



Álgebra Aplicada

Proyecto análisis de sentimientos

Profesor Andrés Abadie

Juan Pablo Cerizola - 5397796-9

León Salvo - 5409486-9

Matías Varela - 5466953-3

Índice

Introducción	3
Marco teórico	3
Análisis de sentimientos:.....	3
Análisis de sentimiento basado en reglas:.....	3
Aplicaciones del análisis de sentimiento:.....	3
Modelos Vectoriales en PLN :	4
Sentimiento Promedio de una Frase :	4
Calidad Promedio de una Frase (Definición de $\text{avg}(w)$):.....	4
Desarrollo	6
Conclusión:	10
Referencias:	11

Introducción

Este práctico tiene como objetivo investigar y trabajar con el análisis de sentimientos de frases utilizando un enfoque basado en vectores. El análisis de sentimientos es una técnica fundamental en el procesamiento del lenguaje natural (PLN) que se utiliza para identificar y cuantificar el tono emocional expresado en un texto. En este caso, nos enfocaremos en cómo transformar una frase en un vector y utilizar esa representación para evaluar su carga emocional.

La metodología de análisis de sentimientos basada en vectores es especialmente útil porque permite traducir la complejidad del lenguaje natural en una forma que las máquinas pueden procesar y analizar de manera eficiente. Esta técnica es ampliamente utilizada en diversas aplicaciones, desde la clasificación de reseñas en sitios web hasta el monitoreo de redes sociales, donde la capacidad de evaluar automáticamente el sentimiento de grandes volúmenes de texto es invaluable.

A lo largo de este práctico, exploramos no solo la implementación de esta técnica, sino también las posibles mejoras que podrían hacerse.

Marco teórico

Análisis de sentimientos:

El análisis de sentimientos es una rama del procesamiento del lenguaje natural (PLN) que se enfoca en la identificación y extracción de opiniones subjetivas dentro de un texto. Este tipo de análisis es ampliamente utilizado en áreas como la minería de opiniones, donde el objetivo es determinar si una revisión, comentario o cualquier fragmento de texto expresa una actitud positiva, negativa o neutral. Según Pang y Lee (2008), el análisis de sentimientos puede realizarse a nivel de documento, oración o aspecto, dependiendo de la granularidad que se desea alcanzar.

Análisis de sentimiento basado en reglas:

En el enfoque basado en reglas (el cual es utilizado para este práctico), el software está entrenado para identificar y clasificar ciertas palabras clave en un bloque de texto. Por ejemplo, pueden clasificarse palabras positivas como “excelente” o “gran”, palabras negativas como “luto” o “muerte”, y palabras neutrales como “pérdida”, esto con el objetivo de buscar en oraciones y/o párrafos completos estas palabras, calcular una puntuación total del sentimiento de cada categoría, calcular la puntuación total y así determinar si un volumen de palabras es categorizado como positivo, neutro o negativo.

Aplicaciones del análisis de sentimiento:

Una de las aplicaciones más destacadas del análisis de sentimientos es el monitoreo de redes sociales. Las empresas y organizaciones utilizan esta técnica para analizar las opiniones y emociones expresadas por los usuarios en plataformas como Twitter, Facebook e Instagram. Esto les permite evaluar la percepción pública de sus productos, servicios o incluso de sus campañas de marketing. Mediante el análisis de grandes volúmenes de datos generados en tiempo real, las empresas pueden identificar rápidamente crisis de reputación, ajustar estrategias de comunicación y mejorar la satisfacción del cliente.

Otra muy común es para evaluar las opiniones de los consumidores sobre un producto o servicio, las empresas pueden obtener información muy importante sobre las preferencias y comportamientos de sus clientes. Este análisis puede extenderse también a la evaluación de la competencia, permitiendo a las organizaciones comparar su desempeño y la percepción de sus productos con los de sus competidores. Esta información es invaluable para la toma de decisiones estratégicas, como el lanzamiento de nuevos productos o la entrada a nuevos mercados.

Modelos Vectoriales en PLN :

Los modelos vectoriales son una técnica fundamental en el procesamiento del lenguaje natural para representar palabras o frases en un espacio multidimensional. Estos modelos permiten transformar texto en datos numéricos que pueden ser procesados mediante algoritmos matemáticos o de aprendizaje automático, facilitando tareas como el análisis de sentimientos.

En el modelo presentado en este proyecto, se utiliza un vector binario " w " para representar la presencia o ausencia de palabras clave específicas en una frase. Cada componente de w indica si una palabra clave está presente (1) o ausente (0) en la frase. Además, con el vector " s ", cuantificamos la cantidad de palabras positivas, neutrales y negativas en la frase.

Sentimiento Promedio de una Frase :

De esta manera, definimos al sentimiento promedio de una frase como la suma de cada una de las palabras clave que contenga (donde las palabras positivas lo aumentan y las negativas lo disminuyen) y dividiéndola por la cantidad de palabras clave que contenga la frase.

Calidad Promedio de una Frase (Definición de $avg(w)$):

En este contexto, la calidad promedio de una frase, representada por $avg(w)$, es una medida que cuantifica la relevancia de una frase en relación con las palabras clave seleccionadas para el análisis. El vector w es un vector binario de dimensión n , donde n es el número total de palabras clave. La calidad promedio $avg(w)$ se define como la suma de los elementos del vector w dividida por n , es decir:

$$avg(w) = \frac{\sum_{i=1}^n w_i}{n}$$

Este valor proporciona una medida de la densidad de palabras clave relevantes dentro de la frase:

- Si $avg(w)$ es cercano a 1, indica que la mayoría de las palabras clave están presentes en la frase, sugiriendo una alta relevancia respecto a los términos seleccionados.
- Si $avg(w)$ es cercano a 0, sugiere que pocas o ninguna de las palabras clave están presentes, indicando una menor relevancia.

Esta medida es crucial en el análisis de sentimientos, ya que permite evaluar la importancia o representatividad de una frase en el contexto específico del análisis. Por ejemplo, frases con una alta calidad promedio pueden tener un mayor impacto en la evaluación del sentimiento general, ya que contienen más términos críticos seleccionados para el análisis.

Desarrollo

1. Las frases a estudiar escogidas fueron:

- 1.1. *“Excelente en su área, su muerte es una enorme pérdida y debería ser luto nacional”*
- 1.2. *“Viví así, solo, nadie con quien poder hablar verdaderamente, hasta cuando hace seis años tuve una avería en el desierto de Sahara”*
- 1.3. *“No comprendí por qué era tan importante para él que los corderos se comieran los arbustos.”*
- 1.4. *“Estoy feliz por el logro y el beneficio obtenido en el trabajo.”*
- 1.5. *“El ambiente fue fantástico y optimista durante toda la reunión.”*
- 1.6. *“Tener una actitud fantástica puede llevarte al éxito y a la felicidad.”*
- 1.7. *“La empresa está en una fase de transición y cambio debido a las circunstancias.”*
- 1.8. *“Es importante analizar cada situación y necesidad de manera moderada.”*
- 1.9. *“El procedimiento fue un simple trámite, sin demasiados detalles.”*
- 1.10. *“Estamos evaluando los detalles y condiciones actuales para tomar decisiones.”*
- 1.11. *“La necesidad de un cambio fue evidente en las circunstancias actuales.”*
- 1.12. *“La noticia de la muerte lo dejó en luto y lleno de sufrimiento.”*
- 1.13. *“Fue un día lleno de sufrimiento, dolor y tristeza.”*
- 1.14. *“El proyecto terminó en un fracaso debido a un error y causó mucho dolor.”*
- 1.15. *“Sentí un profundo miedo por la posible caída y el fracaso del proyecto.”*

2. Palabras clave:

- 2.1. Positivas: excelente, gran, positivo, maravilloso, increíble, feliz, fantástico, espléndido, optimista, alegría, éxito, genial, felicidad, beneficio y logro
- 2.2. Negativas: muerte, luto, fracaso, depresión, problema, triste, negativo, dolor, sufrimiento, miedo, pérdida, caída, daño, error y derrota

2.3. Neutrales: pérdida, cambio, necesidad, circunstancia, situación, espera, moderado, simple, promedio, condición, rutina, transición, detalles, acción, procedimiento

3. Vectores w y s de cada frase:

3.1. *“Excelente en su área, su muerte es una enorme pérdida y debería ser luto nacional”*

w : [0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 1 0 0 0]

s : [1 1 3]

3.2. *“Fue un día maravilloso lleno de éxito y alegría.”*

w : [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]

s : [2 0 0]

3.3. *“El proyecto fue genial y el resultado fue increíble.”*

w : [0
0 0 0 1 0 0 0 0]

s : [1 0 0]

3.4. *“Estoy feliz por el logro y el beneficio obtenido en el trabajo.”*

w : [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0]

s : [3 0 0]

3.5. *“El ambiente fue fantástico y optimista durante toda la reunión.”*

w : [0 1 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]

s : [2 0 0]

3.6. *“Tener una actitud fantástica puede llevarte al éxito y a la felicidad.”*

w : [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]

s : [1 0 0]

3.7. *“La empresa está en una fase de transición y cambio debido a las circunstancias.”*

w : [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]

s : [0 2 0]

- 3.8. “Es importante analizar cada situación y necesidad de manera moderada.”
w: [0 0 1 0
0 1 0 0 0 0 0 0]
s: [0 2 0]
- 3.9. “El procedimiento fue un simple trámite, sin demasiados detalles.”
w: [0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]
s: [0 2 0]
- 3.10. “Estamos evaluando los detalles y condiciones actuales para tomar decisiones.”
w: [0
1 0 0 0 0 0 0 0]
s: [0 1 0]
- 3.11. “La necesidad de un cambio fue evidente en las circunstancias actuales.”
w: [0 1 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0]
s: [0 2 0]
- 3.12. “La noticia de la muerte lo dejó en luto y lleno de sufrimiento.”
w: [0 0 0 1 0
0 0 0 0 1 0 0 0]
s: [0 0 2]
- 3.13. “Fue un día lleno de sufrimiento, dolor y tristeza.”
w: [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0]
s: [0 0 1]
- 3.14. “El proyecto terminó en un fracaso debido a un error y causó mucho dolor.”
w: [0 0 0 0 0 1 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0]
s: [0 0 2]
- 3.15. “Sentí un profundo miedo por la posible caída y el fracaso del proyecto.”

[illegible]

4. Tabla de la calidad promedio, el valor de sentimiento para cada frase:

Frase	Calidad	Positivo	Neutral	Negativo
<i>Excelente en su área, su muerte es una enorme pérdida y debería ser luto nacional</i>	0.09090909 090909091	0.2	0.2	0.6
<i>Fue un día maravilloso lleno de éxito y alegría.</i>	0.04545454 5454545456	1.0	0	0
<i>El proyecto fue genial y el resultado fue increíble.</i>	0.02272727 2727272728	1.0	0	0
<i>Estoy feliz por el logro y el beneficio obtenido en el trabajo.</i>	0.06818181 818181818	1.0	0	0
<i>El ambiente fue fantástico y optimista durante toda la reunión.</i>	0.04545454 5454545456	1.0	0	0
<i>Tener una actitud fantástica puede llevarte al éxito y a la felicidad.</i>	0.02272727 2727272728	1.0	0	0
<i>La empresa está en una fase de transición y cambio debido a las circunstancias.</i>	0.04545454 5454545456	0	1.0	0
<i>Es importante analizar cada situación y necesidad de manera moderada.</i>	0.04545454 5454545456	0	1.0	0
<i>El procedimiento fue un simple trámite, sin demasiados detalles.</i>	0.04545454 5454545456	0	1.0	0
<i>Estamos evaluando los detalles y condiciones actuales para tomar decisiones.</i>	0.02272727 2727272728	0	1.0	0
<i>La necesidad de un cambio fue evidente en las circunstancias actuales.</i>	0.04545454 5454545456	0	1.0	0

<i>La noticia de la muerte lo dejó en luto y lleno de sufrimiento.</i>	0.04545454 5454545456	0	0	1.0
<i>Fue un día lleno de sufrimiento, dolor y tristeza.</i>	0.02272727 2727272728	0	0	1.0
<i>El proyecto terminó en un fracaso debido a un error y causó mucho dolor</i>	0.04545454 5454545456	0	0	1.0
<i>Sentí un profundo miedo por la posible caída y el fracaso del proyecto.</i>	0.06818181 818181818	0	0	1.0

5. La frase más positiva es: *“Estoy feliz por el logro y el beneficio obtenido en el trabajo.”*, con un valor positivo de 3.

6. La frase más negativa es: *“Sentí un profundo miedo por la posible caída y el fracaso del proyecto.”*, con un valor negativo de 3.

7. La calidad promedio representa la cantidad de palabras clave que se encuentran presentes en la frase. Si la frase contiene pocas palabras clave, nos es poco útil el análisis de sentimiento de esa frase en particular. Cuantas más palabras clave contenga la frase, más relevante será la frase y el sentimiento general de la misma.

La definición matemática tiene sentido porque al dividir la cantidad de palabras clave contenidas entre la cantidad de palabras clave totales se obtiene la razón entre estas dos cantidades, que es, como se explicó anteriormente, un índice de qué tan relevante es para nosotros el sentimiento que se atribuya a esta oración particular y nos permite ponderar adecuadamente al compararlo con el sentimiento de otras frases.

8. El promedio de sentimiento representa, como su nombre lo indica, el sentimiento general de una frase, y se calcula como la razón entre palabras clave positivas, neutras y negativas que esta contenga. De esta manera, el sentimiento de una frase es independiente de la cantidad de palabras que contenga.

9. La selección inicial de las palabras clave es un componente crítico en el desempeño de este algoritmo de análisis de sentimientos. La calidad de las palabras clave elegidas determinan directamente la precisión y la utilidad de los resultados obtenidos. Si se seleccionan palabras clave que no representan adecuadamente los sentimientos o emociones que se desean analizar, el algoritmo podría generar clasificaciones incorrectas o irrelevantes. Esto podría llevar a interpretaciones erróneas del contenido, haciendo que los resultados sean poco útiles o incluso engañosos en el contexto de la aplicación deseada. Una selección cuidadosa y bien informada de palabras clave permite que el algoritmo capture de manera más precisa las sutilezas del texto, mejorando significativamente su capacidad para clasificar frases o párrafos según su sentimiento.

10) Se podría mejorar el algoritmo tratando mejor los textos, por ejemplo pasarlos por un preprocesado que saque las palabras ambiguas o que no benefician al estudio de las frases como “y” o “el”. Además también pensamos en que otra buena forma es crear una especie de contexto de palabras ya que hay palabras las cuales dependiendo del contexto pueden ser negativas o positivas. por ejemplo con la palabra “pérdida”, una frase negativa podría ser “Su pérdida fue muy dolorosa para ella” mientras que por otro lado una forma positiva podría ser “Su pérdida de peso ayudo con los dolores de espalda”

Conclusión:

Al transformar frases en vectores se obtiene una medida objetiva de cuán positiva o negativa es una frase, esta es una muy buena forma de conseguir analizar y clasificar frases de cualquier ámbito dado un contexto.

El análisis de sentimientos mediante vectores es una técnica útil para cuantificar y evaluar el tono general de una frase vista en el práctico, ya que se podría mejorar como por ejemplo dando contexto a las frases y a las palabras.

El uso de vectores en este tipo de análisis es una herramienta valiosa que sigue evolucionando y ofreciendo nuevas posibilidades en el campo del análisis de datos

computacional y el procesamiento del lenguaje natural.

Referencias:

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.

¿Qué es el análisis de sentimiento? | IBM. (n.d.). <https://www.ibm.com/es-es/topics/sentiment-analysis>

colaboradores de Wikipedia. (2024, August 31). *Vector*. Wikipedia, La Enciclopedia Libre.

<https://es.wikipedia.org/wiki/Vector>