

Report

Right wing Opinion leader's influences on followers. a Sentiment analysis from Youtube Comments

Francisco Villarroel-Riquelme - María Paz Raveau

Objetivos y motivación

La idea principal de este proyecto es observar cómo los líderes de opinión generan contagio emocional en sus seguidores, así como comprender la estructura de interacciones que los líderes generan dentro de su público que recibe los mensajes.

Métodos

Se eligieron distintos líderes de opinión de medios de comunicación alternativos y tradicionales, que crean videos de columnas de opinión; De esta forma podemos analizar el contenido emocional del mensaje emitido del líder manteniendo otras características (música, fondo, imágenes, etc) constantes para enfocarnos sólo en el texto.

Además, utilizaremos los comentarios de youtube para analizar la carga emocional y la relación de ella con los videos realizados con los líderes de opinión. La idea es armar la red de comentarios, analizar la carga emocional de video vs comentarios, y analizar en el tiempo si hay nodos que se repiten cómo ese link se intensifica.

Tipos de análisis de sentimiento

Una comparación de métodos de S.A hechos por Van Atteveldt y otros investigadores (2020) señala que:

El análisis de sentimiento se usa en varios contextos, como por ejemplo estudiar la calidad del contacto intergrupar, efectos de medios hostiles, cobertura de guerras, violencia y agresión, comentarios de usuarios en base a esos artículos; diferencias de género en reportaje, etc.

El problema es que el sentimiento se expresa con un lenguaje muy creativo y ambiguo.

Las aproximaciones computacionales pueden ser una buena alternativa a modelos manuales de etiquetado pero tienen una validez muy baja. Por ejemplo métodos de diccionario caen porque están hechos para tareas específicas, y las personas no revalidan los modelos para las tareas que piensan usarlo, llevando a conclusiones erróneas (ej: usar diccionarios para calificar sentimiento de reviews de películas). Tanto diccionario esconde el problema de validez.

Este paper usa triple codificación para analizar cuál método de análisis de sentimiento es mejor.

Los resultados no son muy alentadores: valen callampa. sobre todo los métodos de diccionario. Sin embargo los de deep learning son bastante mejores, pero aún así no tiene el nivel generalmente requerido para análisis de texto. Aún es necesario chequeo humano para análisis de sentimiento

Métodos de análisis de sentimiento

La forma común de medir sentimiento es con anotación manual y libro de código, pero es caro y no siempre se logran buenos niveles de fiabilidad. Otra solución es el crowd coding en vez de sólo un experto, lo que aumenta su validez ya que muchas personas evalúan el mismo texto, dándole más fiabilidad y también mostrando la ambigüedad de la emoción.

Las otras formas de análisis es de automatizado. Se pueden calificar como positivos, negativos o neutros. hay dos grandes ramas para esto: **1) Diccionarios y 2) Machine Learning.**

Los métodos de diccionario se usa una lista de palabras y un modelo entrenado para buscar las valencias emocionales de esa lista de palabras, sin embargo la construcción de esa lista es más compleja de lo que parece y nos lleva a conclusiones divergentes.

En el caso del machine learning, este puede captar algunas sutilezas del texto. En esta técnica, en vez de tener diccionarios, los computadores usan manualmente datos entrenados codificados para aprender el link entre el input y el output deseado. Este método es mucho mejor que diccionarios y la mayoría de los trabajos actuales

Antes usaban muchos modelos BOW pero no consideraba el orden, y ahora se usan modelos de deep learning y redes neuronales combinados con word embeddings. Estos modelos usan grandes cantidades de texto sin etiquetar para aprender los significados completos de una palabra. "esto reduce el número de características de inputs y permite capturar significados de palabras que no estaban presentes en los datos iniciales."

Importante señalar es que en los métodos automatizados, siempre decae en fiabilidad cuando se usan en dominios diferentes para los cuales fueron usados.

Método

Crown coding

Se hizo un crown coding de 5 personas por línea, se pidió calificar si la información tenía que ver con economía, y si era negativo, positivo, neutro/ambiguo. Se calificaron 300 frases. Se les dio 70 frases con el resultado para que las vieran como ejemplo a ver si entendieron las instrucciones.

Diccionarios

Se usaron los diccionarios de Boukes et. al (2020) y se agregaron diccionarios específicos para "esperanza" y "miedo" en contextos económicos. Se aplicaron los diccionarios ANEW Dictionary, Pattern y Polyglot en Python. En R se usó un diccionario de Damstra y Boukes (2018), el NRC Emotion lexicon y el diccionario customizado con la aproximación sugerida por Middiman et. al (2019).

También se tradujo con Google Translate y DeepL los textos al inglés para usar todos los diccionarios de ese lenguaje. En R se usó el Affective Norms for English words (AFINN), que originalmente es polar (-5/+5) pero en R es binario. También se usó el Augmented General Inquirer positive and negative dictionary, y un diccionario de Hu y Liu (2014) Y UN MONTÓN DE OTROS DICCIONARIOS QUE ME DA LATA SUMARIAR

Machine learning

Se probó dos métodos "clásicos": Naive Bayes (NB) y Support Vector Machine (SVM); para deep learning se usaron redes neuronales Convolucionales. Para entrenar los modelos se usaron 6038 titulares manualmente codificados por Boukes et. al (2020).

Para el NB y SVM se usó scikit-learn y ponderación tf-idf. Para entrenar el CNN se usaron embeddings usando el Amsterdam Embeddings Model hecho con noticias holandesas.

Resultado principal

Table 2. Overall performance of the tested sentiment analysis approaches.

section	name	Acc.	alpha	Positive			Neutral			Negative		
				Pr.	Re.	F1	Pr.	Re.	F1	Pr.	Re.	F1
Manual Coding	Single Coder	0.82	0.82	0.88	0.86	0.87	0.76	0.81	0.78	0.84	0.80	0.82
Manual Coding	Vote (3 Coders)	0.88	0.90	0.97	0.91	0.94	0.82	0.88	0.85	0.87	0.84	0.86
Crowd-Coding	Single Coder	0.72	0.75	0.69	0.84	0.76	0.69	0.58	0.63	0.78	0.78	0.78
Crowd-Coding	Vote (3 Coders)	0.77	0.81	0.73	0.89	0.80	0.74	0.65	0.69	0.83	0.81	0.82
Crowd-Coding	Vote (5 Coders)	0.77	0.81	0.73	0.90	0.81	0.73	0.65	0.69	0.84	0.80	0.82
Machine Learning	CNN	0.63	0.50	0.68	0.49	0.56	0.58	0.78	0.66	0.72	0.57	0.63
Machine Learning	NB	0.58	0.39	0.74	0.34	0.47	0.52	0.83	0.64	0.65	0.47	0.55
Machine Learning	SVM	0.57	0.41	0.69	0.37	0.48	0.52	0.79	0.62	0.64	0.48	0.55
Dictionaries	DANEW	0.42	0.10	0.75	0.08	0.15	0.40	0.97	0.57	0.80	0.04	0.08
Dictionaries	DamstraBoukes	0.41	0.05	0.83	0.07	0.13	0.40	0.99	0.57	0.00	0.00	0.00
Dictionaries	Muddiman	0.49	0.31	0.53	0.38	0.44	0.46	0.64	0.53	0.53	0.39	0.45
Dictionaries	NRC	0.47	0.32	0.39	0.53	0.45	0.46	0.44	0.45	0.59	0.46	0.52
Dictionaries	Pattern	0.39	0.07	0.43	0.08	0.14	0.39	0.90	0.54	0.38	0.03	0.06
Dictionaries	Polyglot	0.42	0.26	0.38	0.32	0.34	0.39	0.55	0.45	0.53	0.33	0.41
English Dictionaries	AFINN	0.43	0.27	0.35	0.38	0.37	0.40	0.50	0.45	0.58	0.38	0.46
English Dictionaries	DamstraBoukes	0.42	0.07	0.67	0.08	0.15	0.40	0.98	0.57	1.00	0.02	0.04
English Dictionaries	GenInq	0.41	0.26	0.31	0.37	0.34	0.38	0.38	0.38	0.54	0.47	0.51
English Dictionaries	HuLiu	0.46	0.34	0.40	0.30	0.34	0.42	0.63	0.50	0.65	0.40	0.50
English Dictionaries	LoughranMcDonald	0.50	0.29	0.50	0.14	0.22	0.46	0.79	0.58	0.62	0.43	0.51
English Dictionaries	LSD	0.46	0.33	0.39	0.40	0.39	0.42	0.54	0.48	0.62	0.41	0.50
English Dictionaries	Muddiman	0.48	0.27	0.48	0.38	0.43	0.46	0.71	0.55	0.57	0.30	0.39
English Dictionaries	NRC	0.42	0.23	0.34	0.62	0.44	0.43	0.32	0.37	0.57	0.39	0.46
English Dictionaries	RID	0.42	0.06	0.00	0.00	0.00	0.41	0.97	0.57	0.82	0.09	0.16

Acc.: Accuracy expressed as percentage correct; α : Krippendorff's alpha (ordinal), Pr.: Precision for the specified class (Positive/Neutral/Negative); Re.: Recall for that class; F1: F1-Score for that class. When more than one measurement or prediction (respectively for manual/crowd annotation and machine learning) was available, we averaged the score.

La interpretación de los autores indica que en el caso de CNN y SVM y los modelos automatizados tienen α muy bajos, por lo que no podrían ser usados directamente para análisis sustantivos. Para qué hablar de las traducciones al inglés para usar otros diccionarios.

Además, la correlación entre diccionarios es bien poca, por lo que podemos entender que todos están midiendo cosas diferentes.

Si bien el CNN es de los mejores, falla en cosas que hacen desistir, como por ejemplo calificar como positivo el "más bancarrotas", aunque para otras palabras es más difícil entender por qué se equivocó.

Recomendaciones para clasificación de textos

1. Formalizar la conceptualización y operacionalización para anotaciones manuales de la cantidad de interés. Este paso es de extremo importante y a menudo requiere codificación de pilotaje de materiales y discusión entre investigaciones
2. Califiquen una muestra lo suficientemente grande para el gold standard. al menos dos anotadores para la fiabilidad intercódigo. Entre 100 a 300 unidades son requeridas.

3. Utilice cualquier diccionario off-the-shelf. Si el diccionario logra una fiabilidad similar al gold standard, usa el diccionario porque es más transparente.
4. Si no es el caso, considere crear un diccionario o customizar uno hecho. Se puede hacer enlistando las palabras más sorprendentes del corpus que se utilizará.
5. Si el diccionario customizado no funciona, las opciones que quedan son codificación humana o machine learning. Codifique al menos mil artículos con crown coding.
6. Entrene un modelo de machine learning usando los documentos codificados. Use *cross validation* para hacer una primera validación y después usar otros modelos. Si el modelo está bueno, entrenelo de nuevo usando toda la data codificada y valide contra el gold standard
7. Si es necesario, repita el paso 5 hasta que el modelo sea suficientemente válido o con suficientes unidades que hayan sido codificadas.

Con esta guía se asegura que los resultados sean válidos y replicables con un mínimo de codificación manual necesario.

Finalmente, el problema principal es que no se tienen marcos teóricos compartidos de lo que es un "sentimiento"

Otro paper del mismo año de Mehta y Pandya (2020)

Según este autor, Sentiment analysis es lo mismo de "opinion mining". El análisis de sentimiento es básicamente la percepción de si lo que se dice es positivo o negativo sobre algo (un review de películas, una persona, un partido político, una política, o un review de las características de un producto).

¿Qué es un sentimiento y el análisis de sentimiento?

Los sentimientos se pueden expresar a través de distintas emociones (!?), pasando por juicios, visiones, insights o visiones de las personas. Esto puede ser analizado de dos formas distintas:

1. Como el impacto del escritor: Las formas en que elige las palabras para expresar emociones,

2. El impacto en el lector: Cómo esa persona interpreta el contenido escrito dependiendo de su estado mental y su habilidad para analizar cosas