

Dynamic SLOD-BI: Infraestructura Dinámica de Inteligencia de Negocio Social

Rafael Berlanga¹, María José Aramburu², Indira Lanza¹, Dolores M. Llidó¹,
Lledó Museros², and Ismael Sanz²

¹ Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I
{berlanga,lanza,dllido}@uji.es

² Departamento de Ingeniería y Ciencia de los Computadores. Universitat Jaume I
{aramburu,museros,isanz}@uji.es

Resumen Este proyecto propone nuevas perspectivas de análisis y nuevas extensiones en la funcionalidad de la infraestructura de datos SLOD-BI (*Social Linked Open Data for Business Intelligence*). SLOD-BI se define como una infraestructura de datos enlazados y abiertos (LOD) orientada a capturar y publicar hechos extraídos de las redes sociales que son relevantes para los objetivos estratégicos. La principal limitación de este tipo de infraestructura es su naturaleza estática, ya que los datos son generados y publicados como conjuntos de datos RDF. Sin embargo, los hechos generados en las redes sociales son altamente dinámicos, y muchas veces requieren ser analizados en tiempo real. En el proyecto *Dynamic SLOD-BI* se aborda la generación dinámica de hechos para el cálculo a demanda de indicadores de redes sociales. El sistema propuesto descansa en el modelo conceptual de SLOD-BI, y plantea nuevos desafíos de investigación tales como la generación dinámica de conocimiento y el mantenimiento de su coherencia.

Keywords: Inteligencia de Negocio, Modelos Cognitivos, Redes Sociales, Indicadores Clave de Desempeño

1. Problema a resolver

El Modelo de Negocio Social (*Social Business Model*) es un nuevo paradigma empresarial que consiste en la utilización de las redes sociales en las actividades cotidianas de una organización. Las redes sociales constituyen un medio ideal donde desarrollar el marketing digital, así como para captar la Voz del Mercado y la Voz del Cliente. El alcance social y económico de estos medios es cada vez mayor, y las empresas no pueden obviarlos en sus planes estratégicos si quieren ser competitivas.

En este contexto, surge con fuerza una nueva forma de realizar la Inteligencia de Negocio (*Business Intelligence -BI*), que traslada el foco desde la ejecución del servicio (por ejemplo, ventas y promociones) al contenido generado por las personas en las redes sociales (por ejemplo, gustos, opiniones, experiencias, etc.). De este modo surge el concepto de la Inteligencia de Negocio Social (Social BI

-SoBI-), que consiste en capturar y analizar la información relevante y actual que se difunde en las redes sociales para ayudar a las empresas en la toma de decisiones. Muchas de las aproximaciones SoBI tratan de extender los resultados de los modelos BI al ámbito de las redes sociales [2]. Sin embargo, la falta de flexibilidad de estos modelos los hacen poco adecuados en escenarios con demandas analíticas muy cambiantes. Por otro lado, existen pocas aproximaciones que aborden el análisis social desde la especificación de indicadores surgidos de los objetivos estratégicos [9].

Este proyecto va dirigido a las pequeñas y medianas empresas (pymes) que, por su escala, carecen de los medios necesarios para abordar un análisis efectivo de las redes sociales en su contexto de negocio. Las herramientas comerciales existentes están muy limitadas a determinados indicadores de desempeño social (por ejemplo impacto, seguidores, etc.), que generalmente suelen estar poco relacionados con los objetivos estratégicos propios de cada empresa.

Por otro lado, existe un gran interés en promover iniciativas para publicar datos abiertos, no solo desde las administraciones, sino también desde organizaciones públicas y privadas que necesitan compartir sus esfuerzos para obtener plataformas tecnológicas de mayor calidad. En *Dynamic SLOD-BI* planteamos el uso de datos abiertos enlazados, pero en forma de flujos de datos (*streams*) en lugar de la visión estática concebida en la Web de los Datos. Estos flujos de datos estarán constituidos por hechos extraídos de las redes sociales, los cuales tendrán una validez temporal determinada. La extracción del flujo de hechos estará guiada por las demandas dinámicas que los usuarios especifiquen en función de sus objetivos estratégicos. Para analizar los hechos y realizar el seguimiento de los indicadores estratégicos, pueden seguir utilizándose las distintas herramientas BI disponibles (ej. Knime, Tableau o Power-BI).

2. Arquitectura y tecnologías aplicadas

La arquitectura que se requiere para este proyecto es de tipo Kappa, donde toda la información se sirve en *streaming* desde el origen hacia la herramienta analítica. Ello implica que todo el procesamiento semántico debe hacerse de forma dinámica, manteniendo una ventana temporal en constante movimiento dónde se va recogiendo la información necesaria para el etiquetado apropiado de los hechos. Cabe mencionar que existen muy pocos antecedentes sobre infraestructuras semánticas en *streaming* [6,7], y éstos están principalmente orientados al procesamiento de información de sensores en la Internet de las Cosas (IoT). En nuestra arquitectura, la componente para el procesamiento semántico de los datos dinámicos aplica grandes recursos de conocimiento existentes, muchos de los cuales también son altamente dinámicos (ej. Wikipedia, DBpedia Live y BabelNet Live).

Otra componente importante de esta arquitectura es la que proporciona nuevas funcionalidades útiles para el análisis en las redes sociales, principalmente:

- **Perfiles de usuario**, que consiste en clasificar a los usuarios en categorías a partir de la información pública disponible, generalmente sus mensajes y

descripciones. Las categorías pueden corresponderse con diferentes dimensiones de usuario: género, perfil profesional (periodista, empresa, etc.), o su rol social (*influencer*, *micro-influencer*, *hater*, *fan*, embajador, *bot*, etc.).

- **Vocabulario visual**, que consiste en combinar información visual y textual para obtener nuevos indicadores útiles. Por ejemplo, en el análisis de tendencias de diseño en cualquier sector requiere captar información referente a los estilos mencionados (vintage, industrial, etc.), las opiniones emitidas sobre los productos, y las paletas de colores empleadas [4].

Cabe indicar que estas nuevas funcionalidades de análisis se combinan con las anteriormente soportadas por SLOD-BI como son los tópicos, los sentimientos y las métricas sociales.

La naturaleza altamente dinámica de todo el sistema requiere un *módulo de supervisión continua* de los datos y el conocimiento manejados en la infraestructura. Este módulo permitirá el perfilado y la exploración de manera semi-automática para descubrir nuevos tópicos de análisis, tendencias emergentes y nuevos perfiles de usuario. Esta semántica emergente podría tardar cierto tiempo en ser incorporada a las bases de conocimiento utilizadas en la infraestructura (ej.- DBpedia), lo que no impide que pueda ser incorporada a los hechos de análisis actuales. Este módulo también permitirá detectar conceptos obsoletos para las tareas de análisis actuales.

Finalmente, la infraestructura se completa con una componente para la creación de indicadores clave que pueden ser creados, publicados y analizados desde la propia infraestructura. Es decir, los indicadores son parte integral de la infraestructura de datos. En esta componente se proporcionan mecanismos automáticos capaces de establecer relaciones útiles entre métricas, dimensiones e indicadores a partir de datos de referencia conocidos en un dominio dado. Para este propósito nos basamos en los trabajos existentes sobre catálogos semánticos de indicadores [3], y métodos automáticos de descubrimiento de modelos de análisis en estrella sobre LOD [8].

	Automotive		Banking		Average	
	P-micro	P-macro	P-micro	P-macro	P-micro	P-macro
LSTM+Glove	0.488	0.153	0.463	0.196	0.476	0.174
MLP+d2v	0.495	0.062	0.340	0.043	0.417	0.052
CNN+d2v	0.495	0.062	0.335	0.058	0.415	0.060
SVM+d2v	0.526	0.243	0.579	0.279	0.553	0.261
LM	0.699	0.205	0.760	0.394	0.730	0.300
Majority class	0.475	0.237	0.410	0.205	0.443	0.221
RepLab'14 best	0.450	-	0.500	-	0.475	-

Cuadro 1. Results for user categorization in the RepLab2014 dataset.

3. Resultados iniciales y conclusiones

Los resultados iniciales obtenidos en el proyecto corresponden a la definición de nuevos perfiles de usuario, los cuales se aplican de forma automática a la infraestructura mediante clasificación automática. Estos resultados se enmarcarían dentro del módulo de supervisión de datos que permite incorporar nuevos perfiles de usuario a la infraestructura semántica. Básicamente cada perfil consiste en un clasificador automático que debe entrenarse con muestras tomadas de la misma infraestructura SLOD-BI. En la tabla 1 se muestran los resultados obtenidos con diferentes clasificadores para el conjunto de datos RepLab2014 [1]. En los experimentos se probaron los clasificadores de última generación basados en redes neuronales: Multi Layer Perceptron (MLP), Convolutional Neural Networks (CNN) y Long Short Term Memory (LSTM). Estos toman como entrada dos tipos de codificaciones (*embeddings*): Glove y Doc2Vec (d2v). El método propuesto en el proyecto se basa en modelos de lenguaje (LM) [5], el cual forma parte de los mecanismos de anotación semántica de SLOD-BI. Los resultados son muy prometedores ya que de momento solo se ha utilizado los textos de los tweets, y podrían mejorar añadiendo las métricas de SLOD-BI.

Agradecimientos. Este trabajo ha sido financiado con el proyecto del MINECO con identificador TIN2017-88805-R.

Referencias

1. E. Amigó et al. Overview of RepLab 2014: Author Profiling and Reputation Dimensions for Online Reputation Management. In Proc. CLEF 2014, pp 307-322 (2014)
2. R. Berlanga, L. García-Moya, V. Nebot, M.J. Aramburu, I. Sanz, D.M. Llidó. SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence. IJDWM 11(4): 1-28 (2015)
3. C. Diamantini, D. Potena, E. Storti. SemPI: A semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators. Future Generation Comp. Syst. 54: 352-365 (2016)
4. Z. Falomir, L. Museros, I. Sanz, L. Gonzalez-Abril. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-Learn) Expert Systems with Applications 97(1): 83-94 (2018)
5. L. García-Moya, H. Anaya-Sánchez, R. Berlanga. Retrieving Product Features and Opinions from Customer Reviews. IEEE Intelligent Systems 28(3): 19-27 (2013)
6. A. Mauri, J.P. Calbimonte, D. Dell'Aglio, M. Balduini, M. Brambilla, E. Della Valle, K. Aberer. TripleWave: Spreading RDF Streams on the Web. International Semantic Web Conference 140-149 (2016)
7. S. Nadal, V. Herrero, O. Romero, A. Abelló, X. Franch, S. Vansummeren, D. Valerio. A software reference architecture for semantic-aware Big Data systems. Information and Software Technology, Volume 90, pp. 75-92 (2017)
8. V. Nebot, R. Berlanga Statistically-driven generation of multidimensional analytical schemas from linked data. Knowl.-Based Syst. 110: 15-29 (2016)
9. A. Maté, J. Trujillo, J. Mylopoulos. Specification and derivation of key performance indicators for business analytics: A semantic approach. Data Knowl. Eng. 108: 30-49 (2017)