

Estudio preliminar de la anotación automática de códigos CIE-10 en informes de alta hospitalarios

Preliminary Study of the Automatic Annotation of Hospital Discharge Report with ICD-10 codes

Mario Almagro¹, Raquel Martínez¹, Víctor Fresno¹, Soto Montalvo²

¹Universidad Nacional de Educación a Distancia (UNED), Juan del Rosal, 16, 28040 - Madrid

²Universidad Rey Juan Carlos (URJC), Tulipán, S/N, 28933 - Móstoles
{malmagro, raquel, vfresno}@lsi.uned.es, soto.montalvo@urjc.es

Resumen: En la actualidad, la cantidad de recursos que se destinan a la codificación de informes médicos es enorme. Con la reciente implantación del estándar CIE-10 en el sistema sanitario español se acrecientan las dificultades, ya que se incrementa el número de posibles códigos CIE por cada informe, disponiendo de una escasa cantidad de datos generados y codificados. En este artículo se describen los retos que plantea esta tarea y se propone una primera aproximación de combinación de técnicas para implantar un sistema capaz de, dado un informe, recomendar automáticamente códigos CIE-10 a los codificadores.

Palabras clave: Clasificación automática multietiqueta, códigos CIE-10, recuperación de información, divergencia de Kullback-Leibler

Abstract: Nowadays, the amount of resources dedicated to encode medical reports is huge. In spite of this, an automatic solution for annotation does not appear to be consolidated. With the recent implantation of the CIE-10 standard, encoding becomes increasingly complex, since the number of possible CIE codes for each report is increased and at the same time a minimal quantity of generated and encoded data is available. In this paper, it is described the challenges posed by this task and proposed a first combination of techniques for implementing a system capable of automatically assisting coding specialists by recommending CIE-10 codes regarding a medical report.

Keywords: Automatic multilabel classification, ICD-10 codes, information retrieval, Kullback-Leibler divergency

1 Introducción

La codificación médica según la Clasificación Internacional de Enfermedades (CIE) consiste en la asignación de códigos estandarizados a informes médicos en representación de diagnósticos y procedimientos, cuya finalidad es la generación de estadísticas de morbilidad y mortalidad. En 2015 se publicó el Real Decreto 69/2015, por el que se establece la obligatoriedad de utilizar la nueva versión CIE-10-ES como sistema de codificación clínica a partir del 1 de enero de 2016 en España.

CIE-10 es una clasificación alfanumérica jerárquica que contiene entre 3 y 7 dígitos, y aporta una información clínica más detallada que la anterior codificación CIE-9-MC. El número de diagnósticos aumenta de 14.315 a 69.099, y el número de procedimientos de 3.838 a 72.000. Al mismo tiempo, su recopilación en la historia clínica digital requiere un mayor esfuerzo para

los profesionales, aún no familiarizados con el nuevo estándar. En estos dos años, el enorme coste de los recursos que los centros hospitalarios dedican a la anotación manual (o mínimamente asistida) de códigos CIE-10 ha hecho evidente la necesidad de desarrollar herramientas automáticas que asistan en dicha tarea.

Sin embargo, el proceso de codificación automática no es trivial. Además del gran número de códigos existentes, los informes médicos están escritos en lenguaje natural, sujetos a la variabilidad inherente al lenguaje libre. Es común que los facultativos utilicen terminología distinta a la reflejada en las descripciones de los códigos CIE-10 para expresar información relacionada al desorden o procedimiento de forma más detallada. Asimismo, el texto presente en los informes puede incluir errores ortográficos, estructuras sintácticas incorrectas, uso de jerga, sinóni-

mos, acrónimos y abreviaturas, generando ambigüedad, lo que dificulta aún más su procesamiento automático.

Un posible enfoque del problema es el establecimiento de correspondencias entre los códigos CIE-9 y CIE-10, aprovechando así las aproximaciones ya existentes a la anotación con códigos CIE-9. La administración sanitaria americana ha descrito equivalencias entre ambas versiones a través de los Mapeos GEMs (General Equivalence Mappings)¹. Aunque éstos están disponibles en varias lenguas, la mayor exhaustividad y especificidad de los CIE-10 imposibilita una correspondencia directa entre ambas codificaciones para una parte considerable de los casos. Consecuentemente, los sistemas de anotación de códigos CIE-9 no son directamente adaptables al nuevo estándar, especialmente aquellos basados en reglas.

Otras aproximaciones abordan el problema por medio de enfoques basados en aprendizaje automático supervisado, por lo que requieren un entrenamiento con ejemplos manualmente anotados. Sin embargo, las anotaciones manuales disponibles indican que solo una pequeña fracción del total de códigos es utilizada con frecuencia, mientras que la inmensa mayoría de códigos o son poco frecuentes o no aparecen en ningún informe; dicha distribución limitará irremediablemente las aproximaciones basadas en enfoques supervisados. Aunque este problema se irá atenuando con el paso del tiempo a la par que aumenta la disponibilidad de documentos anotados, los centros hospitalarios necesitan soluciones tecnológicas efectivas a corto y medio plazo para asistir a los anotadores. Por todas estas razones, consideramos que es necesario explorar además tanto las aproximaciones basadas en aprendizaje semi-supervisado como los enfoques no supervisados.

El objetivo principal de esta propuesta es analizar la eficacia de los enfoques básicos; en concreto, se han comparado los enfoques supervisado y basado en Recuperación de Información (RI) aplicados al contexto de la codificación CIE-10 en español. Una vez analizados los resultados, planteamos una mejora del enfoque de RI a través del enriquecimiento de los índices, la selección de diferentes elementos estructurales en

los informes médicos y la combinación de distintos tipos de consultas.

En el segundo apartado del artículo se revisa el estado del arte, mientras que el tercero describe las características de la colección utilizada en la experimentación. En el apartado cuarto se presentan los *baselines* para el problema, y en el quinto nuestra propuesta de mejora del enfoque basado en RI. Finalmente, el apartado sexto resume las conclusiones y el trabajo futuro.

2 Trabajos relacionados

En general, los enfoques utilizados en el estado del arte para la recomendación y asignación de códigos CIE se pueden dividir en dos tipos: los basados en Procesamiento de Lenguaje Médico (PLM) y los basados en técnicas de clasificación.

En cuanto a los primeros, emplean bases de conocimiento y ontologías médicas para identificar los conceptos médicos en informes, y posteriormente asociarlos a los conceptos del esquema de clasificación. Por ejemplo, Ning et al. (2016) utilizan un modelo basado en ejemplos generado a partir de una base terminológica en chino con correspondencias con los códigos CIE-10 de 4 dígitos, por lo que aprovechan la estructura jerárquica del esquema de codificación; Chen et al. (2017) exploran la similitud semántica mediante el concepto de Longest Common Subsequence (LCS) entre los diagnósticos y los nombres dados por los CIE-10. Ambas propuestas basadas en aprendizaje no supervisado.

Por otro lado, en el segundo enfoque utilizado se generan clasificadores mediante aprendizaje automático supervisado para relacionar los informes con los códigos CIE. Por ejemplo, Subotin y Davis (2014) proponen un clasificador para anotar los procedimientos CIE-10, complementando el reducido tamaño de su corpus de entrenamiento mediante informes anotados con CIE-9 y las correspondencias del GEMs; Jatunarapit et al. (2016) emplean clasificadores basados en corpus ingleses y una serie de técnicas de RI para establecer similitudes con los términos tailandeses; Miftakhutdinov y Tutubalina (2017) emplean word embeddings entrenados a partir de un corpus de opiniones médicas de usuarios, junto con redes neuronales recurrentes para asignar los códigos.

A su vez, pueden encontrarse propuestas mixtas que combinan ambos enfoques. Por ejemplo, Boytcheva (2011) propone un clasificador supervisado multiclase, donde cada clase se corres-

¹<https://www.asco.org/practice-guidelines/billing-coding-reporting/icd-10/general-equivalence-mappings-gems>

ponde con un código CIE-10 de 4 dígitos, y utilizan repositorios de terminología médica para enriquecer la representación de los documentos; Zweigenbaum y Lavergne (2016) combinan dos clasificadores: uno entrenado con un conjunto de informes, y otro a partir de distintos diccionarios médicos. Seva et al. (2017), en cambio, emplean un enfoque de RI para buscar posibles códigos CIE-10 candidatos en distintos diccionarios, junto con una serie de clasificadores para filtrarlos.

En lo que respecta al castellano, los trabajos existentes se han centrado en la anotación del esquema CIE-9 (Goicoechea et al., 2013; Perez et al., 2015), no habiendo hasta donde conocemos ninguna propuesta que no se base en los mapeos GEMs para recomendar o clasificar códigos CIE-10 a partir del texto libre de los informes médicos. Sí hay propuestas para otras lenguas, como el inglés (Subotin y Davis, 2014; Miftakhutdinov y Tutubalina, 2017), el búlgaro (Boytcheva, 2011), turco (Arifoglu et al., 2014), el chino (Ning et al., 2016; Chen et al., 2017) o el tailandés (Jatunarpit et al., 2016). Desgraciadamente, los resultados entre todos estos enfoques no son comparables entre sí dado que utilizan diferentes corpus y lenguas. A ello se suman las limitaciones de acceso a las colecciones de datos que impiden su distribución.

Aun siendo éste un problema típico de clasificación supervisada multiclase y multietiqueta, también se ha abordado desde enfoques no supervisados basados en técnicas de RI, fundamentalmente debido a los problemas que conlleva tener un número de clases tan elevado. Algunas aproximaciones plantean el análisis lingüístico del documento para la generación de diferentes consultas, así como el uso de recursos externos, como Wikipedia, manuales de CIE, PubMed, SNOMED-CT u otras ontologías médicas, lo que permite aplicar distintos métodos de expansión de consulta (Rizzo et al., 2015; Zhang et al., 2017). Otra posible aproximación consistiría en acceder a bases de conocimiento con los códigos CIE y representadas por grafos RDF mediante consultas SPARQL (Chiaravalloti et al., 2014).

En relación a la anotación de informes con códigos CIE-10 y los sistemas de RI, se han desarrollado proyectos para informes en turco (Arifoglu et al., 2014), japonés (Chen et al., 2014), francés (VanMulligen et al., 2016) y alemán (Schmidt et al., 2017; Ho-Dac et al., 2017). Los dos primeros casos siguen la tendencia marcada por los trabajos anteriores, explotando especial-

mente la estructura jerárquica de la codificación CIE-10, mucho más precisa que su versión anterior. En general, todas estas aproximaciones recurren al enriquecimiento de las descripciones de los códigos CIE con recursos externos, como diccionarios, otros corpus o sinónimos.

Actualmente existen aplicaciones comerciales que asisten a los anotadores en el proceso de asignación de códigos CIE-10-ES. Aunque éstas emplean los mapeos GEMs y las descripciones asociadas a los códigos, en su mayoría requieren palabras clave propuestas por los codificadores. Además, tras recibir la recomendación de los códigos por parte de la herramienta, el anotador tendrá que filtrar y seleccionar aquellos relativos al informe. Nuestra propuesta consiste en otras técnicas, basadas en clasificadores automáticos y RI; por tanto, una futura aplicación con codificadores reales requeriría modelos pre-entrenados y una aproximación de aprendizaje activo para ir mejorándolos a la vez que los codificadores van anotando nuevos informes.

3 Colección

Disponemos de los informes de alta del Hospital Universitario Fundación de Alcorcón (HUFA) del año 2016 anotados con códigos CIE-10; en concreto son 13.177 informes. El corpus generado sobre estos informes contiene un total de 106.304 códigos asociados, de los cuales 82.020 representan diagnósticos y 24.284 procedimientos. A su vez, estos 106.304 códigos asignados se corresponden con 8.445 códigos CIE-10 diferentes, lo que supone un 6 % de todas las posibles codificaciones de esta clasificación. Las ocurrencias de los códigos responden a una ley de potencias; es decir, la mayoría de ellos aparecen escasas veces, al contrario que una minoría, que presenta un amplio número de ocurrencias. En concreto, el 90 % de las codificaciones existentes en el corpus comprende el 20 % de todos los códigos anotados, y de ese 90 % la mitad se han anotado una única vez.

Aunque el número de códigos CIE-10 en cada informe varía considerablemente, en promedio alcanza una media de 8 diagnósticos y 2 procedimientos. Además, en el proceso de anotación se establece un orden a la hora de asignar códigos, de forma que los códigos CIE-10 que ocupan la primera posición en los diagnósticos anotados en cada informe suelen representar el factor de riesgo principal, mientras que el segundo código suele ser menos relevante, y así sucesiva-

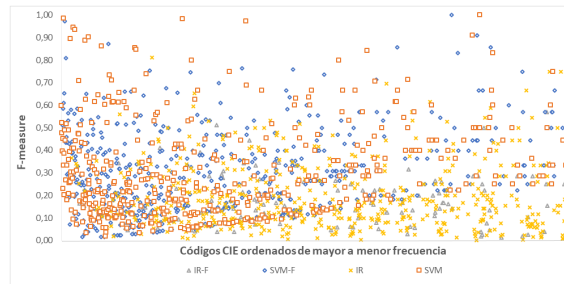


Figura 1: Baselines: distribución de las mejores aproximaciones (F -measure) por cada código CIE

mente.

Los informes médicos de alta de nuestra colección suelen seguir una estructura basada en secciones, en las que se agrupan diferentes tipos de información. A pesar de que existen varias plantillas para organizar estas secciones, los facultativos no siempre las utilizan; en ocasiones no definen las secciones o no marcan los límites de éstas. Según nos indicó el personal codificador del HUFA, considerando esa estructura, las secciones *Juicio Clínico*, *Procedimientos* y *Antecedentes* contienen la información más relevante para la codificación CIE-10. Se ha observado que, en general, *Procedimientos* y *Antecedentes* agrupan la información relativa a los procedimientos empleados, mientras que en la sección *Juicio Clínico* se resumen los principales diagnósticos de manera breve.

4 Baselines basados en clasificación supervisada y en RI

Con el fin de establecer un punto de partida y reflejar la complejidad que encierra la asignación automática de códigos CIE-10 en español, se han establecido dos *baselines*: uno basado en clasificación automática supervisada multietiqueta y otro basado en RI.

Hemos tenido en cuenta las secciones *Juicio Clínico*, *Procedimientos* y *Antecedentes* para representar los informes de alta previamente anonimizados. Por otra parte, hemos utilizado una versión de FreeLing enriquecida con 80.000 conceptos médicos (Ornoz et al., 2013), de la que se han eliminado los términos de la base de datos *BOT Plus* (en su mayoría medicamentos), debido a restricciones en la licencia de uso.

El *baseline* basado en aprendizaje supervisado utiliza un clasificador SVM (Máquinas de Vectores Soporte) mediante una validación cruzada sobre 10 iteraciones. En cada iteración se utiliza un 90% del corpus para entrenamiento

y el restante 10% para test. En cuanto al preprocesamiento, se han aplicado dos configuraciones distintas. Una primera aproximación en la que se ha empleado un proceso de *stemming* y una lista de *stop words*, y otra en la que se han utilizado solo los conceptos médicos detectados por FreeLing. En ambas, la representación se realiza dentro del modelo de espacio vectorial, utilizando *TF-IDF* como medida de pesado.

Por otro lado, el *baseline* de RI se ha diseñado utilizando la biblioteca Apache Lucene, indexando los textos de las descripciones de los códigos CIE-10. Durante la indexación se ha utilizado el mismo proceso de *stemming* y la lista de *stop words*. Las consultas se construyen a partir de cada oración de las secciones consideradas del informe, combinado todas ellas mediante funciones lógicas OR en una única *query*; de esta forma, será el propio motor de búsqueda el que calcule la relevancia de cada código respecto a la consulta por medio de su función de *ranking* estándar. Para generar distintas configuraciones del sistema se han considerado diferentes aspectos en la consulta. Por un lado, se han utilizado dos tipos de *matching*: exacto, en el que la unidad de búsqueda en el campo descripción es la oración, o por palabras, que incrementa las probabilidades de asignación con cada coincidencia entre los términos y la descripción del código. Adicionalmente, se ha empleado FreeLing para

Aproximación	Precision	Recall	F-measure
SVM (SVM con Stemming)	0.5370	0.2115	0.3034
SVM-F (SVM con Freeling)	0.4667	0.1980	0.2780
RI (RI sin FreeLing, matching por palabras y radio de búsqueda 15)	0.0551	0.0799	0.0653
RI-F (RI con FreeLing, matching exacto y radio de búsqueda 15)	0.0415	0.0310	0.0355

Tabla 1: Resultados de los *baselines*

extraer los conceptos médicos. Por último, se han establecido diferentes radios de búsqueda para limitar el número de códigos CIE-10 devueltos por el motor de búsqueda con cada informe.

Aproximación	N. CIEs mejor modelados	F-measure	Δ F-measure	N. ocurrencias
SVM-F	563	0.3587	0.1017	70
SVM	493	0.3427	0.1458	92
RI-F	278	0.2576	0.1488	10
RI	1198	0.2406	0.2037	11
Sin Mod.	5960	-	-	3

Tabla 2: Mejores *baselines* por cada CIE-10. Δ F-measure es el incremento medio de *F-measure* del mejor *baseline* respecto al resto

Las mejores aproximaciones y sus resultados pueden verse en la Tabla 1. Se han empleado las medidas estándar de evaluación *Precision*, *Recall* y *F-measure* para comparar los códigos CIE correctamente asociados en cada enfoque. Por otra parte, la Figura 1 representa la distribución de los mejores *baselines* por cada código CIE de la colección en función de *F-measure*, ordenándolos de mayor a menor frecuencia dentro del corpus. En la Tabla 1 se observa que los mejores resultados se alcanzan con el enfoque de clasificación supervisada, mientras que la aproximación de RI ofrece unos resultados bastante distanciados de éstos. Sin embargo, observando la distribución de los resultados de *F-measure* en la Figura 1, se llega a la conclusión de que el enfoque supervisado principalmente consigue clasificar codificaciones CIE-10 con una alta frecuencia en el corpus, las cuales suponen un reducido número; en concreto, consigue modelar adecuadamente un 10% de todas las codificaciones CIE-10 del corpus, que se corresponden con el 80% de todos los códigos anotados. Como ya se indicó, la mayoría de los códigos CIE-10 del corpus aparecen con poca frecuencia, lo que dificulta su modelado. El enfoque de RI, sin embargo, aunque arroja peores valores, no tiene la misma dependencia de la frecuencia, lo que podría suponer una posible vía para complementar el enfoque supervisado.

En la Tabla 2 se muestra el número de códigos CIE-10 en los que una aproximación representa el mejor sistema de anotación para ese código. El enfoque de RI sin aplicar FreeLing duplica a los sistemas SVM en el número de mejores codificaciones recuperadas en base a *F-measure*. Como era de esperar, observando la media de ocurrencias de esas codificaciones se aprecia que los

sistemas SVM modelan mejor aquellos códigos con mayor número de apariciones en la colección de entrenamiento, mientras que el enfoque de RI no refleja esa dependencia presentando además el mayor incremento medio de *F-measure* con respecto al resto de *baselines*.

El objetivo de este estudio preliminar es establecer mediante qué técnicas y con qué configuraciones sería posible asistir al proceso de anotación manual de informes médicos, por lo que se persigue la búsqueda de valores altos de *Precision* y *F-measure*. Por ello, y a la vista de los resultados anteriores, parece claro que las aproximaciones de clasificación pueden proporcionar mejores resultados en los códigos CIE-10 más frecuentes, mientras que el enfoque de RI puede funcionar mejor en los casos de códigos más inusuales. Centrándonos ahora en el sistema de RI, en el siguiente apartado se van a presentar distintos enfoques para mejorar sus resultados.

5 Mejora de la propuesta basada en RI

El *baseline* de RI se basa únicamente en la indexación de la descripción de los códigos CIE-10. Para mejorar sus resultados se han estudiado: el enriquecimiento de los índices con terminología asociada a cada código, y la combinación de consultas obtenidas a partir de distintas secciones del informe.

5.1 Enriquecimiento de los índices con KLD

Se propone emplear la *Divergencia de Kullback-Leibler* (KLD) como método para extraer terminología relacionada con cada código y añadirla al índice. Con esta técnica se pretende encontrar qué términos aparecen con mayor probabilidad en los informes anotados con un determi-

Aproximación	Precision	Recall	F-measure
RI sobre la Descripción, sin FreeLing, matching por palabras y radio de búsqueda 15	0.0551	0.0799	0.0653
RI sobre KLD con 30 términos, sin FreeLing, matching por palabras y radio de búsqueda 15	0.2093	0.0189	0.0346
RI sobre Descripción y KLD con 30 términos (pesado en 2), sin FreeLing, matching por palabras y radio de búsqueda 10	0.1327	0.0924	0.1090

Tabla 3: Resultados mediante el enriquecimiento de los índices con KLD

nado código, y que a su vez tengan una baja probabilidad en el resto. Para ello, se ha empleado el mismo 90 % de los informes usados en el entrenamiento del SVM para calcular cuáles serían los términos que mejor caracterizan los documentos en los que aparece cada código CIE-10 del corpus. Con el objetivo de estudiar distintas configuraciones en esta aproximación, se ha aplicado KLD sobre el conjunto de términos presente en las secciones *Juicio Clínico*, *Procedimientos* y *Antecedentes*, así como únicamente sobre los términos del *Juicio Clínico*. Posteriormente, por cada código CIE-10, además de la descripción del código, se indexan en campos separados los N términos más representativos de cada código, variando N de 1 a 80. Por último, se ha considerado asignar diferentes factores de empuje a los campos del índice para estudiar su influencia en el proceso de anotación.

La Tabla 3 presenta las configuraciones que han obtenido los mejores valores de *F-measure*. Aunque se observa una mejora de la *Precision* al realizar las consultas solo contra el campo KLD que contiene los 30 términos más representativos de cada código CIE-10, la mejora más importante de los resultados se produce al combinar este campo con la descripción del código en una única consulta.

5.2 Consultas considerando la estructura del informe

Como no se dispone de evidencias para averiguar qué parte específica del informe determina cada código, se ha querido analizar el efecto de reali-

zar las consultas con las secciones *Juicio Clínico*, *Procedimientos* y *Antecedentes* de forma individual y combinadas. Además, se han estudiado los resultados para diagnósticos y procedimientos de forma independiente.

Tras el análisis de múltiples configuraciones, en la Tabla 4 se muestran los mejores valores de *Precision* y *F-measure* por cada tipo de código. Se puede ver que el enfoque de RI consigue anotar más diagnósticos que procedimientos, lo que puede indicar que los diagnósticos están mejor caracterizados en los informes. Además, se muestran también la *Precision* para los primeros (P_1), segundos (P_2) y terceros códigos (P_3), cuantificando así los códigos correctos recuperados en los informes en dichas posiciones.

Si nos fijamos en los valores de *Precision* según la relevancia de los códigos (P_1 , P_2 y P_3), se puede observar la existencia de diferentes configuraciones que favorecen la obtención de códigos en distintas posiciones y, por lo tanto, sería de esperar que estableciendo una combinación de ellas se pudiera incrementar la precisión global del sistema. Por otro lado, observando de forma separada los altos valores de *Precision* obtenidos para diagnósticos y procedimientos, sería conveniente aplicar diferentes técnicas de RI en ambos tipos de código CIE-10. En definitiva, se puede pensar en diseñar un método basado en la combinación de algunas de las configuraciones mostradas para potenciar este enfoque de anotación basado en RI.

En la Tabla 4 también se muestra esa posible combinación (Q7), alcanzando unos valores de

Tipo	Query	Secciones	Campos de búsqueda	Aproximación	P_1	P_2	P_3	P_R	Precision	Recall	F-measure
Diagnósticos	Q ₁	Juicio clínico y Antecedentes	KLD con 30 términos (pesado en 2)	RI sin FreeLing, matching por palabras y radio de búsqueda 10	0.0013	0.5449	0.3015	0.2636	0.1608	0.1796	0.1697
	Q ₂	Juicio clínico	Descripción y KLD con 10 términos del juicio clínico	RI sin FreeLing, matching por palabras y radio de búsqueda 1	0.1740	0.0634	0.0028	0.0000	0.3655	0.0405	0.0729
	Q ₃	Juicio clínico y procedimientos	Descripción y KLD con 1 término del juicio clínico (pesado en 2)	RI sin FreeLing, matching exacto y radio de búsqueda 1	0.0486	0.0028	0.0000	0.0000	0.4926	0.0102	0.0200
Procedimientos	Q ₄	Juicio clínico y Antecedentes	KLD con 30 términos (pesado en 2)	RI sin FreeLing, matching por palabras y radio de búsqueda 10	0.0150	0.1926	0.0685	0.1576	0.1139	0.0708	0.0874
	Q ₅	Procedimientos	KLD con 10 términos	RI sin FreeLing, matching por palabras y radio de búsqueda 2	0.0039	0.0043	0.0005	0.0000	0.2259	0.0028	0.0055
	Q ₆	Juicio clínico, procedimientos y antecedentes	KLD con 3 términos	RI sin FreeLing, matching por palabras y radio de búsqueda 2	0.0553	0.0985	0.0016	0.0000	0.2012	0.0482	0.0777
Q₇ = Q₁ + Q₂ + Q₃ + Q₄ + Q₅					0.0077	0.4298	0.2578	0.2870	0.1465	0.1165	0.1298

Tabla 4: Resultados con otras secciones del informe desglosados por tipo de código. P_1 , P_2 y P_3 representan la fracción de códigos correctos recuperados en las posiciones de anotación 1, 2 y 3 respectivamente. P_R indica esa fracción en el resto de posiciones

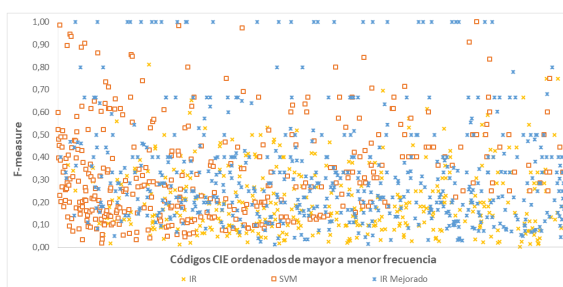


Figura 2: Mejora de RI frente a los *baselines*: distribución de las mejores aproximaciones (*F-measure*) por cada código CIE

F-measure de 0.13, duplicando los valores obtenidos por el *baseline*. Se ha representado su distribución (Figura 2) frente a los resultados de los *baselines*. Como se aprecia, esta combinación de consultas logra mejorar a las otras aproximaciones en una de cada dos codificaciones CIE. Por último, en la Tabla 5 se observa que la nueva aproximación del sistema de RI duplica el número de códigos CIE mejor caracterizados con respecto al sistema que genera las consultas sobre la descripción de los códigos. Al mismo tiempo, el número de códigos CIE sin modelar desciende de 5960 a 4334.

Aproximación	N. CIEs mejor modelados	$\overline{F\text{-measure}}$	$\Delta \overline{F\text{-measure}}$	N. ocurrencias
SVM	427	0.3649	0.1487	97
RI	914	0.2751	0.1948	10
RI Mejorado	2054	0.2980	0.2636	8
Sin Mod.	4334	-	-	3

Tabla 5: Mejores aproximaciones por cada CIE-10: comparación con RI mejorado. $\Delta \overline{F\text{-measure}}$ es el incremento medio de *F-measure* del mejor *baseline* respecto al resto

6 Conclusiones y trabajo futuro

Los métodos de clasificación basados en un aprendizaje supervisado demuestran su eficacia en problemas con un gran número de datos disponibles. En nuestro caso concreto, resultan útiles para anotar los códigos CIE presentes en un número considerable de informes del corpus. Sin embargo, el comportamiento obtenido para los códigos CIE-10 con pocas ocurrencias es opuesto, ya que el sistema en la mayoría de los casos no es capaz de generar modelos de clase a partir de los pocos informes disponibles.

El problema de la asignación de códigos CIE-10 viene condicionado por el enorme número de

códigos y por la distribución real de diagnósticos y procedimientos. Llevado a la práctica, los desórdenes que padecen las personas, así como los procedimientos que llevan a cabo los facultativos, suelen seguir una tendencia general que va a repercutir sobre los datos finalmente disponibles. Si bien es cierto que un método que se centre en los códigos más frecuentes puede responder con unos niveles de *Precision* considerables, supone un sesgo nada despreciable, ya que cualquier código menos usual prácticamente no se contempla. Siendo éste un problema en el que las soluciones se basan en códigos bien definidos y estructurados, no parecería lógico desarrollar un sistema incapaz de asignar una gran parte de los códigos, aún cuando las condiciones lo demandasen; por ello, en este artículo se estudian técnicas complementarias al aprendizaje automático, específicamente desde un enfoque de RI.

Como muestra la experimentación realizada, la generación de consultas a partir del informe y su aplicación únicamente sobre la descripción de los códigos no arroja buenos resultados. No obstante, dicha consulta en combinación con la terminología extraída a partir de la *Divergencia de Kullback-Leibler* logra recuperar más códigos CIE-10 afines a los informes. Un análisis más detallado de esos resultados desvela la distinta naturaleza de diagnósticos y procedimientos, así como los códigos más relevantes para el anotador y aquellos asignados en una segunda o tercera posición. Todas estas pautas parecen indicar la necesidad de seguir explorando la combinación de aproximaciones complementarias, así como de combinar distintas consultas, cada una orientada a esa fracción de códigos CIE que persigue recuperar.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación a través del proyecto

PROSA-MED: TIN2016-77820-C3.

Bibliografía

- Arifoglu, D., Deniz, O., Aleçakir, K., Yöndem, M. 2014. CodeMagic: Semi-Automatic Assignment of ICD-10-AM Codes to Patient Records. En *Information Sciences and Systems 2014*, páginas 259–268.
- Boytcheva, S. 2011. Automatic Matching of ICD-10 codes to Diagnoses in Discharge Letters. En *Proceedings of BioNLP*, páginas 11–18.
- Chen, S., Lai, P., Tsai, Y., Chung, J., Hsiao, S., Tsai, R. 2014. NCU IISR System for NTCIR-11 MedNLP-2 Task. En *Proceedings of the 11th NTCIR Conference*, páginas 9–12.
- Chen, Y., Lu, H., Li, L. 2017. Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity. En *PloS one*, vol. 12(3).
- Chiaravalloti, M.T., Guarasci, R., Lagani, V., Pasceri, E., Trunfio, R. 2014. A Coding Support System for the ICD-9-CM standard. En *Proceedings of ICHI 2014*, páginas 71–78.
- Goicoechea, J.A., Nieto, M.A., Laguna, A., Canto V.D., Rodríguez, J., Murillo, F. 2013. Desarrollo de un sistema de codificación automática para recuperar y analizar textos diagnósticos de los registros de servicios de urgencias hospitalarios. En *Emergencias*, vol 25:430–436.
- Ho-Dac, L. M., Fabre, C., Birski, A., Boudraa, I., Bourriot, A., Cassier, M., Delvenne, L., Garcia-Gonzalez, C., Kang, E. B., Piccinini, E., Rohrbacher, C., Séguier, A. 2017. LITL at CLEF eHealth2017: automatic classification of death reports. En *CLEF*.
- Jatunaratit, P., Piromsopa, K., Charoanlap, C. 2016. Development of thai text-mining model for classifying ICD-10 TM. En *Proceedings of ECAI 2016*, páginas 1–6.
- Miftakhutdinov, Z., Tutubalina, E. 2017. KFU at CLEF eHealth 2017 Task 1: ICD-10 Coding of English Death Certificates with Recurrent Neural Networks. En *CLEF*.
- van Mulligen, E. M., Afzal, Z., Akhondi, S. A., Vo, D., Kors, J. A. 2016. Erasmus MC at CLEF eHealth 2016: Concept Recognition and Coding in French Texts. En *CLEF Working Notes*, páginas 171–178.
- Ning, W., Yu, M., Zhang, R. 2016. A hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation. En *BMC Medical Informatics and Decision Making*, vol. 1:16–30.
- Pérez, A., Gojenola, K., Casillas, A., Oronoz, M., Díaz de Ilarraza, A. 2015. Computer aided classification of diagnostic terms in Spanish. En *Expert Systems with Applications*, vol. 42(6):2949–2958.
- Rizzo, S.G., Montesi, D., Fabbri, A., and Marchesini, G. 2015. ICD Code Retrieval: Novel Approach for Assisted Disease Classification. En *Data Integration in the Life Sciences, LNCS*, vol. 9162:147–161.
- Schmidt, D., Budde, K., Sonntag, D., Profitlich, H. J., Ihle, M., Staeck, O. 2017. A novel tool for the identification of correlations in medical data by faceted search. En *Computers in Biology and Medicine*, vol. 85:98–105.
- Seva, J., Kittner, M., Roller, R., Leser, U. 2017. Multi-lingual ICD-10 coding using a hybrid rule-based and supervised classification approach at CLEF eHealth 2017. En *CLEF*.
- Subotin, M., Davis, A. 2014. A System for Predicting ICD-10-PCS Codes from Electronic Health Records. En *Proceedings of BioNLP*, páginas 59–67.
- Zhang, D., He, D., Zhao, S., Li, L. 2017. Enhancing Automatic ICD-9-CM Code Assignment for Medical Texts with PubMed. En *Proceedings of BioNLP*.
- Zhao, S., He, D., Zhang, D., Li, L., Meng, R. 2017. Automatic ICD Code Assignment to Medical Text with Semantic Relational Tuples. En *Proceedings of iConference 2017*, vol. 2:156–158.
- Zweigenbaum, P., Lavergne, T. 2016. Hybrid methods for ICD-10 coding of death certificates. En *Proceedings of LOUHI*, páginas 96–105.
- Oronoz, M., Casillas, A., Gojenola, K., Perez, A. 2013. Automatic annotation of medical records in Spanish with disease, drug and substance names. En *Iberoamerican Congress on Pattern Recognition*, páginas 536–543.