

Análisis de sentimiento sobre textos en Español basado en aproximaciones semánticas con reglas lingüísticas

Sentiment analysis of texts in spanish based on semantic approaches with linguistic rules

Roberto Hernández Petlachi y Xiaoou Li

Department of Computer Science

CINVESTAV-IPN

Av. IPN 2508, col. Zacatenco

C.P. 07360, México DF

rfhernandez@computacion.cs.cinvestav.mx, lixo@cs.cinvestav.mx

Resumen: La aparición de blogs, foros y redes sociales en internet, donde hay una multitud de usuarios expresando opiniones sobre una gran variedad de temas; ha despertado el interés del ámbito empresarial. Se ve en este tipo de recursos una oportunidad de conocer como sus productos son percibidos por los consumidores, es por ello que se han requerido de soluciones para analizar y monitorizar estas opiniones. Nuestra propuesta está basada en aproximaciones semánticas con reglas lingüísticas para la clasificación de polaridad de textos en español. La clasificación de la polaridad en las palabras se hace de acuerdo a un diccionario de orientación semántica donde cada término se encuentra marcado con un valor de uso y valor emocional, junto con reglas lingüísticas para resolver varias contrucciones que podrían afectar la polaridad del texto.

Palabras clave: *análisis de sentimiento, freeling, minería de opinión, clasificación de la polaridad.*

Abstract: The emergence of blogs, forums and social networks sites where there is a multitude of users expressing opinions on a wide variety of topics; has aroused the interest of the business world. We see in this type of resources a chance to meet their products are perceived by consumers, is why solutions are required to analyze and monitor these reviews. Our proposal is based on semantic approaches with linguistic rules for classifying polarity texts in Spanish. Polarity classification in the words is done according to a dictionary of semantic orientation where each term is labeled with a use value and emotional value, along with linguistic rules to solve various constructions that could affect the polarity of text.

Keywords: *sentiment analysis, freeling, Opinion Mining, polarity classification.*

1 Introducción

“Lo que otros piensan” ha sido una pieza importante de información para la mayoría de nosotros durante el proceso de toma de decisiones. Mucho antes de que la conciencia de la World Wide Web se generalizara, muchos de nosotros pedimos a nuestros amigos que nos recomiendan un buen electricista, pedir cartas de referencia con respecto a los solicitantes de empleo a los colegas, o consultar los comentarios de los consumidores para decidir qué producto o servicio comprar. Con la explosión de las plataformas Web 2.0, tales como blogs, foros de discusión y otros tipos de medios de comunicación social, los consumidores

tienen a su disposición una gran variedad de herramientas al alcance sin precedentes para poder compartir sus experiencias de marca y opiniones (positivas o negativas), en relación con cualquier producto o servicio. Como las grandes empresas están creciendo cada vez más empiezan a darse cuenta, que las voces de los consumidores pueden ejercer una enorme influencia en la formación de las opiniones de otros consumidores y, en última instancia, su fidelidad de marca, sus decisiones de compra, y su defensa de la marca. Las empresas pueden responder a las opiniones de los consumidores que se generan a través del monitoreo y análisis de medios sociales mediante la modificación de

sus mensajes de marketing, posicionamiento de marca, desarrollo de productos, y otras actividades en consecuencia.

En este artículo presentamos una clasificación de la polaridad con aproximaciones que combinan el conocimiento semántico con reglas lingüísticas para textos en español. Finalmente debemos destacar el desafío que significa trabajar con textos espontáneos en español, plagados de errores ortográficos y cierto desapego a las normas gramaticales especialmente en la puntuación.

El artículo se organiza de la siguiente manera. En la sección 2 se revisan brevemente los antecedentes referentes a la clasificación de polaridad, en la sección 3 detallamos nuestra solución. En sección 4 se ilustran los resultados experimentales. Finalmente en la sección 5 se muestran las conclusiones y trabajos a futuro.

2 Antecedentes

El análisis de sentimiento en el procesamiento del lenguaje natural se refiere, a la disciplina que comprende la tarea de identificar-clasificar fragmentos de texto que contengan emotiva opinión o subjetiva. Las tareas que se pueden englobar en la disciplina, cabe destacar la detección de polaridad donde los textos dados son clasificados y analizados según la carga emotiva que presenten ya sea positiva o negativa; la detección de la subjetividad cuyo objetivo es la identificación de fragmentos de textos que poseen un significado o una carga subjetiva expresada por parte del autor.

Nuestra propuesta se enfoca principalmente a la clasificación de la polaridad en textos en español. El uso del término sentimiento en relación al análisis de textos apareció por primera vez en los trabajos de (Das y Chen, 2001) y (Tong, 2001) sobre la predicción de juicios para analizar el comportamiento de los mercados. Solamente un año después, el término análisis de sentimiento fue acuñado por (Turney, 2002) en referencia a sus aproximaciones para la clasificación de textos en función de su polaridad, donde se presenta una fórmula matemática para el cálculo de lo que Turney denominó Orientación Semántica. (Zhang et al., 2011) proponen un método supervisado para el análisis de textos cortos (Twitter) que es entrenado con los datos proporcionados por un analiza-

dor semántico base no supervisado. (Bakliwal et al., 2012) presenta un método de puntuación del sentimiento no supervisado y comparan su propuesta con un método supervisado. Por otro lado, Martín-Wanton y Carrillo de Albornoz (2012) proponen un método basado en diccionarios afectivos y WordNet. Trabajos recientes sobre el análisis de español son los trabajos de (Vilares, Alonso, y Gómez-Rodríguez, 2013) donde hacen una clasificación de la polaridad mediante análisis sintáctico de dependencias que les brinda mejores resultados con respecto a los sistemas puramente léxicos. La mayoría de los trabajos recientes acerca de la clasificación de la polaridad se enfocan en el idioma inglés, dándole poca importancia al español. En este último lenguaje el sistema más relevante es SODictionariesV1.11Spa realizado por (Brooke, Tofiloski, y Taboada, 2009) . Donde este resuelve la orientación semántica a nivel individual en sustantivos, verbos, adjetivos y adverbios; trata modificadores de la polaridad como son los intensificadores y la negación. La manera de tratar todas estas construcciones lingüísticas a nivel léxico, con lo cual el SODictionariesV1.11Spa nos ayuda a resolver estas construcciones.

La negación en el análisis de sentimiento ha sido estudiada en los últimos años, los trabajos tradicionales utilizan métodos supervisados que modelan el texto como un conjunto de frecuencia de palabras asignándole un valor binario. (Choi y Cardie, 2008) su hipótesis es asumir que las palabras dentro de una expresión interactúan unos con otros para determinar la polaridad total de la expresión. (Yang et al., 2010) considera la negación como los términos situados a la derecha. Otro trabajo relacionado (Taboada et al., 2011), utiliza información morfológica para identificar la negación.

El uso del lenguaje en plataformas de redes sociales en internet presenta un desafío interesante. Al tratarse de escritura espontánea y creativa con marcados rasgos de oralidad, la ortografía resulta ser sumamente variable, debemos abarcar no solo una corrección ortográfica si no también de caracteres repetidos, palabras abreviadas y de signos de puntuación, haremos uso el algoritmo de Levenshtein, con sus operaciones de edición (inserción, borrado y sustitución) para la corrección ortográfica.

3 Solución

Nuestra solución se basa en aproximaciones semánticas, etiquetación morfológica y orientación semántica con métodos supervisados, para así determinar la polaridad del texto. Como primer paso pre procesamos, seguido de la lematización entrenando las frases en la herramienta *Freeling*, tokenización y segmentación, posteriormente la etiquetación, una vez teniendo estos pasos, aplicamos las reglas lingüísticas para obtener la clasificación de polaridad. El sistema ha sido preparado para trabajar con el idioma Espa nol.

3.1 Preprocesamiento

- *Corrección de signos de puntuación:* es importante hacer resaltar que los signos de puntuación en el análisis de sentimiento son muy importantes debido a que si se colocan de manera incorrecta, pueden afectar negativamente al resto del procesado (por ejemplo: “Hasta la prox . VERACRUZ ... Gracias ... ”)
- *Normalización de URL:* las direcciones web presentes en un tuit son sustituidas por la cadena “enlace”.
- *Reemplazo de emoticonos:* muchas de las herramientas existentes en internet cuentan con un catálogo de emoticonos que ayudan a los usuarios a expresar mejor lo que están sintiendo al momento de escribir sus opiniones. Para poder usar esta información creamos una bolsa de emoticones más usuales en los textos y su respectivo significado, y reemplazamos esta cadena de caracteres por una cadena que pueda ser buscada en el diccionario de orientación semántica y nos ayude a definir la polaridad de cada uno de los textos que contenga dicha característica.
- *Corrección de abreviaturas:* se sustituyen algunas de las palabras abreviadas más habituales por su forma gramatical reconocida (por ejemplo: “q” → “que”, “xq” → “porque”).
- *Corrección de palabras:* para la corrección ortográfica utilizamos el algoritmo de Levenshtein con su noción de distancia. El cual calcula la distancia entre dos términos como la mínima cantidad de operaciones de edición (borrado, inserción y sustitución) para corregir un

término al candidato generado por el algoritmo. Para corregir las palabras se hace uso de un diccionario de palabras, el cual se compone de la lista completa de formas del Corpus de Referencia del Español Actual (CREA) de la Real Academia Española, con frecuencias de uso y con las formas conjugadas más usadas, aproximadamente son 128 000 formas. Si una palabra no es encontrada en nuestro diccionario, el algoritmo tomara la palabra más cercana con distancia 1, y la reemplazara.

- *Corrección de caracteres repetidos:* especialmente en el caso de vocales, reemplazamos la repetición de la misma concurrencia por una sola, con excepción de *cc,rr,ll*. Una vez que se obtiene el texto limpio procedemos a realizar la lematización de las palabras para obtener su lema sin conjugación, junto con la tokenización y la segmentación de las oraciones para así poder clasificar la polaridad (por ejemplo: largooooo → largo”).

3.2 Identificación de conceptos con carga emocional

Una vez el sistema ha determinado todos los conceptos de cada oración, el siguiente paso consiste en traducir cada concepto identificado a su correspondiente carga emocional con un léxico afectivo, en el caso de que el concepto tenga una categoría emocional asociada. Para realizar este paso se precisa, por tanto, de un léxico afectivo llamado SODictionariesV1.11Spa de orientación semántica, este diccionario almacena valores que reflejan el sentimiento de una colección de palabras de distintas categorías gramaticales.

Para obtener la polaridad de los textos se genera una sumatoria de todas las palabras de la opinión que se encuentren en los diccionarios como se muestra en la ecuación 1.

$$OS_{texto} = \sum_{i=1}^N Palabras \quad (1)$$

3.3 Intensificadores en el análisis de sentimientos

Identificar la intensificación permite conocer cuáles son los puntos que intención de destacar una idea. Los intensificadores son un grupo de modificadores emocionales a tener en

cuenta cuando se plantea desarrollar un sistema de análisis de sentimientos. Los intensificadores se definen como términos capaces de ampliar o disminuir la intensidad emocional del texto al que afectan.

Para realizar esta tarea seguimos el enfoque de (Taboada et al., 2011), donde cada cuantificador lleva asociado un porcentaje que varía según la intensidad con la que aumente o disminuya el texto al que afecta. La intensificación basada en porcentajes propone que cada intensificador tenga un porcentaje asociado, positivo si es amplificador y negativo si es decrementador. Esto permite que las oraciones semánticas se modifiquen de forma que se tenga en cuenta las polaridades de los elementos involucrados.

El mecanismo seguido en el sistema para la detección de intensificadores en los textos consiste en la comparación de los términos de cada oración con todas las entradas de la lista de intensificadores proporcionada por el diccionario SODictionariesV1.11Spa. Cuando el sistema detecta una coincidencia, comprueba a su vez que el rol gramatical del término identificado es el adecuado, y si es así, éste es marcado como intensificador. Posteriormente se le suma el porcentaje ya sea para incrementar o decrementar la palabra, y este valor es $Int_{intensificador}$ es la palabra encontrada en nuestro diccionario de intensificadores, que servirá como incrementador o decrementador, multiplicado por la palabra que afecta, el resultado final obtenido se suma junto con los demás términos encontrados para obtener la polaridad final del texto.

3.4 Negación en el análisis de sentimientos

La detección de la negación es un proceso complejo que no solo abarca la identificación de las palabras o términos que expresan la negación, sino también su ámbito de acción; es decir, las palabras afectadas por dicha negación.

La negación es una de las construcciones lingüísticas que más influye en el análisis de polaridad de los textos. La negación es una construcción habitual a la hora de expresar una opinión. Una forma de negar una idea en el español es mediante el término “no”, aunque existen otros negadores como “tampoco”, “sin.” “nunca” que son también utilizados con frecuencia.

La estructura sintáctica para la negación

“no.es prácticamente idéntica en todas las situaciones posibles. La única diferencia es el tipo de dependencia con el que se anota cuando el negador depende de un verbo en vez de cualquier otro tipo de construcción.

Para resolver la negación en vez de invertir la polaridad de la palabra afectada por negación, modificamos su valor en una cantidad fija de signo contrario al sentimiento de la palabra afectada por la negación, con lo que se consigue un enfoque más real de la negación. Tomando en cuenta que las palabras fuertemente negativas tienen un valor de -5 , tomamos el mismo valor para poder generar la polaridad con palabras que contengan esta característica. Esto es para cada palabra positiva afectada por la negación se tomara a dicha palabra y se le restara el valor de 5, la palabra con valoración negativa tendrá una suma de 5 (por ejemplo: horrible -4 , No está horrible = 1, genial 4, No está genial = -1 , querer = 2, No quiero = -3).

Para esto construimos un diccionario propio que contienen las formas posibles en el idioma español de negar un hecho como son :

- La negación se produce usualmente anteponiendo al verbo el adverbio negativo “no”: [No] tengo hambre. Incluso hay que tomar en cuenta el alcance de la negación para verificar que palabra esta afectando, por ejemplo: No está mal.
- También mediante otros adverbios, por ejemplo:
Nunca: **Nunca** lo dije.
Nada: **Nada** traje.
Tampoco: **Tampoco** está mal.
- Con la utilización de conjunciones negativas “ni”:
[Ni] compraré hoy [ni] compraré mañana.

4 Resultados experimentales

4.1 Datos de entrenamiento

El conjunto de datos de entrenamiento C_{train} proporcionado por la organización TASS 2014 consiste en 7,219 mensajes de Twitter. Cada tweet esta etiquetado con una polaridad global, indicando si en el texto se expresa un sentimiento positivo, neutral, negativo o ningún sentimiento. 6 niveles han sido definidos: fuertemente positivo (P+), positivo

(P), neutral (NEU), negativo (N), fuertemente negativo (N+) and sin sentimiento (NONE) como se muestra en la tabla 1. El número de tweets correspondiente a P+ y NONE son los que continene mas alto porcentaje en el corpus.

| Sentimiento | % # Tweets | # de Tweets |
|-------------|------------|-------------|
| P+ | 22.44 % | 1619 |
| P | 4.12 % | 297 |
| NEU | 8.45 % | 610 |
| N | 16.91 % | 1220 |
| N+ | 12.51 % | 903 |
| NONE | 23.58 % | 1846 |

Tabla 1: Distribución por clases de polaridad en corpus train C_{train}

Este corpus fue utilizado para realizar el entrenamiento necesario para probar nuestra solución.

4.2 Evaluación

Nuestra solución fue evaluada utilizando el corpus test C_{test} de TASS 2014 que consiste en 60,798 mensajes de Twitter. Cada tweet esta etiquetado con una polaridad global, indicando si en el texto se expresa un sentimiento positivo, neutral, negativo o ningún sentimiento. Se ha realizado experimentos para las dos tareas de clasificación de polaridad: cuatros categorías (Positivo \rightarrow P, Neutral \rightarrow NEU, Negativo \rightarrow N, NONE) y seis categorías (Fuertemente Positivo \rightarrow P+, Positivo \rightarrow P, Neutral \rightarrow NEU, Negativo \rightarrow N, Fuertemente Negativo \rightarrow N+, NONE) como se muestra en la tabla 2. Los resultados se evalúan siguiendo las métricas de precisión, recall y la medida F. En el cuadro 3 se incluyen los resultados para las dos categorizaciones (3 y 5 niveles).

| Sentimiento | % # Tweets | # de Tweets |
|-------------|------------|-------------|
| P+ | 34.12 % | 20,745 |
| P | 2.45 % | 1,488 |
| NEU | 2.15 % | 1,305 |
| N | 18.56 % | 11,287 |
| N+ | 7.5 % | 4557 |
| NONE | 35.22 % | 21,416 |

Tabla 2: Distribución por clases de polaridad en corpus C_{test}

| | Presición | Recall | Medida F |
|-----------|-----------|--------|----------|
| 3 Niveles | 0.6192 | 1 | 0.7648 |
| 5 Niveles | 0.4558 | 1 | 0.6261 |

Tabla 3: Resultados obtenidos sobre la evaluación del conjunto de test.

Los valores más altos en el diccionario de orientación semántica son: fuertemente positivo +5 y fuertemente negativo -5, dentro de los cuales tomamos los rangos para generar los resultados de las categorías correspondientes, donde podemos notar que en la clasificación de los tweets neutros se alcanza un bajo rendimiento, se puede deber a muchas razones una de ellas es la mezcla de ideas a favor y en contra que dificultan la clasificación de estos textos.

Para poder obtener los resultados utilizamos el rango antes mencionado y mostrado en la ecuación 2, donde para cada categoría los valores fueron tomados de la siguiente manera:

$$X = -5 \dots 5 \quad (2)$$

- Para valores positivos (P) el rango toma fue $X \geq 1$
- Para valores negativos (N) el rango toma fue $X \leq -1$
- Para valores neutrales (NEU) el rango toma fue $-1 < X < 1$
- Para valores que no contienen sentimientos, fue necesario declara una variable que indicara la carencia de esta característica para poder clasificarlos con (NONE).

Analizando a detalle los resultados llegamos a la conclusión que tenemos errores en la subjetividad de los tweets y además en las reglas lingüísticas tomadas para realizar la negación de los mismos. Además de que los valores de precisión son más bajos en los casos de valores neutrales.

5 Conclusiones y trabajo futuro

Es bien sabido que la importancia del análisis de sentimientos se ha visto incrementada debido a la inmensa cantidad de información subjetiva disponible. Es por ello que en este

artículo presentamos una propuesta que emplea conocimiento semántico y lingüístico para detectar el sentimiento de los tweets escritos en español. Los resultados experimentales muestran un buen rendimiento y sugieren que la estructura morfosintáctica de los textos junto con reglas lingüísticas es muy útil para detectar la polaridad de dichos textos. En definitiva, estamos convencidos de que el análisis de sentimiento es un muy reciente campo de aplicaciones práctica de la investigación académica con una clara relación con la industria. En los trabajos a futuro sería interesante integrar en el análisis los signos de interrogación y las risas para tener un acercamiento más real al sentimiento que expresa el usuario, además de rediseñar las reglas lingüísticas para la negación, donde una de las cuestiones mejor estudiadas de la negación es la negación lógica mediante medios sintácticos, atendido al orden. Entre las lenguas del mundo cuatro posibilidades para el orden relativo del sujeto, del objeto y del verbo mezclado con la negación, siendo los más frecuentes por orden: SOV-Neg, S-Neg-VO, Neg-VSO, Neg-VOS. También se deben considerar las herramientas disponibles de la plataforma como el uso de hashtags que en nuestra solución no son tomados en cuenta y que en ocasiones son el único indicador de sentimiento indicado en el tweet: #FelicesFiestas, #muchassonrisaslindas, #horror, entre otros.

Bibliografía

- Abbasi, Ahmed, Hsinchun Chen, y Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.*, 26(3):12:1–12:34.
- Abdul-Mageed, Muhammad, Sandra Kübler, y Mona Diab. 2012. Samar: a system for subjectivity and sentiment analysis of arabic social media. En *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bakliwal, Akshat, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, y Vasudeva Varma. 2012. Mining sentiments from tweets. *WASSA 2012*, página 11.
- Brooke, Julian, Milan Tofiloski, y Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. *International Conference RANLP*, páginas 50–54.
- Choi, Yejin y Claire Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, páginas 793–801. Association for Computational Linguistics.
- Das, Sanjiv y Mike Chen. 2001. Yahoo! for amazon: Extracting market sentiment from stock message boards. En *Proceedings of the Asia Pacific Finance Association Annual Conference*.
- Díaz, Ismael, Grigori Sidorov, y Sergio Suárez-Guerra. 2014. Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. En *TASS 2012 Working Notes*, Mexico.
- Kamps, J. y M. Marx. 2002. Words with attitude. *1st International WordNet Conference*, páginas 332–341, Septiembre.
- Kim, Soo-Min y Eduard Hovy. 2004. Determining the sentiment of opinions. En *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Maas, Andrew L., Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, y Christopher Potts. 2011. Learning word vectors for sentiment analysis. En *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Martín-Watson, T y J Carrillo de Albornoz. 2012. Sistema para la clasificación de la polaridad y seguimiento de temas. En *TASS 2012 Working Notes*, Castellón, Spain.
- Nasukawa, Tetsuya y Jeonghee Yi. 2003. Sentiment analysis: capturing favorability using natural language processing. En *Proceedings of the 2nd international conference on Knowledge capture*, K-CAP '03, páginas 70–77, New York, NY, USA.
- Pang, Bo. 2012. *Opinion Mining and Sentiment Analysis*, volumen 1. Morgan and Claypool, Hanover, MA, USA, Mayo.

- Pang, Bo, Lillian Lee, y Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. En *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, páginas 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Saralegi Urizar, X y I San Vicente Roncal. 2012. Detecting sentiments in spanish tweets. *IEEE, Working Notes*.
- Su, Qi, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, y Zhong Su. 2008. Hidden sentiment association in chinese web opinion mining. En *Proceedings of the 17th international conference on World Wide Web*, WWW '08, New York, NY, USA. ACM.
- Taboada, Maite, Julian Brooke, Milan To-filoski, Kimberly Voll, y Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Comput. Linguist.*, 37(2), Junio.
- Tong, Richard M. 2001. An operational system for detecting and tracking opinions in on-line discussion. 1(6).
- Turney, Peter D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. En *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Vilares, David, Miguel Ángel Alonso, y Carlos Gómez-Rodríguez. 2013. Supervised polarity classification of spanish tweets based on linguistic knowledge. En *Proceedings of the 2013 ACM symposium on Document engineering*, páginas 169–172. ACM.
- Yang, Kiduk. 2008. Widit in trec 2008 blog track: Leveraging multiple sources of opinion evidence. Indiana, USA.
- Yang, Qiang, Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, y Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. En *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, páginas 751–760, New York, NY, USA. ACM.
- Zhang, Lei, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, y Bing Liu. 2011. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *HP Laboratories HPL-2011-89*.