

Algoritmos sobre el impacto de los medios de comunicación en medios sociales: estado de la cuestión

Algorithms on the impact of mass media in social media: state of the art

Joan Francesc Fondevila Gascón

*Doctor en Periodismo, Profesor titular de Universidad,
(Universitat Pompeu Fabra, Universitat Ramon Llull,
Universitat de Girona, Universitat Oberta de Catalunya,
Universitat Autònoma de Barcelona, Universitat de Barcelona)*

Fecha de recepción: 27 de enero de 2016

Fecha de revisión: 10 de julio de 2016

Para citar este artículo: Fondevila Gascón, J. F. (2017): Algoritmos sobre el impacto de los medios de comunicación en medios sociales: estado de la cuestión, *Icono 14*, volumen 15 (1), pp. 21-41. doi: 10.7195/ri14.v24i2.948

Resumen

La relación entre medios de comunicación y los medios sociales está acentuándose en plena Sociedad de la Banda Ancha. Mientras los mass media se plantean la difusión de los contenidos en redes sociales como una obligación, los algoritmos y las métricas de análisis en estas redes se multiplican para objetivizar el rendimiento que la empresa informativa puede obtener de su actividad virtual. En este artículo de innovación teórica se exponen el proceso de uso de redes y medios sociales y las métricas utilizadas en función del objetivo a conseguir, y se incide en el análisis de sentimiento, que permite conocer si la reacción del lector a las unidades de contenido de los medios de comunicación es positiva, neutra o negativa. Se concluye que la medición de la polaridad de los comentarios del lector ayuda a la empresa informativa a mejorar la experiencia del receptor gracias sobre todo a la interactividad.

Palabras clave

Medios de comunicación - Redes sociales - Análisis de sentimiento - Lector - Métrica - Reputación - Interactividad - Algoritmo

Abstract

The relationship between media and social media is accentuated in full Broad-band Society. While the mass media dissemination of content on social networks seems an obligation, algorithms and metrics analysis on these networks try to multiply performance for the media company from its virtual activity. This article, from a theoretical innovation perspective, analyses the process of using social networks and social media and metrics used following objectives. We study the sentiment analysis, which identifies if the reader's reaction to the content units of the media is positive, neutral or negative. It is concluded that the measurement of the polarity of the reader comments to the media company helps to improve the customer experience thanks in particular to receiver interactivity.

Key Words

Media - Social media - Sentiment analysis - Reader - Metrics - Reputation - Interactivity - Algorithm

1. Introducción

La relación entre medios de comunicación y medios sociales (especialmente redes sociales) es creciente en plena Sociedad de la Banda Ancha (Fondevila Gascón, 2013). Para los periódicos digitales es condición *sine qua non* la inclusión de iconos que llevan a las redes sociales disponibles por el medio, y es habitual calcular el impacto en forma de lecturas, *likes*, comentarios y comparticiones.

Los criterios son numerosos, y se prevé que el darwinismo propio de estas propuestas acabe determinando cuál o cuáles son las más aceptadas por la empresa informativa. Así como a la hora de determinar la influencia de un nodo en la red se está imponiendo Klout (Fondevila Gascón y Lamelo Varela, 2015), a la hora de determinar la métrica a seguir sobre el efecto de las inclusiones en las redes sociales las opciones son diversas.

Para un medio de comunicación estar presente de forma activa en redes sociales es el salvoconducto para establecer la necesaria interactividad con el lector, buscando el *engagement* en pleno *cloud journalism* (Fondevila Gascón, 2010) y *social journalism commerce*. La creación de una comunidad virtual de lectores mediante redes sociales requiere contenido de calidad, forjado mediante el uso adecuado de la hipertextualidad, el multimedia y la interactividad (Fondevila Gascón, Rom Rodríguez y Santana López, 2016).

Los periódicos digitales están consiguiendo visitas mediante enlaces posteados desde sus perfiles. Se trata de una manera lógica de optimizar la difusión con la viralidad de los medios sociales. Un estudio (Press Cutting Service, 2015) demuestra que Facebook y Twitter (las redes sociales más seguidas por españoles presentes en redes sociales) se erigen en redes favoritas de los principales diarios españoles (*El País*, *El Mundo*, *ABC*, *La Razón*).

Las cifras en el momento del estudio reflejaban que *El País* contaba con gran cantidad de seguidores en ambas redes (más de 4 millones en Twitter, 1,8 millones de “Me gusta” en Facebook); *El Mundo*, dos millones en Twitter y 1,2 millones en Facebook; *ABC*, 900.000 seguidores y 542.800 fans; y *La Razón*, 200.000 en Twitter y ocho mil *likes*.

En Twitter, la media de unidades de contenido lanzadas era de 126. El plus de interactividad lo obtenían las cabeceras económicas (*Cinco Días*, *El Economista* y

Expansión), a pesar de disponer de menos seguidores. En *Cinco Días* la media de “Me gusta” por contenido colgado en Facebook era de 39,3, y la compartición era de 63 veces (27,3 en *El Economista*). De ahí se infiere que en redes sociales cantidad no es sinónimo de calidad. La cifra de seguidores aumenta o reduce valor en función de la idoneidad (*target*) de la red social en la que se divulgue un medio.

El crecimiento experimentado por las redes sociales en los últimos años está conduciendo a los medios de comunicación, pues, a participar en ellas. En el terreno de las redes sociales hallamos oportunidades y riesgos de todo tipo. A tal efecto hay que diseñar una estrategia definida. Los responsables digitales de los periódicos acostumbran a establecer objetivos a partir de métricas (que pueden ser generales o propietarias) para determinar el éxito de las iniciativas. El ROI (Retorno Operativo de la Inversión) de un diario digital puede ser también un retorno de participación, que a medio y largo plazo se traduce en lectores apegados y evangelizadores (Powell, Groves y Dimos, 2011). Se trata de un indicador de cálculo complejo (Blanchard, 2009).

Uno de los objetivos de los diarios digitales, con más motivo que en otros sectores (Fondevila Gascón, Del Olmo Arriaga y Bravo Nieto, 2012), consiste en conseguir reputación, factor clave de la rentabilidad de una empresa informativa por cuanto se proyecta en forma de posible crecimiento.

En contraste con las técnicas cuantitativas utilizadas tradicionalmente (encuestas de satisfacción de los lectores), una de las opciones para medir la reputación es la lealtad o la fidelidad de los consumidores (Reichheld, 2003). En este sentido, es más sencillo aumentar la cantidad de lectores que promocionan al diario que incrementar el índice de satisfacción general. En este tipo de métricas los resultados no son del todo fiables, ya que se considera satisfecho a todo aquel lector que no esté insatisfecho. La negación, en este caso, no debe implicar afirmación, ya que existe el rango de neutralidad.

No obstante, se detecta una correlación directa entre el porcentaje de consumidores que recomendarían una marca (en este caso, extrapolable a lectores que recomiendan un periódico) y los índices de crecimiento relativos entre competidores para la mayoría de las industrias estudiadas. La lealtad del lector supone compromiso: recomienda contenidos a los otros lectores, que viralmente pueden

atraer a más seguidores. La inclusión de una noticia de impacto en una red social desemboca en más tráfico para ese medio de comunicación.

La lealtad del lector se revela como condición necesaria pero no suficiente para que un diario crezca. La palabra clave “recomendación”, en todo caso, no es eficaz para todos los sectores, por lo que la cuestión es determinar la probabilidad de recomendar. La numeración puede variar del cero al diez hasta escalas de Likert o de Osgood. Los lectores que puntúan nueve o diez son considerados promotores; si la calificación es de seis o menos, se les conceptúa como detractores. Los lectores que aportan puntuación intermedia se consideran pasivamente satisfechos. Para un diario digital el objetivo no es otro que intentar acumular más promotores para convertir a los lectores leales en la principal plataforma de difusión.

La propuesta de Reichheld no está ajena a críticas. Así, East, Hammond y Lomax (2008) critican la métrica de la recomendación si se trata de predecir el comportamiento o el rendimiento de una marca. Los reproches se centran en la autopredicción, ya que los encuestados no pueden anticipar con facilidad las circunstancias en las que llevarían a cabo una recomendación. Con relación al *Word on Mouth* (WOM), apuntan que afecta más en el momento de elegir una marca que en la propensión a recomendar. Por ende, se propone una combinación de experimentos a partir de juegos de rol y de encuestas, más allá de las constricciones de esas técnicas.

2. Método

Esta investigación es descriptiva, no experimental, ya que el objetivo es conocer los indicadores predominantes a través del análisis de las actividades y procesos que implican las métricas y los algoritmos. El fin no se restringe a la recolección de datos, sino a la predicción e identificación de las relaciones que existen entre dos o más variables, en este caso las tipologías métricas y los algoritmos sobre el impacto de los medios de comunicación en medios sociales. La metodología utilizada es cualitativa, mediante la técnica de investigación analítica (Fondevila Gascón y Del Olmo, 2013). Esta técnica se utiliza en este caso para estudiar conceptos analizando una selección de documentos, sintetizando la información para aportar un conocimiento del concepto tras la interpretación crítica de los acontecimientos en su contexto.

Igualmente se utiliza la técnica comparativa: se analizan y se comparan diversas aproximaciones (a partir de revistas científicas indexadas y obras de referencia) a métricas que pueden resultar de utilidad para diarios digitales. Para ellos, con actividad prolífica en redes sociales y deseo de contabilizar el retorno del esfuerzo, las alternativas son diversas. El objetivo consiste en exponer el proceso y la objetivización del uso de medios y redes sociales por parte de los diarios y los medios de comunicación a partir de algoritmos y métricas, con una finalidad de rentabilidad.

Los medios sociales han convertido a los consumidores en participantes activos en la creación, difusión y búsqueda del contenido (Hunter y Soberman, 2010). Las marcas han de ceder buena parte del control a los consumidores y entablar un diálogo con ellos (Bonsón y Ratkai, 2013). Los profesionales de la información y los ejecutivos necesitan métricas precisas para cada tipo de objetivo, pues la información que proporciona cada una variará sustancialmente y afectará a la toma de decisiones.

3. Desarrollo

3.1. Objetivos e indicadores de rendimiento

Un periódico digital puede intentar dar sentido a la cantidad de *inputs* obtenidos en redes sociales creando indicadores de rendimiento (Murdough, 2009). Se puede llevar a cabo un proceso iterativo que permita medir la reputación de un diario en las redes sociales a partir de un conjunto de objetivos y unos indicadores de rendimiento, los *key performance indicators*.

Como la métrica es por definición comparativa, los estándares deben plantearse de forma extrapolable a otros medios, incluso los que no sean propiamente periodísticos. Páginas web de impacto de medios audiovisuales u otro tipo de publicaciones deben ser susceptibles de comparación. La primera fase del proceso es la de concepto.

Para el diario digital, como primera fase, a corto plazo el objetivo se centrará en la generación de ingresos. Los diarios pueden probar fórmulas de pago como el *paywall* total, el *freemium*, el *utility* o *metered*, las franquicias o los micropagos (Fondevila Gascón, Sierra Sánchez y Del Olmo Arriaga, 2011) y cruzarlas con fórmulas de inter-

vención en redes sociales, para detectar posibles incrementos en tráfico o correlaciones con hipertextualidad, multimedia e interactividad, halladas en investigaciones recientes (Fondevila Gascón, Beriain Bañares y Del Olmo Arriaga, 2013) y clasificables métricamente (Fondevila Gascón, 2014). Uno de los indicadores principales es la generación de tráfico, que permite cautivar a más lectores (convertibles en suscriptores digitales) y más publicidad (pese a los modestos réditos del Click Through Rate, los anuncios no dejan de ser una fuente de ingresos necesaria para los periódicos digitales). De esta manera, si se promueve la prueba y la investigación entre diversas variables, se puede proyectar una mejor imagen de marca. A largo plazo, el diario digital puede crear valor de marca mediante el incremento de la satisfacción del cliente.

Como segunda fase, se propone al diario organizar la estrategia a partir de tres columnas: alcance, discusión y resultados. Se trata de monitorizar la calidad y la cantidad de las menciones de los autores de contenido, y conocer los ítems sobre los que trata la red para poder participar y controlar el valor económico de cada acción. En la línea de la trasposición en valor económico del impacto de una inserción en forma de noticia en un diario en papel, se trata de crear fórmulas y algoritmos para cuantificar cada consecución en una red social.

Como tercer escalón se plantea el diseño. El diario debe elegir los canales más adecuados en los que desarrollar la estrategia a seguir. Para un diario con un público eminentemente profesional, la acción puede centrarse en LinkedIn, red social no demasiado frecuentada por los periódicos digitales. Un diario puede asignar a un periodista el rastreo y la documentación de datos. También se pueden utilizar herramientas ya existentes, como Google Analytics. Las fórmulas más evolucionadas permiten conocer el estatus en tiempo real de una marca en términos de alcance, fuerza y tono (Botha, Farshid y Pitt, 2011).

Otro factor es la tecnología social disponible para facilitar a los diarios u otras marcas la participación en las redes sociales y analizar los resultados conseguidos (Lovett, 2012). Con el objetivo del descubrimiento, la búsqueda social proporciona resultados rápidos sobre marcas o consultas específicas. Posibles proveedores son Google, Bing, Yahoo, Social Mention o SearchWiki. Las métricas disponibles son las tendencias, la repetición de marcas y palabras clave y el número de fuentes únicas.

En la fase de análisis, la analítica social permite recoger datos y explorar los canales sociales, asuntos, autores influyentes y competidores. Posibles proveedores son Crimson Hexagon, SAS Social Media, Analytics, Omniture o Sysomos. Como métricas que arrojan esas plataformas hallamos las menciones sociales, la cuota de voz, la opinión, la influencia, la autoridad, la lealtad y el impacto.

En cuanto a la participación, la interacción social permite a los negocios monitorizar, asignar y responder a investigaciones de redes sociales. Empresas a considerar son Radian6, Lithium / Scout Labs, Socialmention o Howsociable. Las métricas son los comentarios, los “Me gusta”, las cuotas, la tasa de interacción y la “viralidad”.

La tecnología social facilitadora permite a los diarios crear destinos en línea para que los consumidores interactúen y participen, mediante empresas como Telligent o Jive. Usuarios registrados, nuevos usuarios o crecimiento de la comunidad son métricas al uso.

Finalmente, con el objetivo de mejorar la gestión, la agregación social, la colaboración interna y las herramientas de control proporcionan soluciones unificadas de gestión. Programas como Yammer o Salesforce Chatter facilitan el acceso a la efectividad de la red social y el valor del ciclo de vida del cliente.

Para desplegar se pretende que los datos que obtengan los diarios sean los esperados en términos de contenido, calendarios editoriales y otros. Ello implica controles de calidad de los métodos de recolección de datos.

En la fase de optimización el diario crea un informe final para evaluar el cumplimiento del programa que hubiera sido diseñado para las redes sociales en función de los indicadores de la fase de concepto. Se diagnostican los impulsores del rendimiento a partir de la fase de definición, y se identifican las oportunidades para la mejora del programa de comunicación en las redes sociales partiendo de análisis cuantitativo y cualitativo (fases de diseño y de despliegue). De ello surge un panel de información con datos relevantes para los intereses a corto y largo plazo (Peters et al., 2013). La participación en las redes sociales debe ir precedida de un análisis sobre objetivos, recursos (humanos y materiales) y esfuerzos.

3.2. Métricas de *social media* adaptables a medios de comunicación

La proliferación de métricas para medios sociales genera diversas opciones pero también, en ocasiones, cierta desorientación para los medios de comunicación, que incluso se pueden plantear la creación de indicadores propios, adaptados a sus características, ámbito geográfico natural, lengua del lector o las características que se desee.

Así, Barger y Labrecque (2013) proponen un conjunto de siete métricas. La primera es el volumen, es decir, la cantidad de menciones que una marca o un medio de comunicación recibe en los medios sociales durante un periodo determinado de tiempo. Esta variable de medición puede ser indicativa de los progresos realizados en cuanto a crear una conciencia o notoriedad de marca.

En segundo lugar plantean el *Share of voice*, es decir, la cuota de menciones, el volumen de menciones de una marca o medio de comunicación como porcentaje del volumen de menciones que reciben todas las marcas en una categoría de producto concreta, en este caso los diarios digitales. Se calcula teniendo en cuenta únicamente el volumen positivo, ya que las menciones negativas no se consideran una ventaja competitiva. Aquí se aplica el análisis de sentimiento. Este indicador permite compararse con la competencia como índice con el que controlar la efectividad con que la marca genera ruido.

El tercer indicador es el *engagement*, que lidera el factor interactividad dentro del modelo de calidad para diarios digitales (Fondevila Gascón, 2014). Por interacción se entiende que el lector lleve a cabo algún tipo de acción más allá de ver o leer. Esa acción puede ser marcar “Me gusta” en un *post*, comentar o responder al *post* de una marca o compartir dicho *post* con otros usuarios. Esta variable de medición permite conocer el interés que suscita cada *post* del diario. Ello puede alumbrar el enfoque de creación de futuras noticias o unidades de contenido en forma de *post*. El *engagement* se puede calcular como medida agregada o como porcentaje, si se quiere llevar a cabo un ajuste de las diferencias entre el *engagement* de unos *post* y otros en relación con las veces que se ve cada uno o el número de seguidores que tiene la marca en los medios sociales en un momento dado. El crecimiento de Facebook está suscitando estudios diversos (Pöyry, Parvinen y Malmivaara, 2013; Fondevila

Gascón et al., 2015), que pueden indicar tendencias a los medios de comunicación a las redes sociales, incluida la radio (Fondevila Gascón y Lamelo Varela, 2015).

El cuarto indicador son los defensores o *advocates*. Hay que diferenciar entre participación “pasiva” y “activa”. En este sentido, expresar un estado de acuerdo (clicando a “Me gusta”) es una participación pasiva; opinar (comentando la noticia en Facebook o en otra red social) y compartir (compartiendo la noticia con los amigos de Facebook u otras redes sociales) son formas más activas de participación. El lector adopta el papel de defensor de la marca, creando y subiendo contenido del diario digital. El hecho de que éste se recomiende o reciba menciones favorables aporta un valor incalculable, especialmente si el defensor es un influenciador con una marca elevada en Klout. Existen concomitancias entre esta métrica con la propuesta de Reichheld (2003). El objetivo consiste en acrecentar el número de promotores. Los amigos y seguidores de estos defensores estarán predispuestos a considerar el diario digital si defensores lo promueven.

El quinto indicador es el retorno operativo de la inversión (ROI), esto es, los ingresos generados por una campaña de marketing llevada a cabo en los medios sociales menos el coste de la campaña, todo ello dividido entre el coste de la campaña. Este indicador de medición es útil a la hora de evaluar objetivos cortoplacistas. En diarios digitales, se puede comunicar un código de descuento para probar una suscripción o para acceder a una sección. Las suscripciones consecuencia de una campaña son atribuibles a la misma y se puede calcular fácilmente su ROI.

El sexto indicador son los clientes potenciales generados (*leads generated*). Se trata de la cantidad de lectores potenciales generados para la base de datos a través de los medios sociales. Esta métrica aporta información de valor, pero no es sencillo atribuir un cliente potencial a la actividad de la marca en los medios sociales.

Finalmente, el séptimo indicador es el tiempo de respuesta, situado en segundos o minutos. De hecho, estos autores lo fijan en menos de treinta minutos. En periodismo digital la paciencia del lector se cifra en segundos o milisegundos.

En el caso de Bonsón y Ratkai (2013), se propone la medición de la interacción de todas las partes interesadas con la marca o el diario digital en términos

de popularidad, compromiso y viralidad, y la medición del sentimiento de esos mismos grupos de interés (comentarios positivos, negativos o neutros). Al tratarse de datos de acceso público, es posible calcular la trayectoria de la competencia y establecer un marco de comparación (Tabla 1).

Métrica	Posibilidades de medición	Fórmula
Popularidad	Porcentaje del total de <i>post</i> marcados con un “Me gusta”	$(\text{Número de } \textit{post} \text{ con "Me gusta"} \div \textit{Post} \text{ totales}) \times 100$
	Media de “Me gusta” por <i>post</i>	$\text{Total de "Me gusta"} \div \text{Número total de } \textit{post}$
	Popularidad de los mensajes entre los fans	$(\text{Media de "Me gusta"} \text{ por } \textit{post} \div \text{Número de fans}) \times 1.000$
Compromiso	Porcentaje de <i>post</i> totales que han recibido comentarios	$(\text{Número de } \textit{post} \text{ con comentarios} \div \text{Número total de } \textit{post}) \times 100$
	Media de comentarios por <i>post</i>	$\text{Total de comentarios} \div \text{Número total de } \textit{post}$
	Compromiso de los fans	$(\text{Media de comentarios por } \textit{post} \div \text{Número de fans}) \times 1.000$
Viralidad	Porcentaje del total de <i>post</i> que han sido compartidos	$(\text{Número de } \textit{post} \text{ que han sido compartidos} \div \textit{Post} \text{ totales}) \times 100$
	Media de veces en que un <i>post</i> ha sido compartido	$\text{Total de veces en que han sido compartidos los } \textit{post} \div \textit{Post} \text{ totales}$
	Viralidad de los mensajes entre los fans	$(\text{Media de veces en que un } \textit{post} \text{ ha sido compartido} \div \text{Número de fans}) \times 1.000$
Humor de los usuarios de social media en sus comentarios	Ratio entre comentarios positivos, negativos y neutros	Número de comentarios con tono positivo, negativo y neutro

Tabla 1: Métricas de social media. Fuente: Bonsón y Ratkai (2013).

La relevancia e incluso la validez de los “Me gusta” en redes sociales y de los seguidores es motivo de controversia. Para algunos un “Me gusta” aporta más que una cifra (Gerlitz y Helmond, 2013), por cuanto su valor reside en los potenciales “Me gusta”, comentarios u otro tipo de respuestas adicionales que puede generar.

INNOVACIÓN TEÓRICA

En el caso de Lovett (2012), se presentan métricas para el análisis social. Se agrupan en interacción (la actividad que acaba en conversión), compromiso (incluye visitas, comentarios, comparticiones y tiempo destinado a estas actividades), influencia, defensores e impacto (ROI). Se diferencian métricas en progreso de métricas ROI (Tabla 2).

Resultado	Métricas en progreso	Métricas ROI
Exposición	Alcance, Menciones en medios, Difusión y Actividad de publicación	Coste por campaña de exposición
Diálogo	Nuevos visitantes, Fuentes de referencia, Enlaces relevantes, Usuarios activos, Tasa de crecimiento de usuario, Temas de tendencia de la marca y Palabras claves de la marca	Coste por Nuevo visitante, Coste por futuro cliente cualificado
Interacción	Tiempo pasado, Páginas visualizadas, Juegos jugados, Concursos presentados, Aplicaciones descargadas, Mensajes enviados y Comentarios emitidos	Coste por interacción, Coste por visitante comprometido
Soporte	Visitantes que regresan, Visitas recientes, Frecuencia de la visita, Puntuación de satisfacción y Porcentaje de revisiones positivas	Coste por cliente satisfecho
Defensa	Distribución de contenidos, Gustos, Comparticiones, Marcadores, Estado de la comunidad, Influencia, Peso y Comentaristas importantes	Coste por campaña de defensa
Ingresos	Clientes totales, Valor medio de un pedido, Media de pedidos por cliente y Valor de vida del Cliente	Coste por ingreso de campaña por ingreso total de cliente

Tabla 2: Resultados para Social Media y métricas asociadas. Fuente: Lovett (2012).

Para un diario digital que esté planteándose las métricas en fase inicial, existen opciones simples, como el número de “Me gusta” en Facebook, Twitter, LinkedIn o Google + o las visualizaciones de un vídeo en YouTube. En estos casos hay que analizar el contexto y relacionarlo con acciones concretas y otras métricas complementarias.

Los profesionales del periodismo y las relaciones públicas necesitan medir el impacto de los medios sociales de una manera más holística, conceptualizando y midiendo la experiencia de los lectores con el contenido del diario los medios sociales y la interacción de los usuarios (Smith, 2013).

En todo caso, un primer paso consiste en elegir en cuántos medios sociales debe estar el diario, ya que el coste de oportunidad penaliza la ausencia en una red social en términos de lectores potenciales no conseguidos y peligros para la marca o cabecera del diario (Barger y Labrecque, 2013).

3.3. Análisis de sentimiento

Para los diarios es esencial estar presentes en redes sociales. Tanto es así que incluso alguna de ellas (Facebook) aspira subrepticamente a devenir un nuevo macroperiódico. Las opiniones son esenciales para el futuro de un diario, producto o servicio, ya que generan confianza (Wright, 2009). Ello justifica la creación de programas para el análisis de sentimiento, de forma que la información que otrora necesitaba encuestas y grupos focales está disponible gratuitamente en Internet (Bai, 2011; Mostafa, 2013). Las herramientas de análisis de sentimiento permiten identificar el objeto de discusión y clasificar opiniones según puntos de vista diferentes (Jacobson, 2009) y tres categorías posibles de comentario: positivo, negativo y neutral (Wijnhoven y Bloemen, 2013).

Los enfoques para llevar a cabo el análisis de sentimiento (Haddi, Liu y Shi, 2013) son los métodos basados en el aprendizaje automático o *machine learning*, basados en el uso de algoritmos que muestran a la máquina cómo llevar a cabo la clasificación según un conjunto de características determinadas a partir de una serie de ejemplos; métodos basados en la utilización de léxicos (dependen de una lista de palabras predefinidas y clasificadas según su polaridad); y análisis lingüístico, que utiliza las características sintácticas de las palabras u oraciones, el uso de negaciones y la estructura general del texto para determinar su orientación positiva, negativa o neutra. Normalmente el análisis lingüístico se suele usar en combinación con algún método basado en la utilización de léxicos.

El análisis de sentimiento precisa detectar qué partes del texto contienen algún tipo de opinión o sentimiento y determinar la polaridad o intensidad del sentimiento (Yu, Duan, y Cao, 2013), además de categorizar las palabras (Mostafa, 2013) y considerar el contexto y la forma de uso de las mismas (Rao et al., 2014).

El resultado del análisis de sentimiento no está garantizado, hasta el punto que un 70% de precisión ya es satisfactorio (Kennedy, 2012). Los rasgos del idioma y el uso de la ironía, además de la jerga de los medios sociales, complican la tarea para los diarios digitales. La cantidad de sentimientos que se expresa en medios sociales es modesta, hasta el punto que en Twitter el 19% de los *tweets* se refieren a una marca, pero apenas uno de cada cinco de ellos contiene alguna expresión de sentimiento u opinión (Bae y Lee, 2012). El hecho de que la mayoría de opiniones sean favorables a las marcas lleva a sospechar que éstas retiran de Internet los comentarios negativos y que muchos de los comentarios positivos los escriben ellas mismas. Se calcula también que aproximadamente el 10% de los comentarios negativos son falsos (Kennedy, 2012; Mostafa, 2013).

En cuanto al tipo de población, Pettit (2013) observa que las mujeres y las personas mayores tienden a expresar sentimientos más positivos. Por ende, cuando el público masculino puntúa una marca o un producto menos que las mujeres, una comparativa implicaría un factor de corrección.

Otro reto para el análisis de sentimiento es la validez externa. Así, si un diario digital obtiene datos de medios sociales, debe extrapolarlos a diferentes grupos de interés en general. Wijnhoven y Bloemen (2013) no hallaron ningún proveedor de servicios en el campo del análisis de sentimiento que aportase a los clientes de informes sobre la validez externa de sus resultados.

Otro condicionante a tener en cuenta y al que se enfrentan los sistemas de análisis del sentimiento es la calidad ortográfica de los textos. Es frecuente que, cuando los textos provienen de la web, los autores omitan acentos o letras, o empleen abreviaturas no reconocidas. A estos efectos la solución es emplear patrones para adaptar el texto.

Aunque The Spanish SO Calculator, siendo la posición dominante, trate todas las construcciones lingüísticas a nivel léxico, se puede plantear la utilización de la estructura sintáctica de la frase para obtener la orientación semántica de un texto. Así pues, se recomienda seguir un conjunto de pasos:

- a. Procesar los textos: se ha desarrollado un procesador *ad hoc* que trata la unificación de expresiones compuestas, como por ejemplo “a menos que”. Como consecuencia, actuarán como una sola unidad de significado. También se requiere la normalización de los signos de puntuación.
- b. Segmentar el texto en oraciones y dividir las en *tokens* para realizar el etiquetado morfosintáctico de cada una de las palabras del texto.
- c. Análisis sintáctico de dependencias: se identifican relaciones binarias dependiente entre los términos de una oración.

Cada uno de estos vínculos binarios da lugar a una dependencia. A la estructura obtenida se la denomina árbol de dependencias. Los modelos aplicados en el análisis del contenido a partir de la estadística son diversos. Así, encontramos modelos bayesianos (Kupiec, Pedersen y Chen, 1995), *machine learning* (Berger y Mittal, 2000; Barzilay y Lapata, 2008) y técnicas de clusterización semántica (Radev, Jing y Budzikowska, 2000). En el enfoque más propiamente lingüístico, las investigaciones se centran en las posiciones dentro del texto (Brandow, Mitze y Rau, 1995; Lin y Hovy, 1997), la estructura del discurso (Marcu, 1997) y las cadenas léxicas (Barzilay y Elhadad, 1999). Aquí encontramos los antecedentes del análisis cualitativo mediante el *big data*, que mejora las técnicas de resumen automático de textos (RAT).

El entorno multimedia propio de Internet está promoviendo el RAT no sólo escrito. El documento puede ser único o múltiplo; el idioma, monolingüe o plurilingüe; el género, técnico, científico o de noticias; la longitud, corta (una o dos páginas) o larga (más de 50 páginas); y, el medio, texto, audio o vídeo. Los formatos de salida son el material (resumen), el formato (categorías temáticas o resúmenes de prensa etiquetados) y estilo informativo, indicativo (señalando el tema sin dar más explicaciones), crítico y agregado (incorporando fuentes diferentes).

Los métodos del RAT pueden ser estadísticos, es decir, centrados en el número o veces que un término aparece repetido en un texto. El *trending topic* de *Twitter* es un buen ejemplo. El clasificador bayesiano indica la probabilidad de que una frase del documento sea incluida en el resumen. De hecho, se puede indicar el número de mayúsculas, la longitud de la frase, la estructura y la posición de la frase en relación al total del texto. También se trata de medir el significado de una palabra en vez de la frecuencia de aparición mediante la herramienta *WordNet*, sumando el recuento del término con la aparición de sinónimos y otros términos que asociamos.

Desde una perspectiva moral, el análisis de sentimiento puede suponer una monetización de la intimidad (Kennedy, 2012), aunque, en contrapartida, permite conocer las preferencias y las opiniones de los usuarios y proponerles servicios más personalizados y relevantes (Rao et al., 2014).

Pese a que los algoritmos desarrollados para los ordenadores no siempre permiten interpretar el lenguaje humano de manera adecuada, la precisión de estos sistemas crece. La mayor precisión de los programas informáticos y la incorporación del *big data* permiten tratar cantidades ingentes de información y desarrollar campañas de marketing y comunicación más adecuadas para los diarios digitales.

4. Conclusiones

Para el periodismo digital la aplicación de algoritmos genéricos o *customizados* puede aportar valor tanto en forma de retorno de la inversión como en forma de fidelización del lector, lo que se traduce en una mayor facilidad de crear una comunidad homogénea de lectores.

La métrica de medios sociales determina el impacto en el usuario final, por lo que crece su interés en el ámbito profesional. Para un diario digital es un objetivo medir la experiencia de los lectores con el contenido del diario en las redes sociales, medir el valor de las interacciones de los lectores en redes sociales, determinar la cantidad de redes sociales en las que participar e intentar apostar por la interactividad.

Para los diarios digitales el análisis de sentimiento se revela como un posible aliado, ya que medir la polaridad de los comentarios en los medios sociales ayuda al periódico a buscar soluciones más precisas y a mejorar la experiencia del lector mediante el diálogo con éste. Detectar el doble sentido, ponderar valoraciones en función del perfil y conseguir mayor validez externa son retos que la praxis puede determinar.

Una posibilidad sería aplicar la métrica marketiniana digital del Committed Visitor Volume (CVV), referente a la duración de la visita del lector al medio. Las cifras que recoge OJD Interactiva establecen una horquilla habitual entre los 5 y los 12 minutos de duración media de las visitas. Se trata de registros mucho más elevados que en otras opciones de publicidad digital, con lo que asociaciones como la promoción o el patrocinio arrojan resultados jugosos para un anunciante que deba aterrizar en la arena digital.

Otra red semiolvidada por los diarios digitales españoles es LinkedIn, la red profesional por antonomasia, más cuidadosa gracias al Air Traffic Controller. En ámbito *business*, no estar activo en LinkedIn es una especie de sacrilegio, el octavo pecado capital, la némesis del raciocinio empresarial. La presencia en esa red es también simbólica.

Por ende, es loable que los diarios digitales se posicionen y acumulen “Me gusta”, comentarios y comparticiones en las dos redes principales (aunque Instagram esté superando a Twitter), pero desterrar a Google + y LinkedIn es un error tildable de muy grave. La elección de las redes sociales donde participar y las métricas a considerar es, pues, un paso necesario para los diarios digitales.

Referencias

- Bae, Y. & H. Lee (2012). Sentiment Analysis of Twitter Audiences: Measuring the Positive or Negative Influence of Popular Twitterers. *Journal of the American Society for Information and Technology*, 63(12), 2521-2535.
- Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*, 50, 732-742.

- Barger, V. A. & Labrecque, L. (2013). An Integrated Marketing Communications Perspective on Social Media Metrics. *International Journal of Integrated Marketing Communications*.
- Barzilay, R. & Elhadad, M. (1999). Using Lexical Chains for the Text Summarization. *Advances in Automatic Text Summarization*, 10-17.
- Barzilay, R. & Lapata, M. (2008). Modeling Local Coherence: An entity-based approach. *Computational Linguistics*, 34(1).
- Berger, A. & Mittal, V. O. (2000). OCELOT: A System for Summarizing Web Pages. *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval*, 144-151.
- Blanchard, O. (2009). How to NOT calculate Social Media ROI. *The BrandBuilder Blog*. <http://thebrandbuilder.wordpress.com/2009/05/29/how-to-not-calculate-social-media-r-o-i/> (consultado el 27 de enero de 2016).
- Bonsón, E. & Ratkai, M. (2013). A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review*, 37(5), 787-803.
- Botha, E., Farshid, M. & Pitt, L. (2011). How sociable? An exploratory study of university brand visibility in social media. *South African Journal of Business Management*, 42(2), 43-51.
- Brandow, R. M.; Mitze, K.; Rau, L. F. (1995). Automatic condensation of electronic publications by sentence selection. *Information Processing & Management*, 31(5), 675-685.
- East, R., Hammond K. & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, 25, 215-224.
- Fondevila Gascón, J. F. (2010). El *cloud journalism*: un nuevo concepto de producción para el periodismo del siglo XXI. *Observatorio (OBS*) Journal*, 4(1), 19-35.
- Fondevila Gascón, J. F.; Sierra Sánchez, J. & Del Olmo Arriaga, J. L. (2011). New communicative markets, new business models in the digital press. *Trípodos* (Extra 2011-VI International Conference on Communication and Reality-Life without Media, Universitat Ramon Llull), 301-310.
- Fondevila Gascón, J. F.; Del Olmo Arriaga, J. L. & Bravo Nieto, V. (2012). Presencia y reputación digital en *social media*: comparativa en el sector de la moda. *Fonseca, Journal of Communication*, 5, 92-116.

- Fondevila Gascón, J. F. (2013). Periodismo ciudadano y *cloud journalism*: un flujo necesario en la Sociedad de la Banda Ancha. *Comunicación y Hombre*, 9, 25-41.
- Fondevila Gascón, J. F. & Del Olmo Arriaga, J. L. (2013). *El Trabajo de Fin de Grado en Ciencias Sociales y Jurídicas. Guía metodológica*. Madrid: Ediciones Internacionales Universitarias.
- Fondevila Gascón, J. F.; Beriain Bañares, A. & Del Olmo Arriaga, J. L. (2013). Hipertexto, multimedia e interactividad: comparativa empírica en el periodismo digital español. En Zilles, Klaus; Cuenca, Joan; Rom, Josep (Eds.). *Breaking the Media Value chain. VII International Conference on Communication and Reality*. Barcelona: Universitat Ramon Llull, 41-50.
- Fondevila Gascón, J. F. (2014). El uso de hipertexto, multimedia e interactividad en periodismo digital: propuesta metodológica de ranking de calidad. *ZER, Revista de Estudios de Comunicación (Journal of Communication Studies)*, 19(36), 55-76.
- Fondevila Gascón, J. F. & Lamelo Varela, C. (2015). La influencia social 2.0 en los programas magazín de las radios generalistas españolas con relación a su audiencia. *Estudios sobre el mensaje periodístico*, 21(2), 813-831.
- Fondevila Gascón, J. F.; Mir Bernal, P.; Crespo, J. L.; Santana López, E.; Rom Rodríguez, J. & Puiggròs Roman, E. (2015). La introducción de Facebook en el aula universitaria en España: la percepción del estudiante. *RELATEC: Revista Latinoamericana de Tecnología Educativa*, 14(3), 21-30.
- Fondevila Gascón, J. F.; Rom Rodríguez, J.; Santana López, E. (2016). Comparativa internacional del uso de recursos digitales en el periodismo digital deportivo: estudio de caso de España y Francia. *Revista Latina de Comunicación Social*, 71, 124-140. DOI: 10.4185/RLCS-2016-1087.
- Gerlitz, C. & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society* 15 (8), 1348-1365.
- Haddi, E., Liu, X. & Y. Shi (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26-32.
- Hunter, M. L. & Soberman, D. A. (2010) 'The Equalizer': Measuring and Explaining the Impact of Online Communities on Consumer Markets. *Corporate Reputation View* 13(4), 225-247.
- Jacobson, L. (2009). Take a sentimental journey: What sentiment analysis means for PR professionals. *Public Relations Tactics* (November), 18.

- Kennedy, H. (2012). Perspectives on Sentiment Analysis. *Journal of Broadcasting & Electronic Media*, 56(4), 435-450.
- Kupiec, J. P.; Pedersen, J. & Chen, F. (1995). A trainable document summarizer. *Proceedings of the 18th annual international ACM SIGIR conference on Research and Development in Information Retrieval*.
- Lin, C. Y. & Hovy, E. (1997). Identifying Topics by Position. *Proceedings of the fifth Conference on Applied Natural Language*, 283-290.
- Lovett, J. (2012). *Social Media. Métricas y análisis*. Madrid: Ediciones Anaya Multimedia.
- Marcu, D. (1997). *The rhetorical parsing, summarization, and generation of natural language texts*. University of Toronto, Toronto.
- Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40, 4241-4251.
- Murdough, C. (2009). Social Media Measurement: It's Not Impossible. *Journal of Interactive Advertising*, 10(1), 94-99.
- Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B. & Pauwels, K. (2013). Social Media Metrics – A Framework and Guidelines for Managing Social Media. *Journal of Interactive Marketing*, 27, 281-298.
- Pettit, A. (2013). Identifying the real differences of opinion in social media sentiment. *International Journal of Market Research*, 55(6), 757-767.
- Powell, G., Groves, S. & Dimos, J. (2011). *ROI of Social Media – How to improve the return on your social marketing investment*. Hoboken: John Wiley & Sons (Asia) Ltd.
- Pöyry, E., Parvinen, P. & Malmivaara, T. (2013). Can we get from liking to buying? Behavioral differences in hedonic and utilitarian Facebook usage. *Electronic Commerce Research and Applications*, 12, 224-235.
- Press Cutting Service (2015). *Redes sociales y medios de comunicación*. Madrid: Press Cutting Service.
- Radev, D. R.; Jing, H. & Budzikowska, M. (2000). Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *Proceedings of the 2000 NAACL-ANLP workshop on automatic summarization*.
- Rao, Y., Li, Q., Mao, X. & Wenyin, L. (2014). Sentiment topic models for social emotion mining. *Information Sciences*, 266, 90-100.

- Reichheld, F. F. (2003). The One Number You Need to Grow. *Harvard Business Review*, 81(12), 46-55.
- Smith, S. (2013). Conceptualising an evaluating experiences with brands on Facebook. *International Journal of Market Research*, 55(3), 357-374.
- Wijnhoven, F. & Bloemen, O. (2013). External validity of sentiment mining reports: Can current methods identify demographic biases, event biases, and manipulation of reviews? *Decision Support Systems*, 59, 262-273.
- Wright, A. (2009). Our Sentiments, Exactly. *Communications of the ACM (Association for Computing Machinery)*, 52(4), 14-15.
- Yu, Y., Duan, W. & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55, 919-926.