan foros debates acerca de algún tema específico. En este caso Reddit se destaca por la variedad de temas que en él se abordan, incluyendo la depresión. Varios conjuntos de datos han sido creados de las publicaciones de sus miembros; en este caso podemos mencionar a [Losada and Crestani, 2016] que construyen un corpus formado por dos grupos depresivos y no depresivos. Para adquirir las publicaciones de los usuarios pertenecientes al grupo depresivo usaron la expresión de auto-declaración del diagnóstico clínico de la enfermedad. Para el grupo de los no deprimidos se buscaron usuarios que estuvieran activos dentro del subgrupo de depresión pero sin haber mencionado en sus

publicaciones algún padecimiento de la enfermedad.

Enfoque basado en uso de palabras claves

Otra forma de recolectar datos para posteriormente usarlos en estudios exploratorios para detectar depresión, es el uso de palabras específicas del dominio que permiten definir si un usuario o una publicación pertenece a un grupo u otro. Un ejemplo de ello se muestra en [Cavazos-Rehg et al., 2016] donde obtuvieron sus datos de la plataforma Twitter, estableciendo como palabra clave *"depresión"* para obtener los tuits cuyo contenido estuviera relacionado con esta enfermedad mental, por lo que fueron descartados los tuits donde apareciera el término pero sin relación alguna con enfermedad mental.

3.1.2. Representaciones de textos para detección de depresión

Estudios recientes reportan modelos de clasificación donde se intenta detectar enfermedades mentales a partir del contenido creado por usuarios [Reece et al., 2017, Nadeem, 2016, Preoţiuc-Pietro et al., 2015]. El análisis automatizado de datos en redes sociales puede proporcionar elementos para la detección temprana [Guntuku et al., 2017] pues los usuarios involucrados revelan sus estados de ánimo, emociones y forma de vida de manera espontánea, de modo que el contenido de los datos recolectados describen no sólo la vida diaria de los usuarios sino también su estado mental [Shen et al., 2017]. Para capturar comportamientos específicos de personas con depresión es de gran utilidad la extracción y representación de varias características que favorecen la distinción entre personas sanas y enfermas. A continuación se detallan las diferentes características y representaciones usadas para la detección de depresión.

BOW o N-gramas de palabras

Las palabras presentes en los textos reflejan su contenido y los temas principales que en él se abordan. Varios estudios relacionados con depresión revelan que las palabras que usan las personas deprimidas permiten distinguirlos entre el resto de las personas. En este sentido [Tsugawa et al., 2013,Nadeem, 2016] realizan un análisis respecto a la frecuencia de uso de palabras en un tuit; obteniendo como resultado que la frecuencia de las palabras es de gran utilidad para mostrar el comportamiento de una persona enferma con depresión.

En [Pedersen, 2015] utilizan los n-gramas de palabras en la construcción de listas de decisión para predecir usuarios con depresión o desorden por estrés post-traumático (PTSD). Los n-gramas son cuantificados para cada grupo donde los que pertenecen al GD (grupo depresivo) reciben un peso positivo y al GC (grupo de control) reciben un peso negativo, de manera que el signo del n-grama determina a que clase pertenece. Los n-gramas pertenecientes a las listas de decisión son buscados en el historial del usuario y de acuerdo a su frecuencia y a la clase de procedencia determina la polaridad para ese usuario, positivo si es deprimido, negativo en caso contrario. Los autores demuestran con sus resultados que los n-gramas y la frecuencia con los que se utilizan ayudan considerablemente a la clasificación de depresión, obteniendo una precisión de 73 %.

Por otra parte, [Preoţiuc-Pietro et al., 2015] utilizan los n-gramas de palabras con un valor de $n = 3$ para el análisis del lenguaje de las personas deprimidas. Combinan además información personal como la edad y el género con los rasgos de la personalidad descritos en el modelo *Big-Five*, el cual en términos de psicología establece cinco dimensiones que describen la personalidad: extrovertido, amable, consciente, neurótico y abierto a experiencias. Ellos observan que el lenguaje es usado de manera diferente cuando cambia la edad, aunque verlo de forma aislada no aporta mucho a los resultados, sin embargo analizar edad y género de manera

conjunta puede ser tomado en cuenta cuando se detecta depresión. Respecto a los rasgos de la personalidad su trabajo muestra que las personas deprimidas tienden a presentar rasgos de tipo neurótico. Finalmente concluyen que existe una gran diferencia en las palabras que utilizan ambos grupos, deprimidos y no deprimidos.

A diferencia de trabajos anteriores, en [Stankevich et al., 2018] combinan el enfoque BOW con características estilo-métricas y morfológicas. Ésto les permitió alcanzar su mejor resultados. Utilizan además *word-embeddings* como característica para la detección de depresión; seleccionando las n palabras más informativas para cada usuario y obteniendo una representación para ellos usando el promedio de los vectores de esas palabras informativas; estos vectores fueron previamente multiplicados por la frecuencia de la palabra. Este último aportó resultados inferiores a los obtenidos con el enfoque BOW, lo que les permitió reafirmar la utilidad de las palabras y su frecuencia para alcanzar una buena predicción en la tarea de depresión.

Actividad en las redes sociales

Otras características utilizadas para la representación de los textos en la tarea de depresión, se extraen del monitoreo de los usuarios en las redes sociales respecto a las horas del día que dedican a publicar, los horarios más frecuentes, la cantidad de publicaciones y respuestas, y respecto al número de seguidores que tiene cada uno. En este sentido, [De Choudhury et al., 2013] para representar la interacción en las redes sociales lo hacen a través de un grafo, donde la conexión entre nodos representa el vínculo con otros usuarios. Esta conexión está dada ya sea por respuestas o por mensajes emitidos. Además, tienen en cuenta el número de publicaciones diarias de cada usuario. Cada una de estas características extraídas fueron cuantificadas diariamente para luego construir un vector usando media, varianza, desviación estándar y entropía. Concluyen con su estudio que las características que provienen del lenguaje son las que más aportan a la predicción de usuarios con depresión.

También [Tsugawa et al., 2015, Reece et al., 2017] cuantifican la frecuencia de actividad sobre Twitter respecto al número de publicaciones, RT (re-publicar un tuit), seguidores y seguidos por el usuario se refiere; además de la cantidad de horas del día que dedican a publicar. Estas características son vinculadas con tópicos que extraen usando LDA (Latent Diriechlet Allocation). Una observación importante es la utilidad de los tópicos como característica distintiva para usuarios depresivos alcanzado resultado superiores sobre el resto de las características extraídas.

Recursos psicolingüísticos

El uso de recursos psicolingüísticos ha sido factible en el análisis de emociones en textos. Las emociones son un factor que ha sido muy estudiado para la detección de depresión. Por ejemplo [De Choudhury et al., 2013, Reece et al., 2017] analizan las emociones usando *LIWC* y *ANEW* (Affective Norms for English Word): ambos son recursos muy populares en el estudio de las emociones, el primero consta de 64 categorías y cada una de ellas es representada por un conjunto de palabras, el segundo lexicón está formado por 1034 palabras calificadas en términos de valencia y excitación, cada una de estas cualidades se cuantifican en una escala de 0 a 9. En [Reece et al., 2017] observan el comportamiento de las emociones de los usuarios a través del tiempo con series temporales, modelando el problema con HMM (Modelos ocultos de Markov).

En este sentido [Shen et al., 2017] utilizan LIWC para evaluar la polaridad de las emociones (positiva o negativa). Proponen un enfoque multimodal que combina emociones, información personal, tópicos, palabras específicas del dominio y relacionadas con síntomas de la enfermedad e información de la imagen de perfil del usuario. Similarmente, [Coppersmith et al., 2014] analizan la polaridad de las palabras presentes en un tuit. Resaltan en su investigación que el lenguaje usado por usuarios depresivos aporta evidencia de las diferencias entre sanos y enfermos.

Este planteamiento los lleva a coincidir con [Tsugawa et al., 2015] y [De Choudhury et al., 2013] que existe un aumento significativo en el uso de palabras que responden a emociones negativas.

En [Gkotsis et al., 2016] extraen los datos de la plataforma Reddit para su estudio; analizando publicaciones y comentarios de manera separada. A través de un análisis de sentimientos se enfocan en la detección de felicidad a partir de las palabras que se usan. Observan que los usuarios que padecen de depresión expresan infelicidad y sentimientos negativos en sus publicaciones, coincidiendo así con las observaciones de trabajos anteriores.

Apoyados en la idea de la utilidad de los recursos léxicos, [Losada and Gamallo, 2018] usando el corpus creado por [Losada and Crestani, 2016] proponen varios modelos para representar los datos a través de lexicones y explotar las ventajas que ellos ofrecen. Para estos métodos principalmente sugieren el uso de lemma-POS, *word embeddings*. Evalúan cada modelo con varios recursos léxicos y obtuvieron que utilizar POS(partes de la oración) es útil para aprovechar los lexicones, haciendo una especial observación en los adjetivos, calificándolos como un indicador útil en las expresiones relacionadas con depresión.

De manera general las investigaciones coinciden en que los textos generados por usuarios que padecen de depresión muestran una existencia frecuente de emociones negativas y un constante uso de palabras referidas a síntomas y medicamentos relacionados con la enfermedad. Esto permite que sea un elemento distintivo en el análisis de textos para detectar depresión.

Palabras del dominio

Entre las palabras que se destacan por su frecuente uso en personas deprimidas se encuentran aquellas que se refieren tanto a síntomas como a medicamentos rela-

cionados con la enfermedad. Varias investigaciones [Shen et al., 2017, Reece et al., 2017] han tomado ventaja de esto para encontrar diferencias entre personas sanas y enfermas. Por ejemplo, en [Cavazos-Rehg et al., 2016] hacen un estudio exploratorio de los tuits donde de acuerdo a las palabras que contiene, estos se asocian cada uno con un síntoma. En el estudio, los autores observan que los síntomas más revelados en las publicaciones son aquellos relacionados con la culpabilidad, enojo y depresión, incluso una parte representativa menciona ideas de auto-lesión y además, destacan que en la mayoría de los casos se expresa más de un síntoma. Otra observación muestra que las publicaciones relacionadas con depresión son mayormente hechas por mujeres en su mayoría con edades por debajo de los 25 años. Su análisis va dirigido a publicaciones de la plataforma Twitter, pues revelan que es una de las redes sociales más usadas por los jóvenes.

3.2. Enfoques de clasificación translingüe

La clasificación translingüe consiste en asignar clases a documentos escritos en un idioma objetivo usando recursos de un idioma fuente [Gliozzo and Strapparava, 2006]. Ésta ofrece la ventaja de tomar datos previamente etiquetados en un idioma para ser utilizados en otro donde se carece de datos, ofreciendo así una solución factible para problemas de este tipo.

Hasta el momento ningún estudio que involucra detección de depresión a partir de textos ha sido guiado hacia una perspectiva de clasificación translingüe. Sin embrago, en otras áreas cercanas como el análisis de sentimientos se han propuesto varios estudios guiados por este enfoque y utilizando diferentes técnicas para la transferencia de conocimiento entre los idiomas involucrados. A continuación se describen algunas investigaciones.

Aprendizaje por correspondencia estructural (SCL)

SCL es una técnica de adaptación de dominios que aplicado al enfoque de clasificación translingüe, los dominios corresponden a los idiomas en cuestión. En este sentido [Prettenhofer and Stein, 2010, Wei and Pal, 2010] han realizado varias investigaciones. El objetivo de esta técnica es encontrar relaciones entre palabras de ambos idiomas a partir de pivotes. Los pivotes son pares de palabras que representan a cada uno de los idiomas y tiene significados similares; en algunos casos los pivotes son elegidos haciendo traducción automática. En ambos trabajos el clasificador se construye calculando la co-ocurrencia de características ordinarias que en este caso serían todas las palabras de los textos respecto a los pivotes. La diferencia consiste en que [Prettenhofer and Stein, 2010] entrenan el clasificador en ambos idiomas, donde el vector característica resultante es el vocabulario de ambos idiomas. En [Wei and Pal, 2010] en cambio, representan las características en un mismo idioma a través de una traducción automática, pero sólo seleccionan las características pivotes de los textos traducidos. Este tipo de técnica es útil cuando se tienen suficientes datos para los idiomas involucrados.

Traducción automática

Otra de las técnicas que se ha utilizado es un traductor automático para hacer la transferencia de un idioma a otro. Basado en esto, [Al-Shabi et al., 2017] exploran la posibilidad de usar la traducción automática para generar un corpus de entrenamiento confiable sobre reseñas en Internet. Extraen n-gramas de palabras con $n = (1, 2, 3)$ utilizando los diferentes pesados existentes para generar sus representaciones aplicándoles varios clasificadores. Ellos concluyen que el rendimiento de los clasificadores varían en dependencia del tipo de producto que se analiza, así como mencionan que se deben buscar alternativas para minimizar el error que es introducido por la traducción automática debido a las diferencias existentes entre

idiomas.

Word embeddings y alineación

Buscando soluciones a los errores que son introducidos por la traducción automática [Abdalla and Hirst, 2017, Yang et al., 2017, Artetxe et al., 2017] proponen el uso de *word embeddings* bilingües para capturar relaciones entre palabras que ocurren en similares contextos a pesar de su idioma, y de este modo reducir las diferencias existentes entre los idiomas. Todos ellos se basan en la observación realizada por [Mikolov et al., 2013b] que palabras similares en diferentes idiomas tienen una representación similar en el espacio de los *embeddings*, para proponer un método que permita el mapeo al mismo espacio de embeddings en diferentes idiomas. En [Abdalla and Hirst, 2017] los autores usan los vectores de las palabras una vez que ya han sido mapeados al mismo espacio para construir sus modelos. Tratándose de análisis de sentimientos proponen como primer modelo definir la polaridad de la palabra, positiva o negativa, a partir de su vector, y como segundo paso la clasificación de reseñas de productos a partir de los vectores de las palabras que la componen usando el recurso psicolingüístico ANEW. En todos los casos los modelos fueron entrenados sólo en uno de los idiomas y probados con el otro. Observan que es posible tener palabras relacionados pero con pobre predicción y que no es necesario un mapeo exacto para lograr buenos resultados de clasificación, incluso puede ser utilizado en otras tareas.

Para lograr la alineación o el mapeo entre los *word embeddings* de los diferentes idiomas, [Artetxe et al., 2017] proponen un método que partiendo de un diccionario de semillas y los *word embeddings* de ambos idiomas, a través de un proceso iterativo y de optimización establecen una relación contextual entre las palabras. Uno de los aportes más importantes de esta investigación es que alcanzan buenos resultados partiendo de un diccionario de semillas pequeño (25 palabras), diferenciándose así

de otros métodos propuestos donde es necesario tener un gran número de palabras en el diccionario. En la presente investigación se hace uso de este método para alinear las palabras apoyados en las ventajas y los buenos resultados mostrados por el método antes mencionado.

3.3. Discusión del capítulo

Los trabajos expuestos anteriormente reportan modelos que permiten distinguir rasgos entre personas que padecen de depresión y las que no. Éstos, muestran en sus investigaciones la utilidad de los n-gramas de palabras y su frecuencia en la tarea de detección de depresión, resaltando que existe una gran diferencia entre las palabras que utilizan ambos grupos, depresivos y no depresivos. Resaltan además el uso de recursos psicolingüísticos para la captura de emociones, lo cual es un aspecto esencial en el comportamiento de personas con depresión; la mayoría de los estudios coinciden en que las personas con este padecimiento presentan un aumento significativo de emociones negativas y pensamientos de culpa e infelicidad, las cuales son plasmadas en sus publicaciones. Se destaca el análisis de las características que provienen del monitoreo de la actividad en las redes sociales, que aunque aportan ciertos detalles en la descripción y caracterización de usuarios con depresión, no tiene gran influencia sobre una correcta predicción. En cambio, las características provenientes del lenguaje si tienen una influencia positiva sobre la predicción de usuarios con depresión.

Una característica peculiar de estos usuarios es el uso de constante de términos referidos a síntomas y tratamientos relacionados con la enfermedad. El presente trabajo está enfocado en la extracción de características provenientes del texto; tomando en consideración las que se han obtenido previamente en otros estudios y que han resultado exitosas en la detección de la presencia o no de depresión; tales como: n-gramas y el uso de recursos psicolingüísticos para detectar la presencia de

determinadas categorías del recurso contenidas en el texto. Sin embargo, todos estos trabajos han sido guiados hacia una perspectiva de clasificación monolingüe, restringiéndose de este modo al idioma de los datos que se usan, provocando que no puedan ser utilizados en otros idiomas con escasos recursos orientados a esta tarea. Específicamente no se reporta ninguna investigación orientada al idioma Español el cual se presenta con una gran carencia de datos para el dominio de depresión. Motivados por la efectividad del enfoque de clasificación translingüe para la tarea de análisis de sentimientos, guiaremos la investigación a utilizar este enfoque; el cual permite tomar un corpus ya existente y previamente etiquetado con amplios recursos en un idioma específico, para utilizarlo en otro que carece de recursos.

El enfoque de clasificación translingüe se caracteriza por hacer una transferencia entre los idiomas involucrados usando diferentes mecanismos de alineación. Estudios reportan el uso de técnicas de adaptación de dominios tomando cada idioma como los dominios en cuestión. Sin embrago estas técnicas son útiles cuando se manejan grandes cantidades de datos por lo cual no es aplicable a nuestro problema. Otro mecanismo de transferencia utilizado es a través de traductores automáticos, una solución sencilla pero sujeta a errores debido a la diferencia existente entre los idiomas. Los errores de este mecanismo se agudizan cuando se trata de textos provenientes de redes sociales, que por la brevedad de sus textos no existe suficiente contexto para una buena traducción, abundan errores ortográficos y gramaticales, así como el uso frecuente de acrónimos. Debido a que este mecanismo es una solución muy básica y sencilla será la utilizada para obtener los resultados del baseline en la presente investigación.

Por otro lado los *word embeddings* bilingües han sido utilizados para encontrar relaciones entre palabras de diferentes idiomas de acuerdo a su contexto, y de este modo obtener una alineación a nivel de palabra que no necesariamente tiene que ser un mapeo exacto para lograr buenos resultados de clasificación. Esta forma de alinear ofrece como ventaja que permite relacionar palabras aunque no sean una

traducción exacta, lo que favorece incluir más elementos contextuales. En el proceso de alineación además de los embeddings intervienen diccionarios de palabras de diferentes tamaños. El enfoque propuesto se caracteriza por usar una alineación a nivel de palabra para transferir el conocimiento entre idiomas, utilizando un método de alineación a nivel de palabra que permite tener diccionarios de palabras de tamaños pequeños.

Capítulo 4

Método Propuesto

Esta sección describe el enfoque translingüe propuesto para la detección de usuarios de Twitter que padecen de depresión. El método propuesto consta de dos partes fundamentales, una primera que usa un corpus etiquetado en un idioma para detectar depresión en otro idioma realizando la clasificación desde una perspectiva translingüe, y una segunda parte que toma ventaja de esta clasificación y mejora el etiquetado haciendo uso de diccionarios específicos del dominio.

4.1. Descripción general

El método propuesto de forma general usa un conjunto de documentos etiquetados, escritos en un idioma, para detectar depresión en textos escritos en otro idioma. La Figura 4.1 muestra de manera resumida las diferentes etapas del método donde se toma un conjunto de documentos escritos en un idioma fuente, se extraen sus características y se construye la representación, luego se entrena un clasificador que se utiliza posteriormente para predecir clases de un conjunto de prueba escrito en un idioma objetivo diferente. Para llevar a cabo este proceso y lograr transferir el conocimiento de un idioma a otro las características extraídas de los documentos

escritos en el idioma objetivo son sometidas a un proceso de alineación. Una vez asignadas las etiquetas de clase a estos documentos se realiza una etapa de re-etiquetado con el objetivo de obtener etiquetas más acertadas para cada documento. A continuación se detallan las etapas de entrenamiento y prueba, así como las representaciones obtenidas.

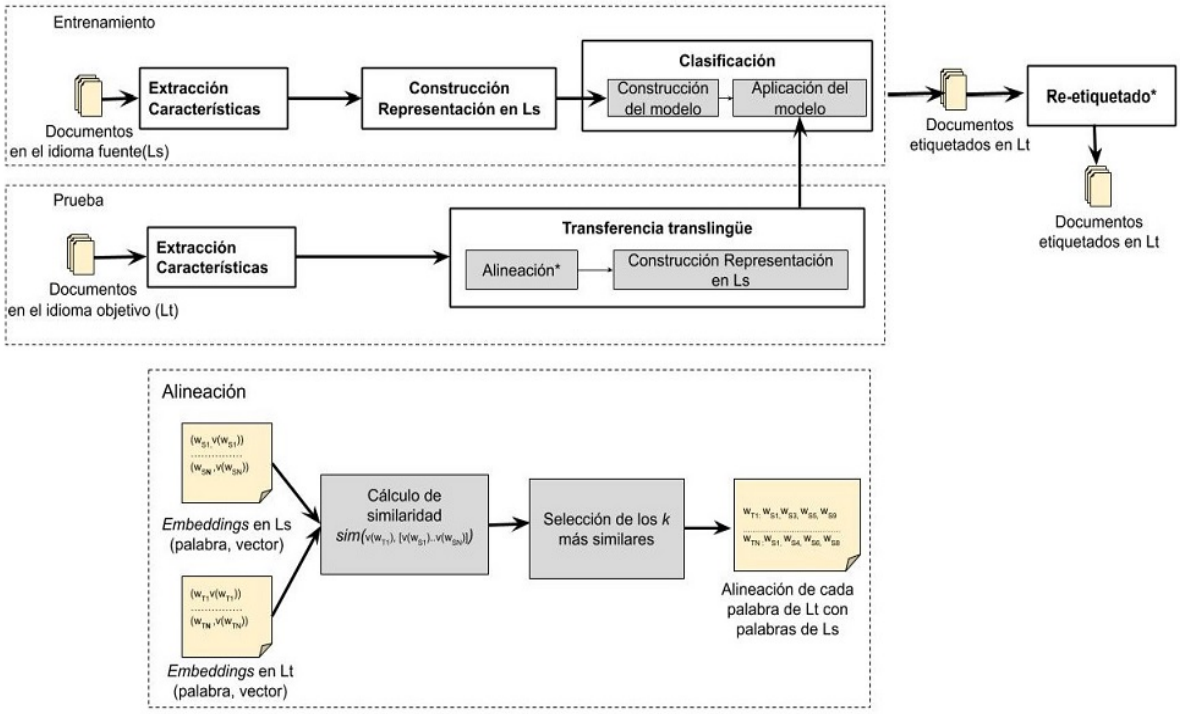

Figura 4.1: Esquema general del método propuesto.

Formalmente basado en [Prettenhofer and Stein, 2010], se tiene $D^s = \{d_1^s, ..., d_{|D^s|}^s\}$ que denota la colección de documentos etiquetados (conjunto de entrenamiento). También se tiene a $D^t = \{d_1^t, ..., d_{|D^t|}^t\}$ que representa el conjunto de documentos no etiquetados (conjunto de prueba). Estos conjuntos están escritos en diferentes idiomas. Comúnmente, el idioma de los documentos de entrenamiento y prueba son llamados idioma fuente e idioma objetivo respectivamente. En este contexto translingüe, el vocabulario V es dividido en $V^s = \{w_1^s, ...w_{|V^s|}^s\}$ y $V^t = \{w_1^t, ...w_{|V^t|}^t\}$, llamados vocabulario del idioma fuente y vocabulario del idioma objetivo respec-

tivamente por lo que el clasificador Φ_s generado de D^s con V^s no puede ser directamente aplicado para clasificar documentos del idioma objetivo. Una solución para minimizar las brecha que existe en los lenguajes sería hacer una traducción automática de los documentos, siendo esta una alternativa básica la cual presenta varias desventajas. Particularmente, respecto a las diferencias presentes entre los idiomas, este trabajo propone dos representaciones translingüe, una basada en un recurso psicolingüístico multilingüe y la otra en una alineación de palabras bilingüe con el objetivo de capturar una correspondencia entre las palabras de los idiomas involucrados. A continuación se describen las representaciones.

4.2. Etapa de clasificación

4.2.1. Representación basada en LIWC

LIWC es un recurso psicolingüístico con origen en el campo de la psicología [Pennebaker et al., 2007]. Actualmente este recurso está disponible en varios idiomas y tiene entre sus ventajas que permite analizar emociones y componentes estructurales y cognitivos presentes en textos. De manera general, las palabras son clasificadas en varias categorías lingüísticas y psicológicas. Formalmente, se tiene $C^s = \{C^s_1, ..., C^s_{|C^s|}\}$ y $C^t = \{C^t_1, ..., C^t_{|C^t|}\}$ que representan el conjunto de categorías de LIWC en las versiones para el idioma fuente y para el idioma objetivo respectivamente. Cada categoría está formada por un conjunto de palabras (unigramas léxicos) denotado por $C_f = \{w_1, ..., w_{|C_f|}\}$.

Alineación. Las versiones de inglés y español presentan una correspondencia entre el conocimiento para ambos idiomas. Específicamente, las categorías de la versión en inglés comparten el mismo índice con las categorías de la versión en español. De manera que C^s_1 corresponde a la traducción de C^t_1 y así con el resto. Para alinear

y crear el vector de los documentos no etiquetados fueron usados los índices de las categorías.

Representaciones de entrenamiento. En esta representación cada documento etiquetado (instancias del entrenamiento) d_i^s es representada por un vector $\mathbf{d_i^s}$, para la cual su espacio de características está determinado por las categorías correspondientes a LIWC:

$$\mathbf{d_i^s} =< v_{i,1}, ..., v_{i,|C^s|} > \tag{4.1}$$

donde $v_{i,j} = \sum_{\forall w \in C_j^s} f(w, d_i^s)$ representa la suma de las ocurrencias de la palabra que pertenece a la categoría C_j^s del diccionario en el documento d_i^s.

Representaciones de prueba. Considerando el mecanismo de alineación, cada documento no etiquetado (instancias de prueba) d_i^t es representado por un vector característica $\mathbf{d_i^t}$, el cual está en el mismo espacio dimensional del espacio de los documentos de entrenamiento:

$$\mathbf{d_i^t} =< v_{i,1}, ..., v_{i,|C^s|} > \tag{4.2}$$

donde $v_{i,j} = \sum_{\forall w \in C_j^t} f(w, d_i^t)$ representa la suma de las ocurrencias de las palabras que pertenecen a la categoría C_j^t del diccionario del idioma objetivo. Como el índice de las categorías de los diccionarios en el idioma fuente y objetivo coinciden, se asumió que los vectores para los documentos en el entrenamiento y la prueba están representados en el mismo espacio.

La representación antes propuesta basada en el recurso psicolingüístico LIWC, ofrece como ventaja que permite hacer un mapeo entre las categorías de los idiomas presentes en la investigación (Inglés y Español), apoyados en que dicho recurso cuenta con diccionarios en varios idiomas, que en su mayoría son traducciones a los diferentes

idiomas. Su desventaja radica, en nuestro caso, que en el idioma español hay muchas conjugaciones que no las hay en inglés y por tanto la misma categoría en ambos idiomas no tienen exactamente la misma cantidad de palabras, lo que puede provocar que varias palabras no se tomen en cuenta.

4.2.2. Representación basada en word embeddings bilingües

Para esta representación, los word embeddings bilingües son usados como una estrategia para alinear los datos entre los idiomas. Se considera una estrategia efectiva debido a su capacidad de relacionar contextos similares y encontrar analogías entre palabras. En este caso, la representación propuesta está basada en los conocidos modelos de bolsa de palabras y n-gramas de palabras. Para obtener el mismo espacio dimensional, las palabras de los documentos no etiquetados fueron mapeadas al idioma fuente a través de un proceso de alineación como se describe a continuación.

Alineación. Se usó el método descrito en [Artetxe et al., 2017] para alinear los *word embeddings* entre diferentes idiomas. Las entradas de este método son X_i, Z_j and D, donde X_i y Z_j son los vectores $n-dimensionales$ de las palabras en el idioma fuente y el idioma objetivo respectivamente, los cuales participan en el proceso de mapeo. D es un diccionario de semillas que actúa como el conocimiento básico bilingüe de los idiomas. Éste está formado por un conjunto de pares de palabras (w_i^s, w_i^t) donde w_i^s es una palabra en el idioma fuente y w_i^t es su traducción al idioma objetivo. Esta traducción fue realizada con la ayuda del Traductor Google. Un ejemplo de lo anterior es: (*depression*, *depresión*), (*movie*, *pelicula*), (*yesterday*, *ayer*). La salida de este método son *word embeddings* en ambos idiomas, los cuales han sido mapeados al mismo espacio. La figura 4.2 muestra el esquema general de la alineación donde se aprende un mapeo W basado en las semillas del diccionario D y los vectores Z y X; este mapeo es usado entonces para aprender un nuevo diccionario D, este proceso

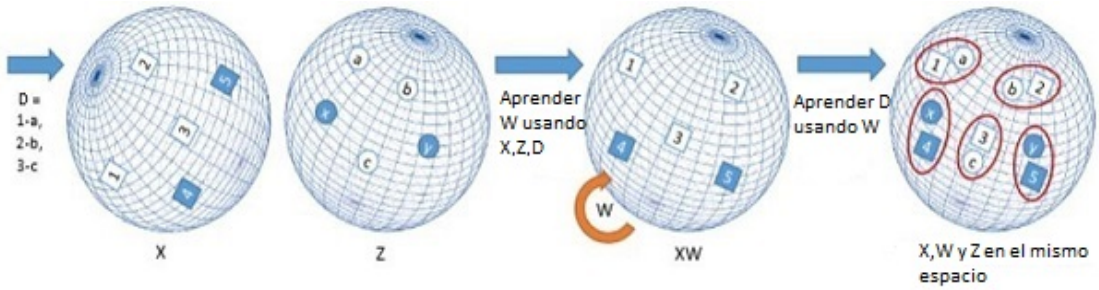

Figura 4.2: Esquema general de la alineación usando *word embeddings*. Fuente: [Artetxe et al., 2017]

es repetido hasta encontrar un criterio de convergencia, el algoritmo 4.1 resume los pasos de este método. Estos vectores resultantes fueron usados para encontrar palabras relacionadas, de manera que por cada palabra en el idioma objetivo su vector es comparado con los vectores de cada palabra del idioma fuente usando la similitud coseno. De este modo las k palabras más similares del idioma fuente son seleccionadas como candidatas para la traducción de la palabra del idioma objetivo. El valor de k puede ser establecido manualmente o automáticamente determinado teniendo en consideración un umbral de similaridad δ.

Algoritmo 4.1 Pasos para la alineación

Entradas: X (Embeddings en el idioma fuente)

Entradas: Z (Embeddings en el idioma objetivo)

Entradas: D (Diccionario de semillas)

repetir:

1. W: Aprender Mapeo (X, Z,D)
2. D: Aprender Diccionario (X,Z,W)

hasta: criterio convergencia

Salidas: X, Z mapeados al mismo espacio

Representaciones de entrenamiento para BoW. Para esta representación propuesta cada documento etiquetado (entrenamiento) d_i^s es modelado por un vector $\mathbf{d_i^s}$ como sigue:

$$\mathbf{d_i^s} =< v_{i,1}, ..., v_{i,|V^s|} > \tag{4.3}$$

done $v_{i,j} = f(w_j^s, d_i^s)$ representa la frecuencia normalizada de la palabra w_j^s en el documento d_i^s.

Representaciones de prueba para BoW. Por otro lado, para construir la representación de los documentos de prueba es necesario mapear cada palabra del idioma objetivo a sus c más similares palabras del idioma fuente, de acuerdo al mecanismo de alineación. Por ejemplo para las palabras en la siguiente oración: "el *dolor explota* en mi *alma*", son transformadas a: (*pain*,*fear*,*emptiness*), (*explode*,*combust*,*suffocate*), (*soul*,*heart*, *mind*). El documento transformado se denota como $\breve{d}_i^t$, cuya longitud máxima es $c * |d_i^t|$. La representación de los documentos de prueba es construida a partir de los documentos transformados, el cual contiene términos del lenguaje fuente. Específicamente cada $\breve{d}_i^t$ es modelado por un vector $\mathbf{d_i^t}$:

$$\mathbf{d_i^t} =< v_{i,1}, ..., v_{i,|V^s|} > \tag{4.4}$$

donde $v_{i,j}$ representa el *pesado* de la palabra w_j^s en el documento transformado $\breve{d}_i^t$. Se consideraron dos diferentes tipos de pesado, uno basado en frecuencia y otro basado en similaridad. A continuación se describen estas dos variantes, las cuales son normalizadas por la suma de los pesos del vector $\mathbf{d_i^t}$.

Pesado basado en frecuencia: En este caso $v_{i,j} = f(w_j^s, \breve{d}_i^t)$ representa la ocurrencia de la palabra w_j^s en el documento $\breve{d}_i^t$.

Pesado basado en similaridad: Como se describió previamente, cada palabra del idioma objetivo podría ser mapeada a *varias* palabras del idioma fuente; como consecuencia, cada palabra w_j^s del idioma fuente está conectada con un conjunto de palabras del idioma objetivo denotado como Q_j. Dado que cada par mapeado tienen diferentes confianzas debido a que la similitud de sus palabras es diferente, el peso de la palabra w_j^s en el documento transformado $\breve{d}_i^t$ es calculado como sigue:

$$v_{i,j} = \sum_{\forall w \in Q_j} sim(w, w_j^s),$$

donde $sim(w, w_j^s)$ representa el valor de la similaridad coseno entre los vectores de los embeddings de las palabras $w \in d_i^t$ y w_j^s.

Esta representación basada en una alineación a nivel de palabra ofrece como ventaja que al encontrar relaciones entre las palabras de ambos idiomas sin que sean una traducción exacta, las palabras del lenguaje objetivo puede ser representadas a través de las del idioma fuente, sin embargo tiene como desventaja que en muchas ocasiones no se alcanza una correcta asociación entre las palabras y esto puede tener influencia negativa sobre la clasificación.

4.3. Depuración de las etiquetas del conjunto de datos del español.

El proceso de crear y recolectar un conjunto de datos para la fase de entrenamiento, con un correcto etiquetado, es un proceso complejo, costoso y que consume tiempo; especialmente si el dominio requiere del análisis de expertos. A ello se suma la carencia de datos para un dominio o un idioma específico. Ésto, afecta también el proceso de clasificación. Específicamente el dominio de depresión para el idioma español, tiene como limitante la escasez de datos y con ello trae aparejado el problema del etiquetado correcto, lo que dificulta la creación de modelos que puedan ser

aplicables al idioma antes mencionado.

La depresión de acuerdo a varios estudios se distingue por el lenguaje específico que usan las personas que lo padecen, ya sea refiriéndose a síntomas y tratamientos relacionados con la enfermedad o a sus sentimientos, que por lo general tienden a ser sentimientos negativos. En la presente investigación se aprovecha la ventaja que ofrece tener clases que se distinguen por el uso de ciertas palabras, para refinar y depurar las etiquetas de clase de los datos que se analizan. Motivados por la idea de que las palabras más frecuentes en los datos de una clase y que presenten menor frecuencia en la otra ayudan a identificar ejemplos que representen mejor a dichas clases, se propone un método para la depuración de las etiquetas de clases.

El punto inicial del proceso de depuración son las etiquetas asignadas en un proceso de clasificación previo, explicado en la sección 4.2. El proceso de depuración de las etiquetas de clase que se propone a continuación, utiliza las etiquetas que se obtienen de la clasificación antes mencionada para refinarlas, y obtener nuevas etiquetadas más acertadas para los documentos.

En la figura 4.3 se resumen los pasos a seguir para el proceso de re-etiquetado que se llevó a cabo. Los documentos con sus respectivas etiquetas son pasados por un proceso de depuración basado en diccionarios, el cual genera un nuevo conjunto de entrenamiento y de prueba. Estos conjuntos son utilizados posteriormente para una nueva clasificación.

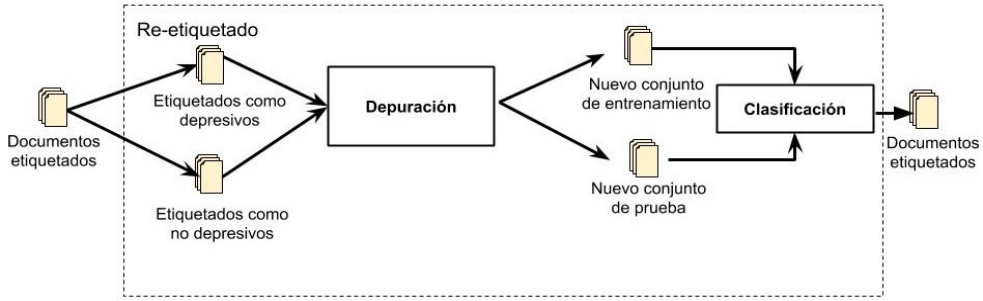

Figura 4.3: Proceso de re-etiquetado

Formalmente, se tiene D^t el conjunto de documentos en español etiquetados como resultado de una clasificación anterior. Se tiene D^t_d es el conjunto de documentos que fueron etiquetados como depresivos y D^t_{nd} el conjunto de documentos etiquetados como no depresivos, de modo que $D^t = D^t_d \cup D^t_{nd}$ y d el diccionario involucrado en el proceso de depuración. Se considera también a D'^t_d como el conjunto de documentos seleccionados para representar los ejemplos de la clase depresiva en el entrenamiento y D'^t_{nd} el conjunto de documentos seleccionados para representar la clase no depresiva en el entrenamiento.

El proceso de depuración consiste en: dentro del conjunto D^t_d se seleccionan para el entrenamiento los k documentos que tuvieran mayor presencia de las palabras del diccionario d formando el conjunto D'^t_d, y en el otro caso dentro del conjunto D^t_{nd} se seleccionaron los k documentos con menor presencia de las palabras del diccionario d formando el conjunto D'^t_{nd}. El resto de los documentos en ambos conjuntos se toman para la fase de prueba. La siguiente fase consiste en el proceso de clasificación, para ello se construye un clasificador basado en n-gramas de palabras usando como pesado la frecuencia normalizada.

Capítulo 5

Experimentos y resultados

En el presente capítulo se describen los experimentos realizados así como la configuración que se utilizó para cada uno de ellos incluyendo la descripción de los datos y demás recursos necesarios para su realización. También se exponen los resultados y se incluye un análisis del proceso de alineación, el cual fue un elemento indispensable en el desarrollo del presente trabajo.

5.1. Colección de documentos

Para hacer la evaluación del método propuesto se utilizaron datos en inglés y en español provenientes de Twitter. Los datos del inglés fueron obtenidos de [Shen et al., 2017]. De acuerdo con los autores los usuarios fueron etiquetados como depresivos si alguna de sus publicaciones contenía la expresión ("*I'm/I was/I've been*)...*diagnosed with depression*". Los no depresivos fueron etiquetados si ninguno de sus tuits contenía la palabra *"depress"*. A continuación se detallan los datos en inglés y español. Siguiendo el mismo enfoque se obtuvieron los datos en español para los usuarios depresivos utilizando la frase ("*Me han diagnosticado/He sido diagnosticado/Me diagnosticaron*)...*depresión"* y para los no depresivos que sus publicaciones no con-

tuvieran la expresión *"depresión"*. La tabla 5.1 resume algunas estadísticas de los datos.

Tabla 5.1: Descripción de los datos Inglés y Español

Idioma	Clase	Número de usuarios	Tamaño del vocabulario	Promedio de Tweets por usuario
Inglés	Deprimidos	2626	15239	220
	No-Deprimidos	5367	70258	1115
Español	Deprimidos	91	14410	4147
	No-deprimidos	225	25231	3844

Como puede observarse en la tabla 5.1 existe desbalance entre ambas clases, predominando los ejemplos de la clase no-depresiva. Este hecho esta relacionado a que de manera general en las redes sociales se encuentran más ejemplos de usuarios que no presentan síntomas de depresión.

Para un mejor detalle de la colección de documentos que se construyó para el idioma Español, las tablas 5.2 y 5.3 muestran un ordenamiento de las 10 palabras con mayores valores de información mutua (PMI) respecto a la clase y las palabras con mayor frecuencia en los documentos en ambas clases respectivamente.

Tabla 5.2: Palabras ordenadas de acuerdo a su PMI para las clases depresiva y no-depresivas.

Palabra	Clase	Valor PMI	Palabra	Clase	Valor PMI
depresión	Depresiva	0.439	vacaciones	No-Depresiva	0.148
abandonado	Depresiva	0.414	manifestación	No-Depresiva	0.122
insultado	Depresiva	0.408	cerveza	No-Depresiva	0.120
medicación	Depresiva	0.393	protagonizan	No-Depresiva	0.102
suicidio	Depresiva	0.388	chingaderas	No-Depresiva	0.100
doloroso	Depresiva	0.256	inconstitucional	No-Depresiva	0.098
deprimido	Depresiva	0.221	cinépolis	No-Depresiva	0.085
destroza	Depresiva	0.205	éxito	No-Depresiva	0.083
amargar	Depresiva	0.187	empoderar	No-Depresiva	0.079
solitario	Depresiva	0.152	asadas	No-Depresiva	0.077

Tabla 5.3: Listado de palabras más frecuentes en los textos del idioma Español para las clases depresivas y no depresivas.

Palabra	Clase	Palabra	Clase
depresión	Depresiva	bien	No-Depresiva
peor	Depresiva	feliz	No-Depresiva
miedo	Depresiva	gracias	No-Depresiva
morir	Depresiva	amigos	No-Depresiva
triste	Depresiva	amor	No-Depresiva
problema	Depresiva	siempre	No-Depresiva
odio	Depresiva	semana	No-Depresiva
difícil	Depresiva	mujeres	No-Depresiva
necesito	Depresiva	historia	No-Depresiva
culpa	Depresiva	bonito	No-Depresiva

5.2. Configuración experimental

5.2.1. Extracción de atributos

El presente trabajo propone representaciones sencillas que capturen la correspondencia entre los idiomas involucrados. Se extrajeron del texto atributos como los *n-gramas de palabras* para *n=1, n=2, n=3*. También se utilizaron las categorías del recurso psicolingüístico LIWC y sus versiones de inglés y español como atributos. Ambas representaciones han demostrado ser efectivas para la tarea que se trata en la presente investigación.

5.2.2. Clasificación

Para el proceso de clasificación se usaron los algoritmos de clasificación *Naïve Bayes (NB)* y *Support Vector Machine (SVM)* por la efectividad que han mostrado ambos en previas investigaciones para la detección de depresión. Fueron evaluados los tres tipos de pesados, *booleano*, *frecuencia(tf)* y *tf-idf*, sin embargo el pesado de frecuencia fue el que mejores resultados alcanzó por lo que se muestran en la sección 5.3 sólo los resultados que se obtuvieron para este tipo de pesado. Para la evaluación se usó el enfoque de validación cruzada con 4-fold. El conjunto fue particionado en 4-fold, estableciendo una proporción 75 % para entrenamiento y 25 % para la prueba, con el objetivo de minimizar las diferencias entre las cantidades de instancias que representan a cada clase en cada partición. Para evaluar la calidad de las representaciones propuestas se utilizó el *F1* de la clase depresiva como medida de evaluación. Todos los experimentos fueron implementados sobre el lenguaje de programación *Python* usando la herramienta *ScikitLearn* y se usó la misma configuración para todos los experimentos.

5.2.3. Alineación con word embeddings

Como se mencionó anteriormente, el método basado en la alineación usando word embeddings requiere de vectores de palabras tanto en el lenguaje fuente como en el lenguaje objetivo, X y Z respectivamente, así como diccionarios de semillas D. Particularmente los vectores en Inglés (X) y Español (Z) fueron calculados sobre 60,000 y 10,000 historiales de usuarios de Twitter. Los historiales en Inglés fueron tomados de [Shen et al., 2017] y los de Español fueron tomados de [Álvarez-Carmona, 2019]. Los vectores fueron entrenados con Word2Vec y usando un modelo basado en skip-gram. Los vectores fueron representados en un espacio dimensional de 200. Referente a las semillas de diccionario se construyeron seis diferentes diccionarios D seleccionando palabras del corpus en Inglés: **d-25**, **d-200**, **d-500**, **d-1000** que contienen 25, 200, 500 y 1000 palabras respectivamente seleccionadas de forma aleatoria, y **d-depres200**, **d-depres25** que contienen las 200 y 25 palabras más frecuentes respectivamente de los usuarios depresivos. Para formar los pares de palabras, las semillas fueron traducidas usando el Traductor Google.

5.2.4. Baseline

Para generar el baseline, con el cual compararemos el resto de los experimentos, se consideró una solución sencilla al problema de la clasificación translingüe [Prettenhofer and Stein, 2010]: se usó un enfoque basado en una traducción automática donde los documentos del lenguaje objetivo D^t (Español) fueron traducidos al lenguaje fuente (Inglés) usando el Traductor Google. De esta manera el clasificador Φ_s es entrenado en los documentos etiquetados del Inglés D^s y probado en los documentos no etiquetados del español D^t que han sido traducidos al Inglés. El método para el baseline fue propuesto con dos representaciones, la primera basada en el tradicional enfoque BoW usando como pesado la frecuencia normalizada. La segun-

da representación es basada en LIWC donde las características corresponden a las categorías C^s de este recurso. En este caso el pesado corresponde al porcentaje de palabras de cada categoría que ocurren en el documento.

5.3. Experimentos monolingües y Baseline

El objetivo principal de este experimento respecto a la clasificación monolingüe es ofrecer una perspectiva general del rendimiento de la clasificación para los idiomas involucrados en la presente investigación (Español e Inglés). Las gráficas a continuación muestran los resultados de los experimentos monolingües, teniendo en cuenta una representación basada en n-gramas de palabras donde $n = 1, 2, 3$ y basada en el recurso psicolingüístico LIWC, así como los resultados obtenidos para el baseline con las mismas representaciones. Los experimentos fueron realizados para ambos idiomas tanto Español (ver gráfica 5.1) como Inglés (ver gráfica 5.2) y se presentan los resultados para dos diferentes clasificadores, *Naïve Bayes (NB)* y*Support Vector Machine (SVM)*. Los resultados obtenidos fueron tomados como el valor ideal esperado del rendimiento para la clasificación. En el caso del baseline (ver gráfica 5.3) como su nombre indica tiene como objetivo establecer el límite inferior esperado en la clasificación. En este caso fueron traducidos los documentos del idioma objetivo (Español) al idioma fuente (Inglés) haciendo uso de un traductor automático como se explica en la sección 5.2.4.

Los resultados obtenidos en la clasificación para el baseline fueron comparados con los obtenidos en los experimentos monolingües considerando las mismas representaciones y teniendo en cuenta la influencia de la traducción automática en la clasificación.

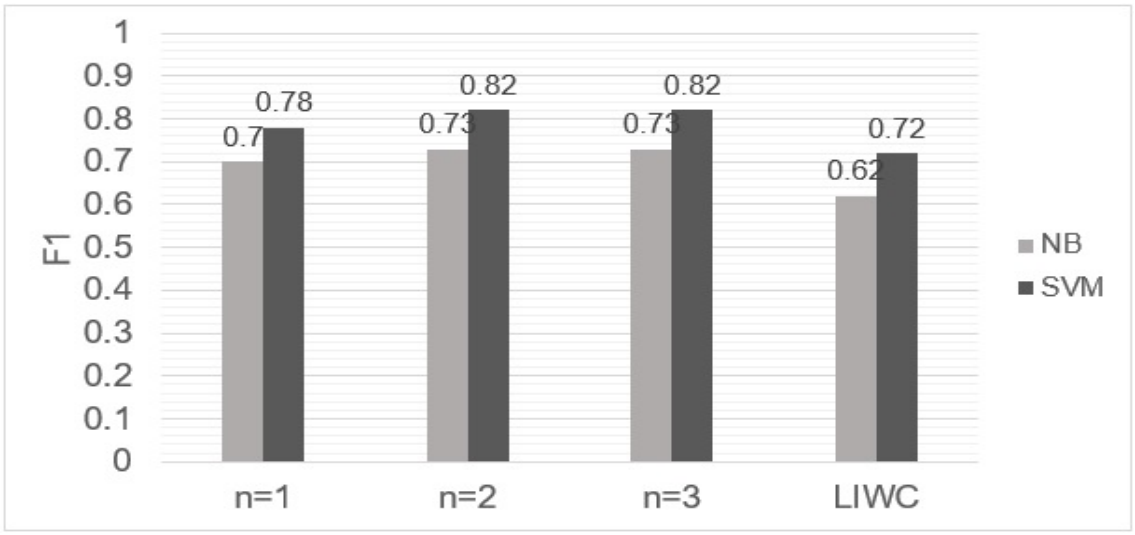

Figura 5.1: Resultados de la clasificación monolingüe para el idioma Español.

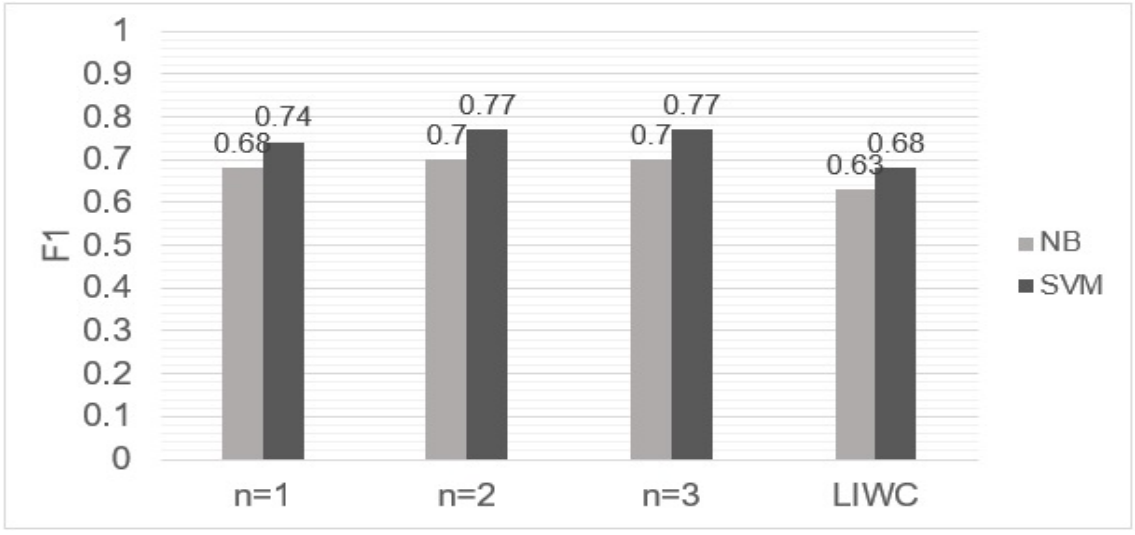

Figura 5.2: Resultados de la clasificación monolingüe para el idioma Inglés.

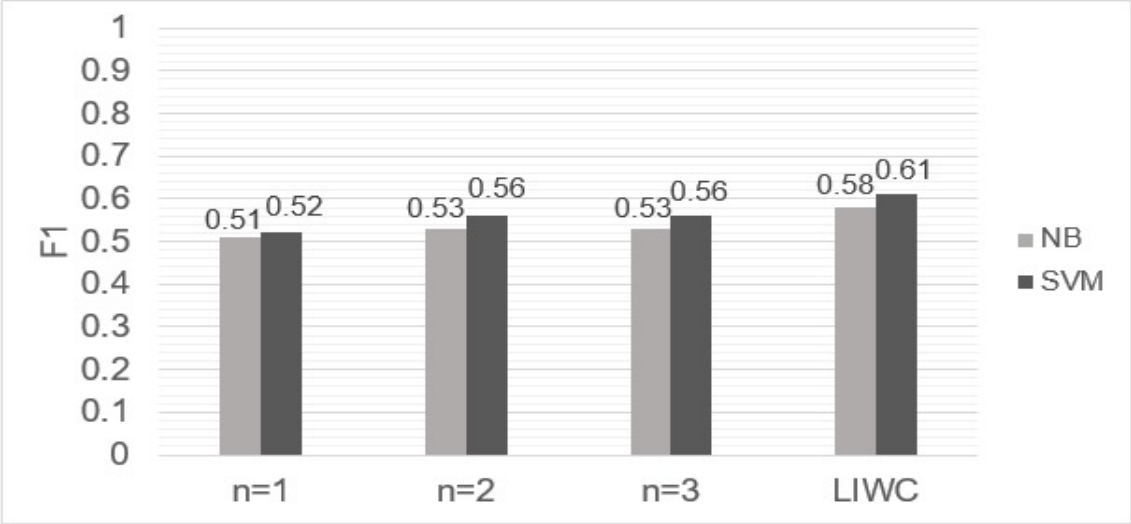

Figura 5.3: Resultados de la clasificación del baseline.

Como se observa, en las gráficas 5.1 y 5.2, se obtienen los mismos resultados para $n = 2$ (bigramas) y $n = 3$ (trigramas), esto significa que la información que ofrecen los tri-gramas no aportan elementos para mejorar el rendimiento del clasificador, sin embargo, si comparamos cuando $n = 1$ (unigramas) y $n = 2$, en este último caso los valores son superiores para ambos clasificadores, demostrando de esta manera la utilidad de los bigramas para la tarea de detección de depresión. Otra observación importante del experimento es que SVM alcanza mejores resultados que NB.

Haciendo una comparación entre los resultados obtenidos por la clasificación monolingüe y el baseline se observa una importante caída en los resultados. Ésto muestra que la traducción automática no es suficiente para transferir ideas y sentimientos de un lenguaje a otro.

La tabla a continuación muestra los unigramas y bigramas con mayor ganancia de información para la clase depresiva.

Tabla 5.4: Valores de ganancia de información (GI) de 1-gramas y 2-gramas para la clase depresiva.

1-gramas	**GI**	**2-gramas**	**GI**
depresión	0.2119	salud#mental	0.2003
suicidio	0.2068	dan#asco	0.1854
ansiedad	0.1809	quiero#morir	0.1742
emocional	0.1543	nadie#habla	0.1669
mental	0.1107	estabilidad#emocional	0.1577

5.4. Experimentos translingües

Teniendo en cuenta los resultados mostrados en la sección anterior y la importante caída de los mismos usando como mecanismo de transferencia entre idiomas una traducción automática, en la presente se presentan los experimentos realizados bajo una perspectiva translingüe utilizando como mecanismo de transferencia una alineación con *word embeddings* a nivel de palabra como se explica en la sección 4.2.2.

La tabla 5.5 muestra una comparación entre los resultados que se obtuvieron para las representaciones basadas en n-gramas de palabras con $n = 1$(BOW) y basada en LIWC en los experimentos monolingües para el idioma español, el baseline y los basados en la alineación con *word embeddings* a nivel de palabra. En este experimento nos referimos como AT-LIWC y AT-BOW a las representaciones obtenidas a partir de la traducción automática, así como BA-LIWC a la representación translingüe (tomando Inglés como idioma fuente y Español como idioma objetivo) basada en LIWC y BA-BOW a la representación basada en *word embeddings* bilingüe, este último con la configuración básica: un diccionario con 500 pares de palabras como

semillas de diccionario, k=1(cada palabra del español se alineó a una sola palabra del inglés), y como pesado la frecuencia normalizada para ambos algoritmos de clasificación *NB* y *SVM*.

Tabla 5.5: Comparación de los resultados de la clasificación entre los diferentes mecanismos de alineación.

Enfoque	Mecanismo de alineación	Representación	NB	SVM	Representación	NB	SVM
Monolingüe	-	BOW	0.70	0.78	LIWC	0.62	0.72
Translingüe	Traducción	AT-BOW	0.51	0.52	AT-LIWC	0.58	0.61
	Alineación	BA-BOW	0.58	0.61	BA-LIWC	0.61	0.65

Los resultados muestran una caída del rendimiento de los experimentos translingües con respecto a los monolingües, indicando que cada idioma tiene ciertas particularidades cuando se expresa la depresión. Respecto a los experimentos translingüe, es importante mencionar que la pérdida del rendimiento introducido por el proceso de traducción automática es mayor que en la representación propuesta. Esto sugiere que las representaciones propuestas basadas en la alineación son útiles para esta tarea.

Adicionalmente, se observa que en los experimentos monolingües, los resultados obtenidos usando una representación basada en uni-gramas ($n = 1$) son superiores que los de la representación con LIWC, mientras que en los experimentos translingües ocurre lo contrario. Con base en estos resultados se considera que la representación basada en la alineación con *embeddings* toma ventaja de alinear varias palabras del inglés con cada palabra del español. El siguiente experimento se enfoca en este análisis.

Una observación importante es que los resultados alcanzados por *SVM* en todos los casos son superiores que los alcanzados con *NB*. En lo adelante los experimentos mostrarán los resultados obtenidos por el clasificador *SVM*.

5.4.1. Parametrización del enfoque de alineación

El objetivo del experimento es evaluar diferentes configuraciones de la representación basada en la alineación con los embeddings descrita en la sección 4.2.2 y en el experimento anterior referida como BA-BOW. Se evalúa además en este experimento la influencia en el rendimiento de dos aspectos: el diccionario D y el número de palabras involucradas en la alineación al lenguaje fuente. Los experimentos consideran además dos esquemas de pesado.

Para evaluar la influencia del diccionario se utilizaron diferentes números de semillas por lo que varía el tamaño del mismo y el dominio, de modo que para variar el tamaño del diccionario se seleccionaron aleatoriamente palabras de los textos escritos en inglés y para especificar un dominio fueron seleccionadas palabras relacionadas con el dominio de depresión. El propósito es identificar las características del diccionario que ayude a maximizar el rendimiento del enfoque propuesto. Se evaluaron varios diccionarios: **d-25**, **d-200**, **d-500** y **d-1000** donde fueron elegidas aleatoriamente 25, 200, 500 y 1000 semillas respectivamente, para diccionarios específicos del dominio se evaluaron tres diferentes diccionarios: **d-depress(25)** y **d-depress(200)** donde fueron elegidas las 25 y 200 palabras más frecuentes de los usuarios depresivos como semillas respectivamente. Para hacer la traducción de las palabras al idioma objetivo (Español) y conformar los pares de semillas se utilizó el *Traductor de Google*.

Para evaluar la influencia del número de palabras alineadas con cada palabra del lenguaje objetivo se consideraron diferentes valores de k. Específicamente se tomaron, la más similar y las cinco más similares respectivamente del lenguaje fuente para cada palabra del idioma objetivo de modo que $k = 1$ y $k = 5$ respectivamente. Adicionalmente a eso se utilizó un criterio de selección dinámica, el cual selecciona todos los elementos alineados cuyos valores de similitud están por encima de un umbral δ determinado. El umbral δ fue definido de la siguiente forma: fue calculado

el promedio de los valores de similitud entre cada palabra del español y su más similar en inglés; a este valor se le restó su desviación estándar. Este umbral tiene dos ventajas: se restringe a seleccionar palabras con altos valores de similitud y es flexible para recuperar varias palabras para aquellas que tienen muchos términos relacionados en el idioma fuente. En la tabla 5.6 se muestran los resultados de este experimento.

Tabla 5.6: Resultados de la clasificación para la representación translingüe propuesta usando la frecuencia como pesado.

Tipo de pesado	Diccionarios	k		
		k=1	k=5	δ
Frecuencia	d-25	0.50	0.50	0.51
	d-200	0.53	0.54	0.57
	d-500	0.58	0.61	0.62
	d-1000	0.58	0.62	0.63
	d-depress(25)	0.56	0.58	0.58
	d-depress(200)	0.58	0.62	**0.65**

Tabla 5.7: Resultados de la clasificación para la representación translingüe propuesta usando la similaridad como pesado.

Tipo de pesado	Diccionarios	k		
		k=1	k=5	δ
Similaridad	d-25	0.57	0.60	0.60
	d-200	0.60	0.6	0.63
	d-500	0.62	0.64	0.66
	d-1000	0.62	0.67	**0.69**
	d-depress(25)	0.62	0.65	0.67
	d-depress(200)	0.62	0.67	**0.69**

Los resultados muestran que la representación propuesta toma ventaja del diccionario que tiene como semillas palabras específicas del dominio cuando se usa un pesado basado en frecuencia. Cuando se usa un pesado basado en similaridad, los resultados no muestran distinción entre usar diccionarios con palabras específicas del dominio o diccionarios con grandes cantidades de palabras; en el caso en que los diccionarios presentan pocas palabras los resultados tienen una gran caída respecto a los que presentan mayor número. Otra de las observaciones derivadas de los resultados arrojados es la importancia de tener semillas con calidad dígase palabras específicas en un dominio que un gran número de semillas.

De forma general los resultados de las tablas 5.6 y 5.7 muestran que el enfoque se favorece cuando varias palabras del idioma fuente son alineadas con cada palabra del idioma objetivo. En la tabla 5.7 se observa además una mejora notable con las representaciones que usan un pesado basado en similaridad mostrando que el mejor resultado con este esquema de pesado mejora considerablemente los resultados del baseline basado en traducción automática (AT-BOW, de la tabla 5.5), lo que representa la conveniencia de esta representación propuesta.

5.4.2. Alineación para bi-gramas y tri-gramas

Los bi-gramas y tri-gramas son representaciones que, como se mencionó en la sección 5.3, son útiles para la detección de depresión, lo que demuestra que además de las palabras también existen pequeñas frases o combinaciones de palabras que ayudan a distinguir a personas con depresión. Motivados por esto en la presente sección se muestran los resultados de la clasificación translingüe para bi-gramas y tri-gramas.

Los resultados de los experimentos que se muestran a continuación se basan en utilizar la alineación previamente obtenida para alinear los bi-gramas ($n = 2$) y tri-gramas ($n = 3$) del idioma fuente con los del idioma objetivo y representarlos.

Se escogió la alineación que mejor resultado mostró en los experimentos anteriores, la cual se obtuvo usando el diccionario **d-depress(200)**, y δ para el número de palabras del idioma fuente alineadas con cada palabra del idioma objetivo.

Por cada palabra del idioma objetivo w_i^t se tiene una lista L^s de palabras en el idioma fuente, las cuales son candidatas a la traducción de w_i^t. Se tiene también el conjunto de n-gramas obtenidos de cada documento del idioma objetivo(prueba) D_i^t. Separando los n-gramas por las palabras que lo forman se utilizó la lista L^s de cada una de ellas para obtener todas las posibles combinaciones de n-gramas que se pudieran formar usando sus traducciones candidatas. Para ello se utilizó el producto cartesiano entre ambas listas, de manera que el primer elemento corresponde a la primera lista y el segundo a la segunda lista, conservando el orden de las palabras. De este modo cada n-grama del lenguaje objetivo tiene asociado una lista de n-gramas en el idioma fuente como candidatos a su traducción. Los resultados de la clasificación para los n-gramas de palabras para $n = 2$ y $n = 3$ se muestran en la tabla 5.8.

Tabla 5.8: Resultados de la clasificación para n-gramas de palabras con n=1, n=2

Representación	F1
n=2	0.62
n=3	0.65

Los resultados muestran una ligera mejora sobre los obtenidos en el baseline pero muy inferiores a los de la representación anteriormente propuesta. Esto demuestra que se debe indagar y explorar otras formas más efectivas para alinear bigramas y trigramas entre ambos idiomas.

5.5. Análisis del proceso de alineación

Para profundizar y comprender mejor los resultados alcanzados se realizó un análisis de las alineaciones generadas tras el proceso. Varios ejemplos son mostrados en la tabla 5.9. Como muestran los resultados de los experimentos es notoria la ventaja de la alineación con *word embeddings* contra la traducción automática.

Primeramente es posible observar que el traductor en ocasiones no puede traducir correctamente varias palabras que se usan comúnmente en las redes sociales, mientras que la alineación de palabras puede encontrar varias palabras relacionadas. Ejemplo de ellos es el caso de la palabra *pendeja* que es un término ofensivo en español; el traductor tiene problemas para reconocerla cuando varía la función sintáctica en la oración, específicamente cuando es usada como sustantivo. Lo mismo ocurre con la expresión *no manches*, la cual es usada por las personas mexicanas para expresar sorpresa, sin embargo la traducción que le asigna el traductor automático proviene del verbo *manchar*, en inglés *stain*, lo cual es incorrecto; en cambio el enfoque propuesto encuentra varias palabras en inglés alineadas con ella que expresan emociones similares.

Por otra parte, se encontraron varias palabras que fueron correctamente traducidas por la traducción automática, sin embrago, la alineación propuesta muestra varios beneficios considerando más palabras relacionadas. Por ejemplo la palabra *chido* es un adjetivo usado por los mexicanos para referirse a cosas buenas, en este caso la alineación la asocia a varios términos en inglés que se refrieren a lo mismo. La tabla 5.9 muestra además palabras en español relacionadas con el dominio de depresión que fueron alineadas con varias palabras en inglés usadas en similares contextos.

En los tres últimos ejemplos mostrados en la tabla anterior se pueden observar palabras en los que falla la alineación lo que ciertamente tiene influencia negativa

sobre los resultados de clasificación.

Tabla 5.9: Ejemplos de varias palabras con sus términos alineados

	Método de alineación	
Palabra en español	Traductor	Embeddings
pendeja	*pendeja/fool*	*bitch, stupid, crazy*
no manches	*do not stain*	*omg, wow, oh*
chido	*cool*	*cool, cute, nice, good, awesome*
tristeza	*sadness*	*sadness,depression, pain, anxiety, loneliness*
ansiedad	*anxiety*	*anxiety, depression, stress,frustration*
miedo	*fear*	*fear, doubts,weakness, frustration*
amarte	*loving you*	*die, cry, disappear*
llorar	*cry*	*signs*
ausencia	*absence*	*presence*

La tabla 5.10 muestra oraciones en las que algunas de las palabras mencionadas como ejemplo en la tabla 5.9 intervienen en un contexto más específico. Se muestra la oración original en idioma español, y la traducción de la misma al idioma inglés a través del traductor automático .

Tabla 5.10: Ejemplos de errores en la traducción de oraciones por el traductor automático

Oración español	Oración traducida al inglés
No manches,te ves bien bonita!	***Do not stain**, you look pretty!*
no manches, pobre de su esposa	***do not stain** , poor of his wife*
La gente se vuelve pendeja para manejar	*People become **pendeja** to handle*
Nadie me gana siendo pendeja	*Nobody wins me being a fool*

Una conclusión general de este análisis es que la traducción automática no es suficiente para transferir sentimientos de un lenguaje a otro. Adicionalmente, se observa que la alineación usando *word embeddings* bilingüe enriquece la traducción

de las palabras. No obstante, a pesar de la efectividad de alineación generada por los *word embeddings*, se encontraron algunos errores que afectan el rendimiento en la detección de depresión. Específicamente se encontraron palabras desalineadas tal como se muestra en los últimos tres ejemplos de la tabla 5.9.

5.6. Re-etiquetado del conjunto de datos de español.

El experimento que se muestra a continuación tiene como objetivo refinar los resultados de clasificación para lograr un etiquetado más acertado del conjunto de documentos en español. La idea principal se basa en tomar las etiquetas asignadas por una clasificación previa y a través de un proceso de depuración obtener un nuevo conjunto de entrenamiento y posteriormente una clasificación. El proceso de depuración consiste en seleccionar los n ejemplos más representativos para ambas clases basado en diccionarios de palabras.

Para definir el nuevo conjunto de entrenamiento se utilizaron cuatro diccionarios, **d-25**, **d-200**. **d-depress(25)** y **d-depress(200)**. Para seleccionar los documentos del entrenamiento que representarían a los ejemplos de la clase depresiva, se tomaron los 20 documentos que mayor número de palabras contuvieran de las presentes en el diccionario, y para representar la clase no depresiva se escogieron los 20 documentos que menos palabras contuvieran; el resto de los documentos con sus respectivas clases fueron tomados para la prueba. Por cada diccionario se generó un experimento de clasificación diferente. La figura 5.4 muestra los resultados de cada diccionario para la representación basada en n-gramas de palabras con $n = 1, 2, 3$.

Como se observa en la gráfica que muestra los resultados, el diccionario más útil para el mencionado proceso de depuración es el que contiene 25 palabras específicas del dominio de depresión aunque, los resultados obtenidos para el diccionario que contiene 200 palabras del dominio de depresión están cercanos a los mencionados an-

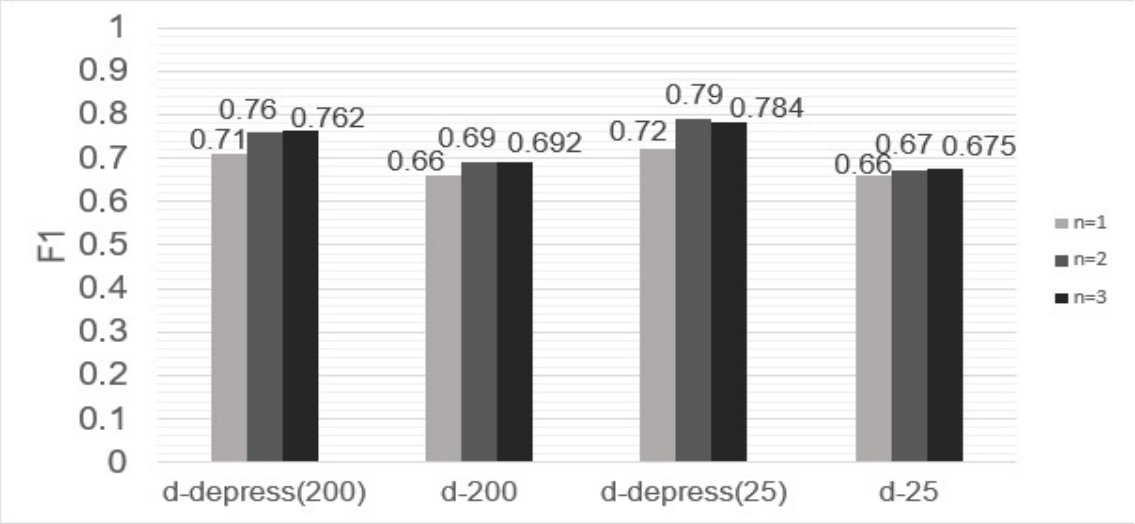

Figura 5.4: Resultados de la clasificación para cada diccionario utilizado en el proceso de re-etiquetado.

teriormente. Esto se debe además a la existencia de intersección no vacía entre ambos diccionarios. En cambio, los resultados alcanzados con los diccionarios que contienen palabras escogidas aleatoriamente **d-200** y **d-25** muestran una caída pues el lenguaje que usan usuarios deprimidos y no deprimidos presentan diferencias bien marcadas. Esto favorece el uso de diccionarios que contengan esas palabras para establecer la diferencia entre los lenguajes de ambos grupos permitiendo así, tomar ventaja en la clasificación. Otra conclusión importante de este experimento es que se observa una recuperación de la clasificación comparado con el mejor resultado obtenido en los experimentos de la sección 5.4 , obteniendo 0.79 sobre 0.69 respectivamente.

5.7. Conclusiones del capítulo

En este capítulo se describieron las colecciones de datos utilizadas en los experimentos, así como el enfoque utilizado para su obtención. También se expusieron las configuraciones utilizadas bajo las que se obtuvieron los resultados mostrados. Se realizaron dos grupos de experimentos: monolingües y translingües. El primero fue

dirigido a la clasificación en los idiomas Español e Inglés, con el objetivo de establecer los valores superiores esperados en la clasificación, y un experimento enfocado a obtener los límites bases de la clasificación donde se usó un traductor automático para transferir conocimiento entre idiomas. El segundo grupo de experimentos fue realizado con el objetivo de transferir el conocimiento entre idiomas usando una alineación a nivel palabra a partir de sus *embeddings*, y las versiones en ambos idiomas (Inglés y Español) del recurso psicolingüístico LIWC.

Los resultados obtenidos mostraron la utilidad de las representaciones propuestas tomando ventaja de la alineación a nivel de palabra. Mostraron además que la traducción automática de textos originados en redes sociales no es suficiente para transferir conocimiento, así como la utilidad de alinear conceptos entre idiomas a partir del recurso psicolingüístico LIWC.

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

Este trabajo está soportado en la idea de que los datos ya etiquetados en un idioma específico pueden ser aprovechados para identificar depresión en otros lenguajes. El enfoque propuesto es basado en un proceso de alineación a nivel de palabra para transferir conocimiento de un idioma a otro. Este proceso de alineación tuvo una importancia notable en la transferencia de conocimiento entre ambos idiomas, donde considerar más elementos contextuales en la clasificación mejoró los resultados. Como una segunda etapa se propuso un proceso de re-etiquetado. En esta etapa los resultados obtenidos muestran la ventaja que ofrece el uso de diccionarios en tareas donde el lenguaje entre las clases presenta una diferencia marcada.

Los experimentos realizados para la clasificación monolingüe mostraron que los bigramas ofrecen mayor información para distinguir usuarios deprimidos que los unigramas en ambos idiomas (Inglés y Español). Demostrando de esta manera que además de palabras específicas también existen frases que resaltan la presencia de la enfermedad y son útiles en la detección de la misma. Se observó además, una

importante caída en los resultados obtenidos para el baseline donde se usa una traducción automática como mecanismo para la transferencia de conocimiento entre los idiomas. En este sentido se observa que este mecanismo no es suficiente para transferir ideas y sentimientos de un lenguaje a otro.

Como parte de los experimentos translingües se observó que a pesar de ser inferiores los resultados a los obtenidos para los monolingües, la pérdida introducida por el proceso de traducción es mayor que en la representación propuesta basada en alineación de palabras. Así como se destaca la utilidad de alinear conceptos entre idiomas apoyado en los resultados alcanzados por la representación basada en LIWC.

De manera general los resultados obtenidos en los experimentos realizados muestran que el enfoque propuesto toma ventaja de aplicar un SVM como clasificador. Reflejan además que la depresión puede ser detectada analizando los textos desde una perspectiva translingüe. Sugiriendo así que a pesar de la diferencia entre los idiomas, usuarios depresivos tienden a expresar su estado de depresión de una forma similar.

Como un aporte importante de este trabajo se menciona la construcción de un corpus de depresión en Español, que aunque pequeño, permitió la evaluación del método propuesto y fue puesto a disposición pública para que posteriormente pueda ser utilizado, sólo con fines científicos, en el estudio y profundización de este trastorno mental.

6.2. Trabajo Futuro

El presente trabajo es pionero en el uso de un enfoque de clasificación translingüe a partir de texto para la detección de depresión. A continuación se exponen algunas ideas que se proponen como trabajo futuro para la presente investigación:

- Explorar el enfoque propuesto para detección de depresión en otros idiomas con el objetivo de observar la existencia de una relación entre las palabras que usan las personas deprimidas en los idiomas involucrados, a pesar de las diferencias culturales que existen, y determinar si el método puede ser adaptado a otros idiomas.

- Aplicación de enfoques translingües en otros desórdenes mentales como es el caso de la anorexia con el objetivo de aprovechar la existencia de datos de entrenamiento existente en algunos idiomas, que puedan ser usados para clasificar documentos en otros donde se carece de recursos.

- En el presente trabajo se extrajeron características usando el recurso psicolingüístico LIWC, así como n-gramas de palabras. Motivados por la fuerte presencia de sentimientos negativos en usuarios con depresión, se propone la extracción de características relacionadas con emociones que basada en la polaridad de las palabras aporten información relevante para distinguir usuarios con depresión.

Bibliografía

[Abdalla and Hirst, 2017] Abdalla, M. and Hirst, G. (2017). Cross-lingual sentiment analysis without (good) translation. *arXiv preprint arXiv:1707.01626*.

[Al-Shabi et al., 2017] Al-Shabi, A., Adel, A., Omar, N., and Al-Moslmi, T. (2017). Cross-lingual sentiment classification from english to arabic using machine translation. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8(12):434–440.

[Álvarez-Carmona, 2019] Álvarez-Carmona, M. Á. (2019). *Author profiling in social media with multimodal information*. PhD thesis, Instituto Nacional de Astrofísica, Óptica y Electrónica.

[Artetxe et al., 2017] Artetxe, M., Labaka, G., and Agirre, E. (2017). Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 451–462.

[Asamblea Mundial, 2013] Asamblea Mundial, d. l. S. (2013). Plan de acción integral sobre salud mental 2013-2020. Technical report.

[Baeza-Yates et al., 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.

[Cavazos-Rehg et al., 2016] Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., and Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, 54:351–357.

[Coppersmith et al., 2014] Coppersmith, G., Dredze, M., and Harman, C. (2014). Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60.

[Coppersmith et al., 2015] Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., and Mitchell, M. (2015). Clpsych 2015 shared task: Depression and ptsd on twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31–39.

[De Choudhury et al., 2013] De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*, pages 128–137.

[Gkotsis et al., 2016] Gkotsis, G., Oellrich, A., Hubbard, T., Dobson, R., Liakata, M., Velupillai, S., and Dutta, R. (2016). The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pages 63–73.

[Gliozzo and Strapparava, 2006] Gliozzo, A. and Strapparava, C. (2006). Exploiting comparable corpora and bilingual dictionaries for cross-language text categorization. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 553–560. Association for Computational Linguistics.

[Guntuku et al., 2017] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., and Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.

[Losada and Crestani, 2016] Losada, D. E. and Crestani, F. (2016). A test collection for research on depression and language use. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 28–39. Springer.

[Losada and Gamallo, 2018] Losada, D. E. and Gamallo, P. (2018). Evaluating and improving lexical resources for detecting signs of depression in text. *Language Resources and Evaluation*, pages 1–24.

[Mikolov et al., 2013a] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[Mikolov et al., 2013b] Mikolov, T., Le, Q. V., and Sutskever, I. (2013b). Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

[Mikolov et al., 2013c] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013c). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

[Nadeem, 2016] Nadeem, M. (2016). Identifying depression on twitter. *arXiv preprint arXiv:1607.07384*.

[Nguyen et al., 2014] Nguyen, T., Phung, D., Dao, B., Venkatesh, S., and Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3):217–226.

[Organization, 2001] Organization, W. H. (2001). *The World Health Report 2001: Mental health: new understanding, new hope*. World Health Organization.

[Pedersen, 2015] Pedersen, T. (2015). Screening twitter users for depression and ptsd with lexical decision lists. In *Proceedings of the 2nd workshop on computational*

linguistics and clinical psychology: from linguistic signal to clinical reality, pages 46–53.

[Pennebaker et al., 2007] Pennebaker, J. W., Booth, R. J., and Francis, M. E. (2007). Liwc2007: Linguistic inquiry and word count. *Austin, Texas: liwc. net*.

[Pennington et al., 2014] Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[Preoţiuc-Pietro et al., 2015] Preoţiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., Schwartz, H. A., and Ungar, L. (2015). The role of personality, age, and gender in tweeting about mental illness. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 21–30.

[Prettenhofer and Stein, 2010] Prettenhofer, P. and Stein, B. (2010). Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1118–1127.

[Reece et al., 2017] Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., and Langer, E. J. (2017). Forecasting the onset and course of mental illness with twitter data. *Scientific reports*, 7(1):13006.

[Rumelhart and McClelland, 1986] Rumelhart, D. E. and McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2. chapter On Learning the Past Tenses of English Verbs, pages 216–271. MIT Press, Cambridge, MA, USA.

[Savard, 2004] Savard, M. (2004). Bridging the communication gap between physicians and their patients with physical symptoms of depression. *Primary care companion to the Journal of clinical psychiatry*, 6(suppl 1):17–24.

[Saxena et al., 2013] Saxena, S., Funk, M., and Chisholm, D. (2013). World health assembly adopts comprehensive mental health action plan 2013–2020. *The Lancet*, 381(9882):1970–1971.

[Sebastiani, 2002] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47.

[Shen et al., 2017] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., and Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844.

[Spitzer et al., 2013] Spitzer, R. L., Md, K. K., and Williams, J. B. W. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Naklada Slap, Jastrebarsko, Croatia.

[Stankevich et al., 2018] Stankevich, M., Isakov, V., Devyatkin, D., and Smirnov, I. (2018). Feature engineering for depression detection in social media. In *ICPRAM*, pages 426–431.

[Tsugawa et al., 2015] Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., and Ohsaki, H. (2015). Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3187–3196. ACM.

[Tsugawa et al., 2013] Tsugawa, S., Mogi, Y., Kikuchi, Y., Kishino, F., Fujita, K., Itoh, Y., and Ohsaki, H. (2013). On estimating depressive tendencies of twitter users utilizing their tweet data. In *2013 IEEE Virtual Reality (VR)*, pages 1–4. IEEE.

[Wei and Pal, 2010] Wei, B. and Pal, C. (2010). Cross lingual adaptation: an experiment on sentiment classifications. In *Proceedings of the ACL 2010 conference short papers*, pages 258–262. Association for Computational Linguistics.

[Wohlgenannt et al., 2016] Wohlgenannt, G., Chernyak, E., and Ilvovsky, D. (2016). Extracting social networks from literary text with word embedding tools. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 18–25.

[Yang et al., 2017] Yang, X., McCreadie, R., Macdonald, C., and Ounis, I. (2017). Transfer learning for multi-language twitter election classification. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 341–348. ACM.

[Yu et al., 2016] Yu, J., Xue, A., Redei, E., and Bagheri, N. (2016). A support vector machine model provides an accurate transcript-level-based diagnostic for major depressive disorder. *Translational psychiatry*, 6(10):e931.

[Yu, 2018] Yu, S. (2018). Uncovering the hidden impacts of inequality on mental health: a global study. *Translational psychiatry*, 8(1):98.