

APLICACIÓN DE LA MINERÍA DE DATOS AL ESTUDIO DE LAS ALTERACIONES RESPIRATORIAS DURANTE EL SUEÑO

Carlos Zamarrón Sanz, Vanesa García Paz, Uxío Calvo Álvarez, Fernanda Pichel Guerrero, José Ramón Rodríguez Suárez
Servicio de Neumología. Hospital Clínico Universitario de Santiago de Compostela

Resumen

En muchas áreas del saber, el conocimiento se ha venido obteniendo por el clásico método hipotético-deductivo de la ciencia positiva. En él es fundamental el paso inductivo inicial: a partir de un conjunto de observaciones y de unos conocimientos previos, la intuición del investigador le conduce a formular la hipótesis. Esta "intuición" resulta inoperante cuando no se trata de observaciones aisladas y casuales, sino de millones de datos almacenados en soporte informático. Las técnicas de análisis estadístico, desarrolladas hace tiempo, permiten obtener ciertas informaciones útiles, pero no inducir relaciones cualitativas generales, o leyes, previamente desconocidas; para esto se requieren otras técnicas de análisis inteligente que están enfocadas a la inducción de conocimiento en bases de datos. Consecuencia de esta creciente necesidad ha aparecido un nuevo campo de interés: la Minería de Datos (data mining), que incluye los nuevos métodos matemáticos y técnicas software para análisis inteligente de datos. Por tanto, la Minería de Datos surge a partir de sistemas de aprendizaje inductivo en ordenadores, al ser aplicados a bases de datos, y su importancia crece de tal forma que incluso es posible que, en el futuro, los sistemas de aprendizaje se usen de forma masiva como herramientas para analizar datos a gran escala.

Se denomina descubrimiento de conocimiento en bases de datos (KDD) al proceso global de búsqueda de nuevo conocimiento a partir de los datos de una base de datos. Este proceso incluye no sólo el análisis inteligente de los datos con técnicas de Minería de Datos, sino también los pasos previos, como el filtrado y preprocesado de los datos, y los posteriores, como la interpretación y validación del conocimiento extraído. El KDD es un proceso iterativo e interactivo. Es iterativo ya que la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad.

Correspondencia:

Carlos Zamarrón Sanz
Servicio de Neumología. Hospital Clínico Universitario
c/Travesía de la Choupana s/n. 15706 Santiago de Compostela
E-mail: carlos.zamarron.sanz@sergas.es

Pneuma 2006; 6: 156 - 166

Es interactivo porque el usuario, o más generalmente un experto en el dominio del problema, debe ayudar en la preparación de los datos, validación del conocimiento extraído, etc.

1. Minería de Datos

La Minería de Datos se caracteriza por el análisis de grandes cantidades de datos con el fin de descubrir relaciones insospechadas y resumir la información mediante nuevas técnicas de forma que pueda ser interpretable y útil a sus usuarios¹⁻³. Las relaciones y resúmenes obtenidos mediante técnicas de Minería de Datos son conocidos a menudo como modelos o patrones. Ejemplos de dichos resúmenes pueden ser reglas, clusters, grafos, estructuras en árbol y patrones recurrentes en series temporales; proporcionando la Inteligencia Artificial un buen número de técnicas de lo que se conoce como Computación Flexible para la realización de un procesamiento inteligente de los datos.

La Minería de Datos es un término relativamente moderno que integra numerosas técnicas de análisis de datos y extracción de modelos. La Minería de Datos tiene como objetivo analizar los datos para extraer conocimiento. Aunque se basa en varias disciplinas, algunas de ellas más tradicionales (como la estadística), se distingue de ellas en la orientación más hacia el fin que hacia el medio. Y el fin lo merece: ser capaces de extraer patrones, de describir tendencias y regularidades, de predecir comportamientos y, en general, de sacar partido a la información computerizada que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades, permite a los individuos y a las organizaciones comprender y modelar de una manera más eficiente y precisa el contexto en el que deben actuar y tomar decisiones. En la práctica, los modelos pueden ser de dos tipos: predictivos y descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes, usando otras variables o campos de la base de datos, a las que nos referiremos como variables independientes o predictivas. Los modelos descriptivos, en cambio, identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos. Algunas tareas de Minería de Datos que producen modelos predictivos son la clasificación y la regresión, y las que dan lugar a modelos descriptivos son el agrupamiento, las reglas de asociación y el análisis correlacional.

En las aplicaciones médicas, donde no se puede obviar la importancia de la componente temporal, las técnicas de Minería de Datos han adquirido gran relevancia⁴. Las aplicaciones de estas técnicas van desde la visualiza-

ción inteligente de grandes cantidades de datos médicos, hasta el control de calidad en centros hospitalarios.

En este proceso de Minería de Datos es importante el uso de modelos de representación y razonamiento temporal, tanto por la naturaleza intrínsecamente evolutiva del seguimiento de pacientes, como por la unidireccionalidad temporal de las relaciones causa/efecto. La Minería de Datos es un campo emergente que está adquiriendo cada vez más importancia. Desde el punto de vista médico, sirve para establecer índices pronósticos y diagnósticos que permitan obtener información útil para predecir incidencias y ajustar los protocolos de actuación clínica.

a) Extracción de patrones y aprendizaje

El aprendizaje puede tener distintas definiciones; como la mejora del comportamiento a partir de la experiencia; capacidad de predecir observaciones futuras con plausibilidad o explicar observaciones pasadas; identificación de patrones, de regularidades, existentes en la evidencia o como comprensión de información. Estas cuatro visiones se conjugan perfectamente de la siguiente manera: el aprendizaje nos permite identificar regularidades en un conjunto de observaciones. Estos patrones pueden ser utilizados para predecir observaciones futuras o explicar observaciones pasadas.

La extracción de conocimiento a partir de datos tiene como objetivo descubrir patrones que, entre otras cosas, deben ser válidos, novedosos, interesantes y, en última instancia, comprensibles. Los seres humanos tenemos una capacidad innata de ver patrones a nuestro alrededor. Las técnicas de Minería de Datos han querido emular, a estas capacidades de aprendizaje.

b) Los patrones son hipótesis

Sea como sea la presentación del problema, una de las características presente en cualquier tipo de aprendizaje y en cualquier tipo de técnica de Minería de Datos es su carácter hipotético, es decir, lo aprendido puede, en cualquier momento, ser refutado por evidencia futura. En muchos casos, los modelos no aspiran a ser modelos perfectos, sino modelos aproximados. En cualquier caso, al estar trabajando con hipótesis, es necesario realizar una evaluación de los patrones obtenidos, con el objetivo de estimar su validez y poder compararlos con otros. Por tanto, la Minería de Datos, mas que verificar patrones hipotéticos, usa los datos para encontrar estos patrones. Por lo tanto, es un proceso inductivo.

2. Almacenes de datos

El concepto de almacenes de datos nace hace más de una década ligado al sistema de información ejecutivo de una organización. Los almacenes de datos no solo se

aplican en ámbitos médicos. Al contrario, en realidad, los almacenes de datos pueden utilizarse de muy diferentes maneras, y pueden agilizar muchos procesos diferentes de análisis.

Un almacén de datos pretende proporcionar metodologías y tecnología para recopilar e integrar los datos históricos de una organización, cuyo fin es el análisis, la obtención de resúmenes e informes complejos y la extracción de conocimiento. Esta tecnología está diseñada especialmente para organizar grandes volúmenes de datos de procedencia generalmente estructurada (bases de datos relacionales, por ejemplo), aunque el concepto general es útil para la organización de pequeños conjuntos de datos en aplicaciones de Minería de Datos más modestas. Debe integrar datos externos, con lo que en realidad debe estar actualizándose frecuentemente de diferentes fuentes. La organización y el mantenimiento de esta información plantean cuestiones técnicas, fundamentalmente sobre cómo diseñar el almacén de datos, cómo cargarlo inicialmente, cómo mantenerlo y preservar su consistencia.

3. Técnicas de Minería de Datos

Dado que la Minería de Datos es un campo muy interdisciplinar, existen diferentes paradigmas detrás de las técnicas utilizadas para esta fase: técnicas de inferencia estadística, árboles de decisión, redes neuronales, inducción de reglas, aprendizaje basado en instancias, algoritmos genéticos, aprendizaje bayesiano, programación lógica inductiva y varios tipos de métodos basados en núcleos, entre otros. Cada uno de estos paradigmas incluye diferentes algoritmos y variaciones de los mismos, así como otro tipo de restricciones que hacen que la efectividad del algoritmo dependa del dominio de aplicación, no existiendo lo el método universal aplicable a todo tipo de aplicación. Puede parecer evidente que un método será mejor que otro si genera mejores modelos.

1. Técnicas bayesianas: son fáciles de usar, muy eficientes, pueden tratar muchos atributos (cientos o miles). No construyen modelos, sólo estiman una serie de probabilidades.

2. Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas: son una de las estrellas de la Minería de Datos. Son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos, los valores faltantes y el ruido. Son bastante eficientes y obtienen resultados para clasificación bastante buenos métodos es su inteligibilidad; los métodos obtenidos se pueden expresar como conjuntos de reglas. Uno de los inconvenientes de los árboles de decisión es su limi-

tada expresividad y que son inestables ante variaciones de la muestra.

3. Técnicas basadas en redes neuronales artificiales: Su ventaja principal es que, cuando están bien ajustadas, obtienen precisiones muy altas. Además son muy expresivas y permiten capturar modelos no lineales. Entre sus inconvenientes se suelen nombrar su sensibilidad a valores anómalos (aunque son robustos al ruido no extremo y a los atributos no significativos), necesitan muchos ejemplos para el aprendizaje y son relativamente lentas y, fundamentalmente, su incomprendibilidad. Son técnicas muy eficientes que permiten trabajar con datos con alta dimensionalidad. Proporcionan modelos muy precisos.

4. Fases en la realización de un proyecto de Minería de Datos

La Minería de Datos no es más que un paso esencial de un proceso más amplio cuyo objetivo es el descubrimiento de conocimiento en bases de datos. Este proceso consta de una secuencia iterativa de etapas o fases: Fases del proceso de extracción de conocimiento: Preparación de Datos, Minería de Datos, Evaluación, Difusión y Uso de Modelos.

La realización de un proyecto de Minería de Datos consta de varios pasos:

1. Análisis de las necesidades y definición del problema, en la que se establecen los objetivos de Minería de Datos. La decisión de implantar un programa de Minería de Datos y el diseño de un plan del mismo deben preceder a cualquiera de las fases. De hecho, establecer cuál es el contexto del problema, los objetivos del mismo y plasmarlos en objetivos de Minería de Datos, es previo a pararnos a pensar en recopilar y preparar los datos, realizar los modelos, evaluarlos y utilizarlos. Para implantar un programa de Minería de Datos: cuándo es conveniente, con qué grado de autosuficiencia, qué fases y cómo planificarlo, qué integración es necesaria con otros subsistemas de la organización y qué recursos materiales y humanos son necesarios.

2. Filtrado de datos: En esta fase se realiza una selección y pre-procesado de los elementos "en bruto" que contienen una base de datos.

3. Selección de variables: Aun a pesar de la fase anterior, es habitual que en los bancos de información haya un exceso de datos que dificulte su manejo.

4. Algoritmos de extracción de conocimiento: Mediante una técnica de Minería de Datos se obtie-

ne un modelo de conocimiento que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. Generalmente cada técnica de minería aplicada a un modelo de conocimiento distinto precisará un filtrado y una selección de variables previa diferente. Es en la construcción del modelo donde vemos mejor el carácter iterativo del proceso de KDD, ya que será necesario explorar modelos alternativos hasta encontrar aquel que resulte más útil para resolver nuestro problema.

5. Interpretación y evaluación: Una vez obtenido el modelo de conocimiento se debe proceder a la validación del mismo, comprobando que las conclusiones obtenidas son válidas y satisfactorias. Medir la calidad de los patrones descubiertos por un algoritmo de Minería de Datos no es un problema trivial, ya que esta medida puede atañer a varios criterios, algunos de ellos bastante subjetivos. Idealmente, los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles (es decir, inteligibles) e interesantes (útiles y novedosos). Según las aplicaciones puede interesar mejorar algún criterio y sacrificar ligeramente otro, como en el caso del diagnóstico médico que prefiere patrones comprensibles aunque su precisión no sea muy buena. Para entrenar y probar un modelo se parten los datos en dos conjuntos: el conjunto de entrenamiento (training set) y el conjunto de prueba o de test (test set). Esta separación es necesaria para garantizar que la validación de la precisión del modelo es una medida independiente. Si no se usan conjuntos diferentes de entrenamiento y prueba, la precisión del modelo será sobreestimada, es decir, tendremos estimaciones muy optimistas.

6. Fase de difusión, uso y monitorización. Una vez construido y validado el modelo puede usarse es necesario su difusión, es decir que se distribuya y se comunique a los posibles usuarios, ya sea por cauces habituales dentro de la organización, reuniones, intranet, etc. El nuevo conocimiento extraído debe integrar el know-how de la organización. También es importante medir lo bien que el modelo evoluciona. Aun cuando el modelo funcione bien debemos continuamente comprobar las prestaciones del mismo. Esto se debe principalmente a que los patrones pueden cambiar. Por lo tanto, el modelo deberá ser monitorizado, lo que significa que de tiempo en tiempo el modelo tendrá que ser re-evaluado, re-entrenado y posiblemente reconstruido completamente.

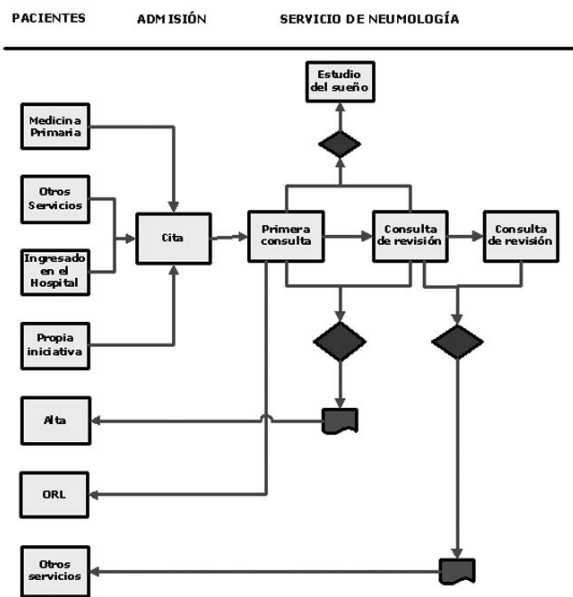
5. Análisis de las alteraciones respiratorias durante el sueño utilizando Minería de Datos

Consideramos alteraciones respiratorias durante el sueño como el término común que nos permite referirnos a un conjunto amplio de trastornos relacionados con pacientes que presentan síndrome de apnea del sueño⁶⁻⁹, EPOC¹⁰, enfermedades restrictivas torácicas o neuromusculares¹¹⁻¹⁵, o insuficiencia cardíaca¹⁶⁻¹⁷. Este tipo de pacientes es objeto de múltiples estudios en las Unidades de Sueño Neumológicas. Aparte de la historia clínica, exploración física, analítica, espirometría y gasometría o ECG estos pacientes son sometidos a estudios del sueño durante los cuales se monitorizan distintas variables respiratorias y neurológicas^{6,7}. Dentro de estos estudios hay que destacar la oximetría, poligrafía respiratoria y polisomnografía. La medicina actual extrae una información muy limitada de los estudios antes referidos¹⁸⁻¹⁹, y las herramientas software que actualmente integran los equipos de polisomnografía resultan del todo insuficientes, al detectar únicamente eventos relativamente simples de analizar como es el de la falta de flujo respiratorio asociado a la apnea del sueño. Por ello se plantea la utilización de técnicas computacionales para la realización de una explotación de los datos disponibles²⁰. Se trata fundamentalmente de involucrar un mayor número de variables fisiológicas en el estudio de los procesos fisiopatológicos que subyacen en las alteraciones cardiopulmonares del sueño, con el fin de encontrar nuevas relaciones causa/efecto que simplifiquen la generación de nuevo conocimiento a partir de la ingente cantidad de datos disponibles.

La aplicación de técnicas de Minería de Datos para el estudio de las alteraciones respiratorias durante el sueño puede plantearse en tres etapas bien diferenciadas, que se corresponden con un incremento en la dificultad y consiguiente grado de innovación planteado: en una etapa inicial se pueden aplicar técnicas convencionales de Minería de Datos, posteriormente técnicas de Minería de Datos supervisada, de modo que el médico guía la búsqueda de hallazgos de interés mediante su asociación con eventos fisiopatológicos conocidos; por último, se pueden utilizar técnicas de Minería de Datos no supervisada, de modo que se busca una agregación en la identificación de manifestaciones a partir de la definición de índices de similitud.

En todas las enfermedades anteriormente referidas, síndrome de apnea del sueño, enfermedades respiratorias obstructivas, enfermedades respiratorias restrictivas e insuficiencia cardíaca, se producen alteraciones respiratorias durante el sueño que tienen un origen multifactorial y son, por tanto, el resultado de diferentes mecanismos fisiopatológicos. La monitorización noctur-

Figura 1. Procesos implicados en la atención de pacientes con síndrome de apnea del sueño.



na de estos pacientes muestra la presencia de eventos comunes (apneas, desaturaciones, arritmias cardiacas, etc.); sin embargo, su asociación y características (duración, intensidad, etc.) muestra patrones diferentes que pueden ser específicos de cada entidad. El análisis de estos patrones puede mostrarnos datos de interés de cada una de los trastornos mencionados, desconocidos hasta ahora mediante la aplicación de la mera inspección visual del neumólogo. Podemos conocer la asociación de diferentes patrones de distribución de eventos a diferentes trastornos, facilitando un diagnóstico precoz y un tratamiento más específico, y por tanto, más eficaz, permitiendo mejorar el pronóstico de un amplio conjunto de pacientes.

Figura 2. Guía clínica general.

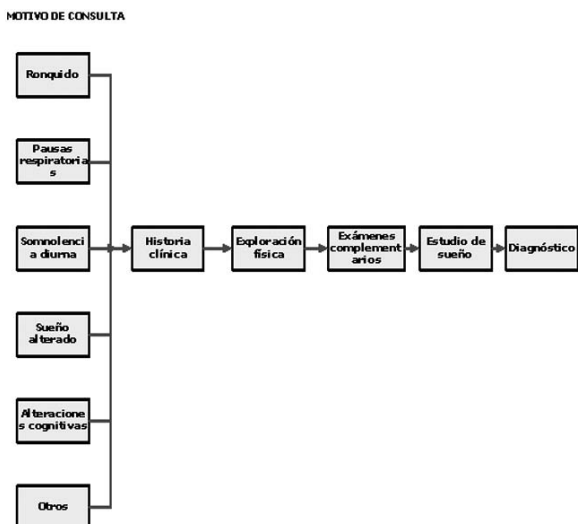
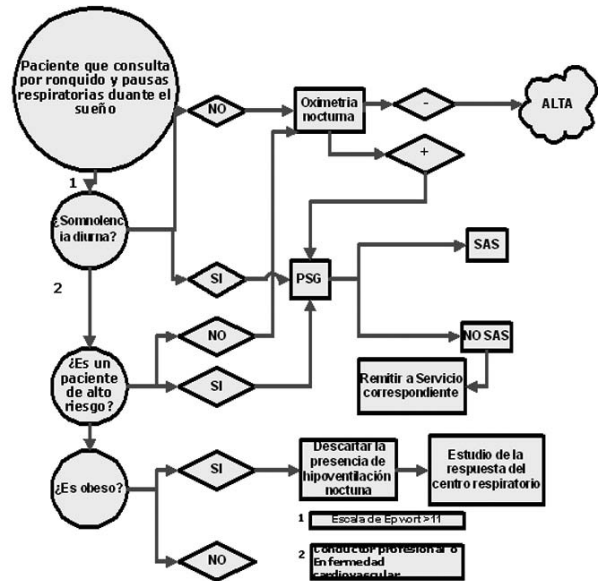


Figura 3. Guía clínica de los pacientes con síndrome de apnea del sueño.



Por otra parte, estos procesos de Minería de Datos son de aplicación en todos aquellos problemas que conlleven un análisis de un elevado número de variables con el fin de descubrir su relación con la aparición de determinados eventos. Así, podrían ser de aplicación también en la monitorización de pacientes en Unidades de Cuidados Intensivos y/o Coronarios,

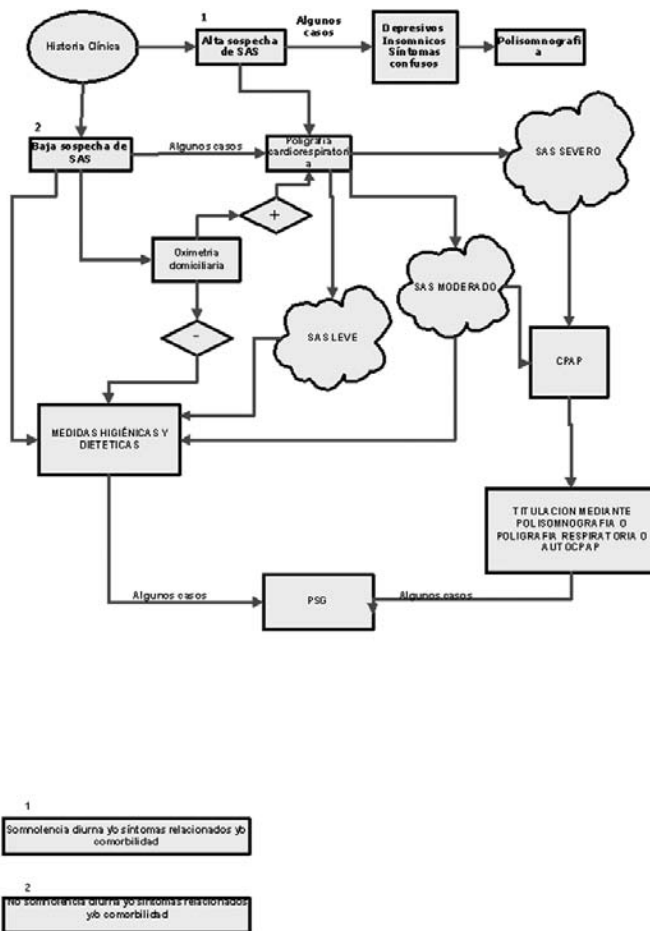
a) Fases en la aplicación de la técnica de Minería de Datos

Lo primero que debemos desarrollar es un almacén de datos teniendo en cuenta los procesos y guías clínicas implicados en la atención de los pacientes con síndrome de apnea del sueño. Hemos de delinear el análisis macroscópico y microscópico de los procesos implicados.

Resultados

Se ha desarrollado un almacén de datos teniendo en cuenta los procesos y guías clínicas implicadas en la atención de los pacientes con síndrome de apnea del sueño. Los pacientes acuden a solicitar una cita desde distintas fuentes, fundamentalmente de Medicina Primaria. Tras ser citados, son atendidos en una primera consulta. Desde esta consulta pueden ser dados de alta o remitidos a ORL para revisión de la vía aérea superior. Desde esta primera consulta también pueden ser remitidos para la realización de un estudio del sueño. Con la información de este estudio se puede decidir el alta, iniciar un tratamiento específico con las consiguientes revisiones dentro del Servicio o ser remitido a otro Servicio (Figura 1).

Figura 4. Guía clínica de los pacientes con síndrome de apnea del sueño.



La Figura 2 muestra una guía clínica general de atención de cualquier paciente, también aplicable al sujeto con síndrome de apnea del sueño.

La Figura 3 es una guía específica aplicable a los sujetos con síndrome de apnea del sueño. Se tiene en cuenta 3 factores clínicos habitualmente asociados a esta alteración, cual es la presencia de somnolencia diurna excesiva, presencia de riesgos asociados a la enfermedad o presencia de obesidad. De acuerdo con estos factores se toman distintas decisiones específicas. La Figura 4 estratifica el nivel de sospecha de síndrome de apnea del sueño y de acuerdo con ello realiza distintos procedimientos diagnósticos.

Las figuras 5, 6, 7, 8 exponen el diseño del almacén de datos. Pueden verse distintas páginas de admisión, estudios complementarios y estudios de sueño.

Las figuras 9, 10, 11, 12 (que sólo aparecen en la página web) exponen los modelos desarrollados en UML (Lenguaje Unificado de Modelado) para el SAS, EPOC, enfermedades restrictivas y fallo cardiaco.

Discusión

Los médicos recopilan datos con apreciaciones subjetivas de los enfermos y usan diferentes teorías, modelos basados en evidencias, representaciones del conocimiento y estrategias de resolución de problemas para basar sus juicios, que no se incluyen generalmente en las descripciones que son guardadas en las historias clínicas.

El aumento del volumen y variedad de información que se encuentra informatizada en bases de datos digitales ha crecido espectacularmente en la última década. Gran parte de esta información es histórica, es decir, representa transacciones o situaciones que se han producido. Aparte de su función de “memoria de la organización”, la información histórica es útil para predecir la información futura.

La Minería de Datos es un conjunto de técnicas de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos así como para predecir comportamientos futuros^{3,21}. En el ámbito médico la aplicación de la Minería de Datos tiene interés en varios campos:

1. En el ámbito clínico resulta de ayuda para la identificación y diagnóstico de patologías. Asimismo tiene importancia para el descubrimiento de posibles interrelaciones entre diversas enfermedades.
2. Al nivel de medicina preventiva, resulta de interés para la detección de pacientes con factores de riesgo para sufrir una patología.
3. Al nivel de gestión hospitalaria, se puede usar para obtener predicciones temporales que permitieran optimizar los recursos disponibles y priorizar el uso de los diversos tratamientos para una misma patología.

Los trastornos cardiopulmonares del sueño consisten en un conjunto de síntomas y signos derivados de la presencia de apneas recurrentes durante el sueño, debidas a la obstrucción parcial o completa de la vía aérea superior. Engloban un amplio espectro de alteraciones de entre las que destacamos el síndrome de apnea del sueño, la enfermedad pulmonar obstructiva crónica, las enfermedades respiratorias restrictivas y la insuficiencia cardiaca. Estos trastornos son muy habituales dentro de la práctica clínica diaria dentro de los Servicios de Neumología⁶⁻¹⁰.

El síndrome de apnea del sueño es una enfermedad común y, durante los últimos años, ha sido asociada a un incremento de morbi-mortalidad de origen cardiovascular y reconocida como un problema importante de salud

Figura 5. Página de admisión del almacén de datos.

Admisión:

Fecha de admisión: día: mes: año:

Procedencia del paciente

Paciente remitido por el médico de cabecera.

Paciente que acude por propia iniciativa.

Paciente remitido por otra unidad del CHUS.

Paciente ya ingresado en el CHUS.

Motivo de consulta

Ronquido. Tos.

Pausas respiratorias. Alteraciones cognitivas.

Somnolencia diurna. Disnea.

Sueño alterado. Dolor torácico.

Otros:

Historia clínica

Historia actual.

Antecedentes personales.

Antecedentes familiares.

SF-36.

Exploración física

Datos antropométricos.

Examen físico.

Exámenes complementarios

Analítica.

Radiografía de tórax.

ECG.

Espirometría.

Tipo de diagnóstico de admisión:

pública. La repetición de apneas durante la noche conduce a hipoxemia nocturna y fragmentación del sueño.

La enfermedad pulmonar obstructiva crónica se caracteriza por la presencia de un trastorno ventilatorio obstructivo crónico y poco reversible con frecuentes episodios de exacerbación, fundamentalmente invernales.

Dentro de las enfermedades respiratorias restrictivas hemos de distinguir las enfermedades estructurales de la caja torácica y las neuromusculares. Las enfermedades estructurales de la caja torácica condicionan una reducción de los volúmenes pulmonares junto a una disminución de la elasticidad del sistema respiratorio. Las enfermedades neuromusculares se caracterizan por

la presencia de debilidad de los músculos inspiratorios que origina hipoventilación alveolar y la aparición de fallo ventilatorio, que habitualmente es más precoz y acusado durante el período nocturno. A consecuencia de ello, se produce desorganización del sueño, con desestructuración y fragmentación del mismo, y sintomatología relacionada como cefalea matutina y somnolencia diurna²¹.

La insuficiencia cardiaca es una entidad nosológica de elevada prevalencia y con una mortalidad que supera la de muchas neoplasias conocidas actuales. El pronóstico de la insuficiencia cardiaca es bastante más desfavorable de lo que habitualmente se suele pensar. Por desgracia, los avances en los tratamientos farmacológicos

Figura 6. Página de exámenes complementarios del almacén de datos.

[Inicio](#) [Admisión](#)

Espirometría:

Fecha de realización: día: mes: año:

Registro computerizado de la espirometría:

Espirometría	Predicho	Actual	% Pred
FVC (L):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEV1 (L):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEV1/FVC (%):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEF25-75% (L/S):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEFmax (L/S):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEF75% (L/S):	<input type="text"/>	<input type="text"/>	<input type="text"/>
FEF50/FIF50 (%):	<input type="text"/>	<input type="text"/>	<input type="text"/>

[Inicio](#) [Admisión](#) [Rx-tórax](#)

Rx-tórax:

Fecha de realización: día: mes: año:

Normal:

Si anormal indicar tipo de anomalía:

[Inicio](#) [Admisión](#) [ECG](#)

ECG:

Fecha de realización: día: mes: año:

Normal:

Si anormal indicar tipo de anomalía:

Figura 7. Página de la poligrafía y oximetría del almacén de datos.

[Inicio](#) [Cuaderno de seguimiento](#) [Nivel de gravedad de la enfermedad](#)

Poligrafía screening:

Fecha de realización: día: mes: año:

Horas de análisis:

Índice de eventos/hora:

Saturación de O2 media noche:

T 90%:

T 80%:

Oximetría:

Fecha de realización: día: mes: año:

Registro computerizado de la oximetría:

Oximetría:

Fecha de realización: día: mes: año:

Registro computerizado de la oximetría:

Figura 8. Página de la polisomnografía del almacén de datos.

Inicio Cuaderno de seguimiento Nivel de alteración pulmonar

Polisomnografía:

Fecha de realización: día: 8 mes: 9 año: 2005

Tiempo de registro (TIB): min.

Tiempo de periodo de sueño (SPT):

Tiempo de latencia NREM:

Tiempo de latencia REM:

Tiempo total de sueño (TST):

Eficiencia de sueño (TST/TIB x 100):

Tiempo de cada fase de sueño:

- Fase 1: () %

- Fase 2: () %

- Fases 3 y 4: () %

- Fase REM: () %

Índice de apnea/hipopnea/hora (IAH):

Nº eventos respiratorios:

Nº apneas obstructivas:

Nº apneas centrales:

Nº apneas mixtas:

Nº Hipopneas:

Cheyne - Stokes

> 50% Cheyne - Stokes

Duración media apneas: seg.

Duración media hipopneas: seg.

Saturación media:

Saturación media durante desaturaciones:

Saturación mínima:

T90 en sueño:

T80:

T70:

Porcentaje de TST con ronquido:

han provocado pocos cambios en la supervivencia. Por otra parte, aunque la presencia de trastornos respiratorios durante el sueño en los pacientes con insuficiencia cardíaca crónica había llamado ya la atención de los médicos hace décadas, su presencia no se había estudiado sistemáticamente hasta hace unos pocos años encontrándose una correlación directa entre la severidad de dichos trastornos respiratorios durante el sueño y el grado de deterioro funcional cardíaco y la presencia de arritmias ventriculares. Algunos estudios independientes sugieren que los pacientes con apneas centrales tienen peor pronóstico que los pacientes con insuficiencia cardíaca sin apneas centrales^{22,23}.

En todas las entidades antes referidas se genera gran cantidad de información procedente de distintas vías.

En primer lugar, en relación con la historia clínica y exploración física, como ocurre en otras enfermedades. En segundo lugar, en relación con los estudios del sueño. La polisomnografía es una prueba fundamental para el diagnóstico, e indispensable para el manejo terapéutico de los enfermos con SAS. Esta prueba aporta una gran cantidad de datos cuyo análisis mediante Minería de Datos puede aportarnos distintos patrones específicos de cada una de las enfermedades.

En el caso que nos ocupa, existe una importante relación de sinergia patológica entre diversos trastornos cardiopulmonares (como ya se ha comentado, en este trabajo nos centramos en la EPOC, insuficiencia cardíaca y los trastornos respiratorios de carácter restrictivo) y las alteraciones respiratorias durante el sueño

(fundamentalmente el síndrome de apnea obstructiva del sueño). Por otra parte, dentro de la fisiología normal de la respiración durante el sueño aparecen modificaciones a nivel del calibre de las vías aéreas, de la contractilidad de la musculatura respiratoria y del control a través del sistema nervioso central. Estos cambios no producen alteraciones significativas a nivel de la respiración de un individuo normal sano, pero sí pueden ser altamente deletéreos para aquellos sujetos que presenten de base las enfermedades antes reseñadas (y mucho más si el sueño está afectado previamente por una patología tan prevalente como es el SAS).

Otro punto a tener en cuenta es que los trastornos respiratorios del sueño pueden no ser una enfermedad primaria independiente, sino secundario a patologías que no tienen en un inicio una clara afectación de la respiración como sucede en la insuficiencia cardíaca o en las neuropatías. La relación entre la insuficiencia cardíaca y trastornos respiratorios secundarios como la respiración de Cheyne-Stockes o la aparición de apnea de predominio central constituye un claro indicativo de que existe una importante relación entre la patología respiratoria y otras patologías. Asimismo se han analizado diversos parámetros no respiratorios tras la instauración de CPAP para tratamiento de SAS objetivando importante mejoría (por ej.: reducción de las cifras tensionales en pacientes con hipertensión previa de mal control con fármacos, mejoría de la sensibilidad a la insulina en pacientes diabético).

Hemos desarrollado un almacén de datos y modelos teniendo en cuenta los distintos procesos y guías clínicas implicadas en la asistencia de los pacientes con alteraciones respiratorias durante el sueño. Una vez introducidos los datos nos permitirá extraer patrones relacionados o explicativos y por tanto extraer conocimiento. Este proceso no es ajeno a la finalidad que tienen los almacenes de datos en cualquier organización. En definitiva un almacén de datos pretende proporcionar metodologías y tecnología para recopilar e integrar los datos históricos de una organización, cuyo fin es el análisis, la obtención de resúmenes e informes complejos y la extracción de conocimiento. Esta tecnología está diseñada especialmente para organizar grandes volúmenes de datos de procedencia generalmente estructurada (bases de datos relacionales, por ejemplo), aunque el concepto general es útil para la organización de pequeños conjuntos de datos en aplicaciones de Minería de Datos más modestas, como es lo que nosotros pretendemos. Debe integrar datos externos, con lo que en realidad debe estar actualizándose frecuentemente de diferentes fuentes. Es evidente, que la organización y el mantenimiento de esta información plantea cuestiones técnicas, fundamentalmente sobre cómo diseñar el almacén de datos, cómo cargarlo inicialmente, cómo mantenerlo y preservar su consistencia.

No debemos olvidar que en el momento actual los avances tecnológicos han conseguido que cada vez haya pruebas complementarias cada vez más complejas y que producen una gran cantidad de información que en algunos casos pueden incluso dificultar el manejo de la misma.

Como ya se ha comentado previamente, dada la importante prevalencia de las enfermedades a estudio, dado que existen factores propios de las patologías que interaccionan entre sí modificando el pronóstico de la enfermedad y dado el inmenso caudal de información que se puede extraer de un paciente (datos de historia clínica, de la enfermedad a estudio, de la exploración física, de las pruebas complementarias, etc.), resulta cada vez más difícil el manejo de todo el conocimiento adquirido por imposibilidad física (tiempo) y capacidad.

Por todas estas razones, si bien el uso de la Minería de Datos se pensó inicialmente como un recurso empresarial de información para mejorar su rendimiento a través de modelos predictivos, en el momento actual es una actividad que por las razones antes esgrimidas se está empezando a aplicar en el campo de la medicina. La Minería de Datos no sólo permite la creación de patrones y guías a través del procesado y análisis de un gran volumen de información, sino que además permite sacar nuevas relaciones e ideas que podrían haber quedado ignoradas debido a la ingente cantidad de datos disponibles. Además este proceso es dinámico e iterativo de tal forma que en función de los datos que nos interesen se puede retomar la operación tantas veces como se desee para sacar nuevas conclusiones.

Agradecemos al Profesor Paulo Félix Lamas, del Departamento de Ciencias de la Computación de la Facultad de Física de la Universidad de Santiago de Compostela, la ayuda prestada para la realización de este artículo.

Bibliografía

1. Fayyad U, Piatetsky-Shapiro G, Smyth P. *From data mining to knowledge discovery: an overview. In Advances in Knowledge Discovery and Data Mining, 1996.*
2. Cios K, Pedrycz W, Swiniarski R. *Data mining methods for knowledge discovery. Kluwer, 1998.*
3. Hernández Orallo J, Ramirez Quintana MJ, Ferri Ramirez C. *Introducción a la Minería de Datos. Editorial Pearson Prentice Hall. Madrid. 2004.*
4. Lavrac N. *Selected techniques for data mining in medicine. Artif Intell Med 1999; 16; 3-23.*

5. DAEDALUS: Aplicación de la Minería de Datos para el diagnóstico de accidentes cerebro-vasculares agudos (ACVA) www.daedalus.es
6. Zamarrón C, Gude F, Otero Y, Alvarez Dobaño JM, Golpe A, Rodríguez Suárez JR. Sintomatología del síndrome de apnea del sueño en población general. *Arch Bronconeumol* 1998; 34: 245 -249.
7. Zamarrón C, Hornero R, Del Campo F, Abasolo D, Alvarez D. Heart rate regularity analysis obtained from pulse oximetric recordings in the diagnosis of obstructive sleep apnea. *Sleep Breath* 2006;10: 83-9.
8. Del Campo F, Hornero R, Zamarrón C, Abasolo DE, Alvarez D. Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea. *Artif Intell Med* 2006;37:111-8.
9. Zamarrón C, Gude F, Barcala J, Rodríguez JR, Romero PV. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. *Chest* 2003;123:1567-76.
10. Togores Solivellas B, Agustí AGN. Trastornos respiratorios durante el sueño en la enfermedad pulmonar obstructiva crónica. *Arch Bronconeumol* 1996; 32: 32-39.
11. Sawicka E, Branthwaite M. Respiration during sleep in kyphoscoliosis. *Thorax* 1987;42:801-8.
12. Sivak ED, Shefner JM, Sexton I. Neuromuscular disease and hypoventilation. *Curr Opin Pulm Med*; 1999;5:355-62.
13. Smith PEM, Calverley PMA, Edwards RHT. Hypoxaemia during sleep and Duchenne muscular dystrophy. *Am Rev Respir Dis* 1988;137:884-8.
14. Ferguson KA, Strong MI, Ahmad D, George CFP. Sleep-disordered breathing in Amyotrophic lateral sclerosis. *Chest* 1996;110:664-9.
15. Shneerson I. Sleep in neuromuscular and thoracic cage disorders. *Eur Respir Mon* 1998;10:324-44.
16. Javeheri S, Parker TJ, Wexler L. Occult sleep-disordered breathing in stable congestive heart failure. *Ann Intern Med* 1995;122:487-92.
17. Staniforth AD, Kinnear WJM, Starling R, Cowley AJ. Nocturnal Desaturation in patients with stable heart failure. *Heart* 1998;79:394-99.
18. Deegan PC, McNicholas WT. Predictive value of clinical features for the obstructive sleep apnoea syndrome. *Eur Respir J* 1996; 9:117-124.
19. Croker BD, Olson LG, Saunders NA, Hensley MJ, Mckeon JM, Allen KM, Gyulay SG. Estimation of the probability of disturbed breathing during sleep before a sleep study. *Am Rev Respir Dis* 1990;142:14-18.
20. Witten & Franck, Clark, P, Boswell, R. *Data mining. Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers. 2000.
21. Kirby S, Eng P, Danter W, Francovic T, Ferguson K. Neural network prediction of obstructive sleep apnea from clinical criteria. *Chest* 1999;116:409-415.
22. Findley LJ, Zwillich CW, Ancoli-Israel S. Cheyne-Stokes breathing during sleep in patients with left ventricular heart failure. *South Med J* 1985;78:11-15.
23. Handly PJ, Zuberi-Khokhar Ns. Increased mortality associated with Cheyne-Stokes respiration in patients with congestive heart failure. *Am J Respir Crit Care Med* 1996; 153:272-6.