



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

CLASIFICACIÓN DE OPINIONES MEDIANTE
APRENDIZAJE DE MÁQUINAS:
EL CASO DE RESEÑAS SOBRE PELÍCULAS

TESIS

QUE PARA OBTENER EL TÍTULO DE:
INGENIERO EN COMPUTACIÓN

P R E S E N T A:

EDMUNDO PAVEL
SORIANO MORALES

DIRECTOR DE TESIS:
DR. ALFONSO MEDINA URREA



Ciudad Universitaria, México D.F.

2011

Resumen

En esta tesis se realiza un sistema capaz de clasificar en positivas o negativas, y a nivel oración, opiniones sobre películas. Esto se logra con ayuda de técnicas del aprendizaje de máquinas y del procesamiento de lenguaje natural. El sistema trabaja con opiniones sobre películas, obtenidas del sitio web IMDb. Este sistema primero separa los enunciados de las opiniones por polaridad, positivos o negativos, dependiendo de la calificación que los usuarios dieron a la película. Segundo, se realiza una selección de rasgos en los enunciados y, con base en ellos, se agrupan para obtener conjuntos de enunciados subjetivos similares. Tercero, se identifican enunciados subjetivos por la presencia en ellos de adjetivos, adverbios o disparadores de presuposición. Finalmente, se entrena, se prueba y se valida un clasificador bayesiano ingenuo (*naïve Bayes*) usando los enunciados obtenidos. Se alcanzó la mayor exactitud usando enunciados subjetivos identificados por la presencia de adjetivos o adverbios. Se concluye analizando las ventajas y desventajas del sistema y el trabajo futuro a realizar.

Érase una vez, ... en que las únicas inteligencias autónomas que conocíamos los humanos eran los humanos. Entonces pensábamos que si la humanidad creaba otra inteligencia sería resultado de un enorme proyecto, una gran masa de silicio y transistores antiguos y chips y placas de circuitos... una máquina con muchos circuitos interconectados imitaría la forma y función del cerebro humano. Por supuesto, las IAs no evolucionaron de ese modo. Se puede decir que se asomaron a la existencia cuando los humanos mirábamos hacia el otro lado.

Dan Simmons, *El Ascenso de Endymion*

Agradecimientos

A mi madre, Sara Micaela, y a mi padre, José Jelvin, por su esfuerzo constante, esfuerzo físico, mental y económico. Por dedicar sus vidas a mis hermanos y a mí. Por dejar pasar sus mejores años y oportunidades por nosotros.

A mi hermana Sara Isabel y a mi hermano Helvin, por acompañarme toda la vida, por ser constantes y estar ahí en todo momento. Por comenzar, antes que yo, este duro cambio hacia la verdadera vida de adulto.

A Alfonso Medina Urrea, mi director de tesis, por su ayuda y paciencia a lo largo de este difícil camino de plasmar en papel lo que se hace en la práctica. Por apoyarme cuando lo necesité, siempre sin dudar. Por confiar en mí y en mis capacidades.

A Rodrigo Alarcón por darme un poco de su tiempo para compartir sus ideas y asesorarme desde los primeros esbozos de la tesis hasta las últimas etapas. Por intentar inculcar en mí el sentido del orden y de los *deadlines*, que casi nunca cumplí en tiempo y forma. Al final me arrepentí de no haber usado esas buenas prácticas.

Al Dr. Gerardo Sierra, por apoyarme siempre también. Especialmente durante la aplicación a las becas de maestría y durante la recta final de este trabajo. Me quedo con su “todo se puede en esta vida”.

A Teresita Mami por revisar mi tesis, en tiempo récord, y hacer de esta un trabajo más digno. Por ser mi compañera del 12 y por enseñarme las reglas de etiqueta de la alta sociedad.

A todos mis amigos, los que lo fueron y los que lo son, porque nunca me han fallado cuando los he necesitado. Especialmente a Alfredo y al Hobbit.

Al Grupo de Ingeniería Lingüística (GIL) que me brindó la oportunidad de aplicar y expandir mis conocimientos de aprendizaje de máquinas y minería de datos.

A la banda del GIL, en especial a Josué, José Luis y Víctor, por jalar (casi siempre) cuando había que trabajar o cuando había que divertirse o cuando había que trabajar y divertirse. También por hacer bastante entretenido este último año con sus ocurrentes y originales formas de actuar. A Alejandro y a Lázaro por llevar siempre la organización del GIL y, más importante, la organización de los eventos recreativos de fin de semana laboral. A Claudia por su paciencia al enseñarme, con ayuda de Jessi, el maravilloso mundo del baile y sus vueltas. A Azury, que siempre tiene pensamientos positivos a pesar de las adversidades. A Irasema por aguantar mis moditos y ahora brindarme el beneficio de la duda. A Brenda por sus correos *spam* que mantienen informado al GIL de casi todo. A Iria por preocuparse de lo que hago. Al resto de la banda que no menciono, sepan que les agradezco todo lo que han hecho por ayudarme. Estoy seguro de que entrar al GIL fue lo mejor que pude haber hecho en el último año de mi licenciatura. A cada uno de ustedes, gracias por todo.

A mis sinodales: Dr. Miguel Moctezuma, Dr. Alfonso Medina, Dr. Boris Escalante, Dr. Gerardo Sierra y Dr. Juan Manuel Torres, quienes me permitieron concluir este proceso de titulación lo más rápido posible. También al Mtro. Carlos Méndez porque sé que me habría apoyado de la misma forma si hubiera sido mi sinodal.

A la Facultad de Ingeniería, la que me genera una mezcla de tristeza y alegría. Tristeza porque invertí demasiados años en ella. Alegría porque al final, después de varios descalabros tempraneros, entendí que lo mío siempre fue la ingeniería. Comprendí en los últimos años que la

ingeniería mueve al mundo: nos permite ir al Espacio en máquinas increíbles, cruzar ríos tremendos, construir estructuras impensables, nos llevó del telégrafo al Internet y me permite escribir esto en una computadora de dos kilos, guardarlo en una memoria de cinco centímetros de longitud y enviarlo a mi *dropbox* como respaldo.

A los compañeros de la Facultad, conocidos o no, que siempre buscaron aprender más, conocer más y no quedarse sólo con lo que los maestros impartieron. A aquellos que hicieron su verdadero mejor esfuerzo y me obligaron a mejorarme a mí mismo para poder alcanzarlos.

A los profesores, a aquellos que están verdaderamente apasionados por enseñar. Afortunadamente la mayoría.

A mi universidad, la UNAM, por permitir todo esto.

A México y a toda la clase media trabajadora cautiva que pagó por mis estudios.

A todos los couchsurferos que he conocido, provenientes de lugares lejanos: desde Estocolmo hasta Buenos Aires y desde Tampere hasta las Canarias. Casi todos me han dejado un poco de ellos con sus experiencias, anécdotas y sueños. Por demostrarme que querer es poder. Por enseñarme que aunque seamos de lugares tan diferentes y distantes, al final tenemos las mismas cosas adentro.

A Golfi, que con todo y sus golpes, me llevó a mí y a otros cuantos sin fallas ni retrasos.

Muchas gracias.

Aquí les regreso a todos un cachito de lo que me han dado.

Por mi raza hablará el espíritu

Este trabajo se llevó a cabo con el apoyo del CONACyT, en el marco del proyecto *Extracción de conocimiento lexicográfico a partir de textos de Internet* con clave de registro 105711.