Desarrollo de una plataforma astroinformática para la administración y análisis inteligente de datos a gran escala

Reglas de Asociación para Líneas Moleculares

Nicolás Miranda (U. De Chile), Guillermo Cabrera (U. De Chile),
Diego Mardones (U. De Chile), Mauricio Del Solar (UTFSM),
Marcelo Mendoza (UTFSM), Karim Pichara (PUC),
Ricardo Contreras (U. de Concepción), Neil Nagar (U. de Concepción),
Victor Parada (U. de Santiago)

Santiago, Chile, 2 de mayo de 2014

Resumen

Dentro del ámbito de la espectroscopía astronómica, los cubos de datos tipo ALMA ofrecen nuevas oportunidades para realizar detección de lineas espectrales de manera automática, dadas las considerables cantidades de pixeles presentes en cada observación, en diferentes rangos de frecuencias dentro de las bandas milimétricas/sub-milimétricas. Esto último supone nuevos desafíos para las herramientas existentes de minería de datos. El presente trabajo apunta a estudiar y aplicar una de estas herramientas, el aprendizaje mediante reglas de asociación, a los cubos de datos tipo ALMA con el fin de encontrar nuevas relaciones entre líneas desconocidas y transiciones moleculares. En este documento, en particular, se presenta el estado del arte en aprendizaje mediante reglas de asociación y los requerimientos del software a desarrollar en el proyecto.

Palabras Claves: Data Mining, Association Analysis, Association Rules, Unsupervised Learning.

1. Resumen Ejecutivo

En este proyecto se busca estudiar algoritmos de *Association Rule Learning (ARL)* con el fin de realizar asociaciones entre transiciones moleculares presentes en espectros de frecuencia obtenidos a partir de cubos de datos tipo *ALMA*. De esta forma, se podrá, entre otras cosas, asociar transiciones previamente no identificadas a moleculas conocidas, agrupar moleculas relacionadas en lugares del espacio en particular, y facilitar la identificación de líneas moleculares desconocidas.

Posteriormente, ya efectuada la implementación de los algoritmos, se realizarán experimentaciones sobre datos simulados; con el fin de evaluar, mediante una serie de métricas, cuál de los algoritmos estudiados se ajusta mejor a este problema en particular. Una vez seleccionado el o los algoritmos más apropiados, se procederá a desarrollar el resto de las características del sistema, incluyendo interfaces de usuario y de aplicación. Esto con el fin de ensamblarse a sistemas que realicen la detección previa de transiciones moleculares y de facilitar el acoplamiento a entornos de observatorios virtuales.

Índice

2. Introducción

En los últimos tiempos, y en gran parte debido al explosivo desarrollo tecnológico, han surgido numerosos campos en los cuales se ha requerido el uso de procesamiento masivo de datos e inteligencia computacional con el fin de automatizar y auxiliar el proceso de generación de nuevo conocimiento. La astronomía es, sin lugar a dudas, uno de ellos.

El *Atacama Large Milimiter/sub-milimiter Array* (ALMA) es un interferómetro radio-astronómico que consiste de 66 antenas que observan el espacio en las bandas milimétricas y sub-milimétricas del espectro electromagnético. Ubicado en el desierto de Atacama, en el norte del país, es parte de uno de los proyectos científicos más importantes del último tiempo a nivel nacional; en el cual se ha hecho uso de tecnologías de punta por parte de investigadores, ingenieros, y técnicos expertos en computación de alto rendimiento, redes de fibra óptica, Machine Learning, minería de datos, entre otros.

Gran parte de los datos obtenidos desde ALMA son guardados en estructuras de datos llamadas cubos de datos tipo ALMA (o *ALMA Data Cubes*), que contienen información de distintos puntos de observación del cielo a distintas frequencias.

Los espectros de frecuencia son una forma de representar la intensidad de la radiación electromagnética, recibida desde un punto del espacio, en un cierto rango de frecuencias. Estos contienen puntos altos de intensidad en ciertas frecuencias en las cuales se sabe que una cierta molécula o átomo conocido efectúa una trancisión cuántica. Por lo tanto, mediante reconocer e identificar estos puntos altos, o *peaks*, se puede saber las transiciónes moleculares que ocurrieron en el objeto del que proviene la radiación electromagnética. Como, a su vez, se sabe de antemano a qué frecuencia específicamente se realiza la transición cuántica de átomos o moléculas conocidas, puede inferirse cuáles son los átomos o moléculas presentes en el objeto de origen1.

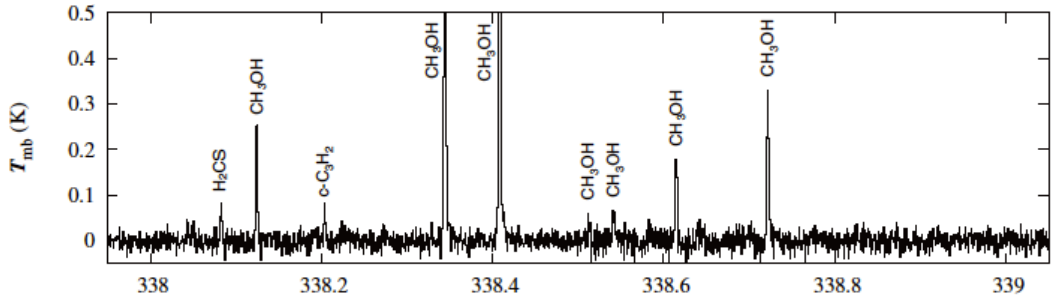

Figura 1: Espectro de frecuencias. Pueden apreciarse claramente los puntos de alta intensidad, que corresponden a transacciones cuánticas de distintas moléculas (obtenido de Watanabe et al. 2012).

A su vez, los cubos de datos tipo ALMA, dentro de su estructura, contienen valores indexados en tres coordenadas2. Dos de las coordenadas son espaciales, y corresponden al equivalente a una imagen normal de dos dimensiones, en el sentido que describen puntos del cielo (o del espacio observable desde la tierra). El tercer eje de coordenadas corresponde al rango de frecuencias del espectro electrómagnético en el que se está observando. Por lo tanto, si se fijan las dos coordenadas espaciales (vale decir, si se fija un punto de observación en el espacio) y se extraen todos los valores en la tercera coordenada de aquel punto, se obtiene el espectro de frecuencias observado en ese punto del espacio, como se mostró en la Figura1.

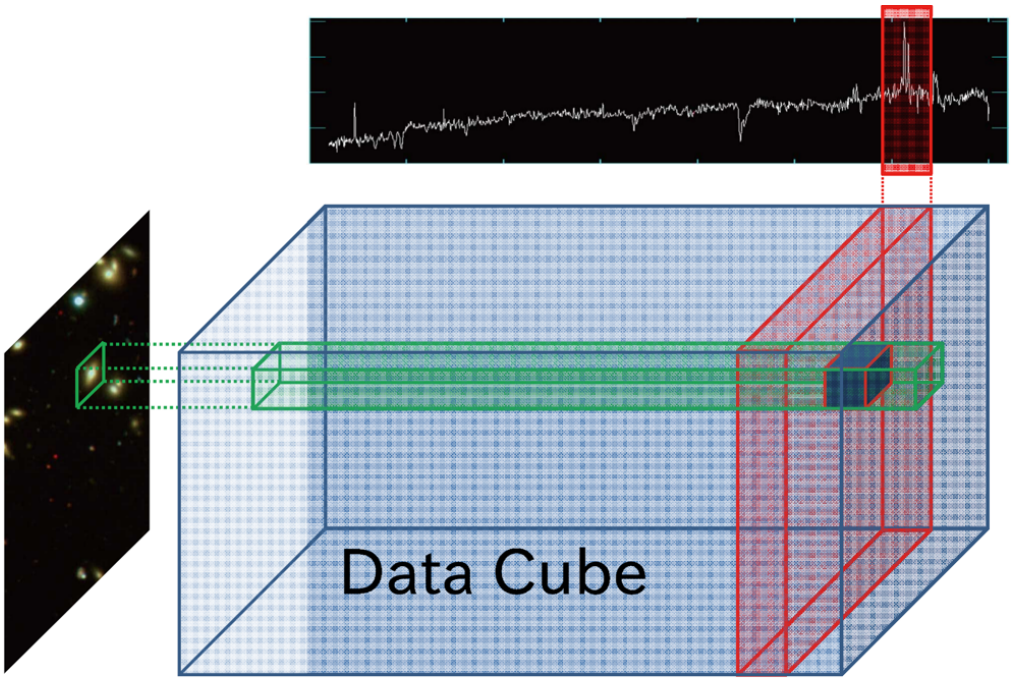

Figura 2: Estructura de un cubo de datos tipo ALMA. Cada punto del espacio tiene asocuado un espectro de frecuencias. (Imagen obtenida de inspirehep.com)

Los cubos de datos tipo ALMA, por tanto, contienen información de los espectros de frecuencia observados en todos los puntos de un sector dado del espacio. Todos los espectros presentes en un cubo se encuentran en un mismo rango de frecuencia. Sin embargo, distintos cubos de datos pueden tener observaciones hechas en distintos rangos de frecuencia entre sí.

A partir de ALMA se generan enormes cantidades de datos, los cuales, debido a su gran tamaño, necesariamente deben procesarse mediante sistemas automatizados de extracción y análisis con el fin de facilitar a los investigadores el extraer información útil a partir de estos. La mayoría de estas herramientas se encuentran dentro de las áreas de investigación en minería de datos y Machine Learning; disciplinas de la computación que han tenido un gran auge en el último tiempo.

2.1. Reglas de asociación

Dentro del área del aprendizaje computacional automatizado, o Machine Learning, existe una técnica que ha sido ampliamente utilizada e investigada desde su concepción. Se trata del aprendizaje mediante reglas de asociación, o *Association Rule Learning* (ARL); la cual se creó con el fin de identificar relaciones entre los productos preferidos por los consumidores en sistemas de punto de venta como supermercados, tiendas de venta al detalle, etc.

La intuición es que, si se posee una base de datos con transacciones, donde cada una de estad posee un

cierto conjunto de ítems que un cliente en particular ha comprado, con la ayuda de un algoritmo puede encontrarse una serie de reglas que indiquen relaciones entre las compras de ciertos ítems en particular. Un ejemplo de regla (bastante intuitiva, por lo demás) sería: "En el 90 % de las transacciones en que se compró pan y mantequilla también se compró lecpercenthe". Los algoritmos de ARL permiten obtener relaciones simples, como la del ejemplo, y otras mucho más difíciles de deducir por otros medios.

Formarmente. se considera un conjunto de variables $\mathcal{X} = \{X_j\}_{j=1}^{p}$. Usualmente, estas variables se consideran como binarias: $X_j \in \{0, 1\}$. Se define, entonces, un conjunto de N transacciones $\mathcal{D} = \{t_i\}_{i=1}^{N}$, donde $t_i = \{x_{i,j}\}_{j=1}^{p}$, y

$$x_{i,j} = \begin{cases} 0 & \text{si el ítem } j \text{ es parte de la transacción } i \\ 1 & \text{de lo contrario} \end{cases}$$

El objetivo principal del análisis de reglas de asociación es obtener aquellos valores conjuntos de valores $(X_1, X_2, ..., X_p)$ que aparezcan de manera más frequente en el conjunto de datos.

A partir de estos, se generan reglas de asociación de la forma

$$I \implies X_j | c$$

donde $I \subset \mathcal{X}$, $X_j \notin I$ y c es la confianza de la regla, o la razón entre el número de transacciones en $\mathcal{D}$ que contienen a $I \cup X_j$ y el número de transacciones que contienen a I. A su vez, la razón entre el número de transacciones que contienen a $I \cup X_j$ y el número total de transacciones se conoce como el *soporte* de la regla de asociación[1].

3. Metodología de Trabajo

El objetivo principal del presente proyecto es implementar una herramienta de ARL (*Association Rule Learning*) para asociar líneas de espectoscopía astronómica a transiciones moleculares, a partir de cubos de datos tipo ALMA. Para ello, esta herramienta recibirá conjuntos de líneas, detectadas con anterioridad, presentes en distintos espectros de frecuencia correspondientes a puntos en el espacio o, de forma equivalente, a pixeles en los cubos de datos.

Durante el transcurso del trabajo se estudiarán distintos algoritmos de ARL, y se llevará a cabo un proceso de evaluación de los mismos. Se implementará y posteriormente se experimentará, en particular, con aquellos que mejor se acomoden a los casos de uso en el contexto de una plataforma en la cual se opera sobre datos provenientes de diversas fuentes. De estos, se seleccionarán aquellos que muestren un mejor desempeño y eficiencia en grandes volúmenes de datos.

En primera instancia, para probar los algoritmos de ARL subyacentes se utilizarán conjuntos de líneas correspondientes a espectros de frecuencia simulados; con datos generados de manera pseudo-aleatoria y otros generados bajo condiciones específicas que por sus características (como el tipo de moléculas presentes) aseguren similitud con datos reales obtenidos a partir de la observación de alguna región del espacio.

Una vez que se comparen los resultados utilizando distintas métricas definidas con anterioridad, los métodos pasarán a ser probados sobre datos provenientes de cubos de datos tipo ALMA reales, publicados por ALMA.

4. Estado del Arte

4.1. Descripción General

El aprendizaje mediante reglas de asociación, o *Association Rule learning (ARL)*, es sin lugar a dudas uno de los métodos más populares y mejor estudiados dentro del área de la Minería de Datos y Machine Learning. Agrawal et al., en su artículo seminal sobre el método[1], sentaron las bases del uso de reglas de asociación con el fin de descubrir relaciones entre productos en una base de datos a gran escala de transacciones, en particular, registradas en sistemas de puntos de venta presentes en supermercados. En este mismo se introdujo el algoritmo *Apriori*, que es el más utilizado para obtener reglas de asociación, haciendo uso de las medidas de *confianza* y *soporte*.

Posteriormente, Agrawal et al. presentaron el algoritmo *AprioriTid*[3], cuyas mejores características fueron combinadas con el algoritmo *Apriori* para crear el algoritmo *AprioriHybrid*, de orden de complejidad lineal en el número de transacciones. Luego se han realizado más desarrollos en ARL orientado a transacciones secuenciales de clientes de puntos de ventas[2].

Savasere et al. introdujeron el algoritmo *Partition*[31] con el fin de extraer reglas de asociación en base de datos, el cual presenta reducciones en las operaciones de la CPU y de entrada/salida, y que además facilita la paralelización. Posteriormente se creó el algoritmo *Dynamic Itemset Counting (DIC)*[6], que realiza menos lecturas sobre los datos que los algoritmos previos, y que utiliza la métrica de *Convicción* a la hora de generar reglas de asociación. Luego, Park et al. presentaron un algoritmo que hace uso de funciones de Hashing con el fin de generar reglas candidatas[26]. Se han realizado, también, adaptaciones de los algoritmos previos con el fin de realizar ARL en datos de tipo cuantitativo[33].

Esfuerzos posteriores se han realizado con el fin de profundizar en los fundamentos teóricos subyacentes en ARL (e.g. definiendo el conjunto de posibles ítemes como una estructura algebráica llamada *retículo*)[35], y con el fin de extender la noción de reglas de asociación a correlaciones[5].

Más recientemente, Han et al. introdujeron el uso de una estructura de datos llamada *Frequent Pattern Tree*[19] en la extracción de reglas de asociación a partir de conjuntos de transacciones. Luego de esto, se han hecho numerosas implementaciones y optimizaciones a los algoritmos más utilizados en ARL, como, por ejemplo, el algoritmo Apriori[4]; así como implementaciones que facilitan el mantener la privacidad de cada una de las fuentes de datos que participan en el proceso[15].

Desde su concepción, el método de ARL ha sido aplicado en numerosas áreas, tales como la detección de intrusiones[24] y anomalías[27][9], educación[29][30], química[13], privacidad de datos[17], búsqueda en la web[16], tráfico en redes[14], computación social[25], búsqueda semántica[11], biología[23][8], salud[21][10], medios de comunicación[12][22], y la investigación forense[20]. Junto con esto, se han realizado numerosas investigaciones sobre el estado actual de ARL y sus posibles desarrollos a futuro dentro del marco de métodos automatizados de generación de conocimiento[18].

4.2. Soluciones Actuales

Si bien existen numerosos esfuerzos por utilizar minería de datos y Machine Learning en diversos ámbitos de la astronomía (en particular, en detección, clasificación y caracterización de líneas moleculares

en espectros de emisión[32]), hasta la fecha no se ha propuesto abiertamente el uso de ARL sobre datos extraídos de espectros de frecuencia.

Sin embargo, se han realizado avances en ampliar los conceptos subyacentes en ARL con el fin de aplicar el método en campos más diversos[5]. Específicamente, una rama de investigación ha desarrollado lo que se denomina *Weighted Association Rule Learning*[34][7]. Este método permite asociar medidas de interés arbitrario a priori a ciertos conjuntos de datos. Si bien esto hace que se pierdan propiedades de clausura que son útiles a la hora de generar algoritmos eficientes, también permite trabajar con distintos conjuntos de transacciones sin que las reglas generadas estos dependan exclusivamente de su soporte u otras medidas estándar.

Esto último, junto con todo el cuerpo de investigación en ARL mencionado anteriormente, suponen una base lo suficientemente fuerte como para abarcar exitosamente el problema en el presente proyecto.

5. Especificación de Requerimientos

5.1. Especificación de Requerimientos

A continuación se enuncian los requerimientos del sistema:

1. **Obtener reglas de asociación entre líneas de emisión espectrales [esencial].**
   El sistema debe generar reglas de asosiación entre líneas de emisión presentes en espectros de frecuencia, independientemente de si estos pertenecen a una misma o a distintas moléculas o átomos, o si no han sido aun identificadas.

2. **Asociar líneas presentes en espectros de distinto rango de frecuencia [esencial].**
   Incluso si las líneas de emisión se encuentran en espectros cuyos rangos no son idénticos, entonces el sistema debe encontrar una forma de extraer reglas de asociación de todos ellos por igual.

3. **Permitir al usuario verificar reglas generadas utilizando criterios tanto estadísticos como astrofísicos [esencial].**
   Una vez extraídas las reglas de asociación, el usuario debe poder revisarlas y verificar su validez mediante métricas y criterios no necesariamente utilizadas por los algoritmos de ARL.

4. **Definir una interfaz de comunicación con sistema de detección de líneas de emisión [esencial].**
   El sistema operará sobre conjuntos de líneas de emisión que han sido ya identificadas con anterioridad dentro de sus respectivos espectros de frecuencia. Por lo tanto, este debe haber una interfaz bien definida entre este y un sistema de detección de líneas de emisión.

5. **El sistema debe ser ejecutable en un ambiente de computación de alto rendimiento [deseable].**

6. **El sistema debe ser compatible con plataformas de observatorios virtuales [deseable].**

7. **Implementar una interfaz gráfica de usuario [opcional].**

5.2. Casos de Uso

En la Figura 3 se muestra un diagrama con los casos de uso preliminares del sistema a desarrollar.

5.2.1. Actores

Para este sistema existe solo un tipo de actor, dado que todos los usuarios finales tendrán acceso a las mismas funcionalidades. Este usuario será el encargado de seleccionar el conjunto de datos que quiere ingresar al sistema, en forma de vectores de líneas moleculares. Cada vector poseerá las líneas identificadas en un espectro de frecuencia en particular. Este usuario ingresará estos datos al sistema y luego seleccionará los parámetros de detección de reglas que desee. Una vez ejecutados los algoritmos

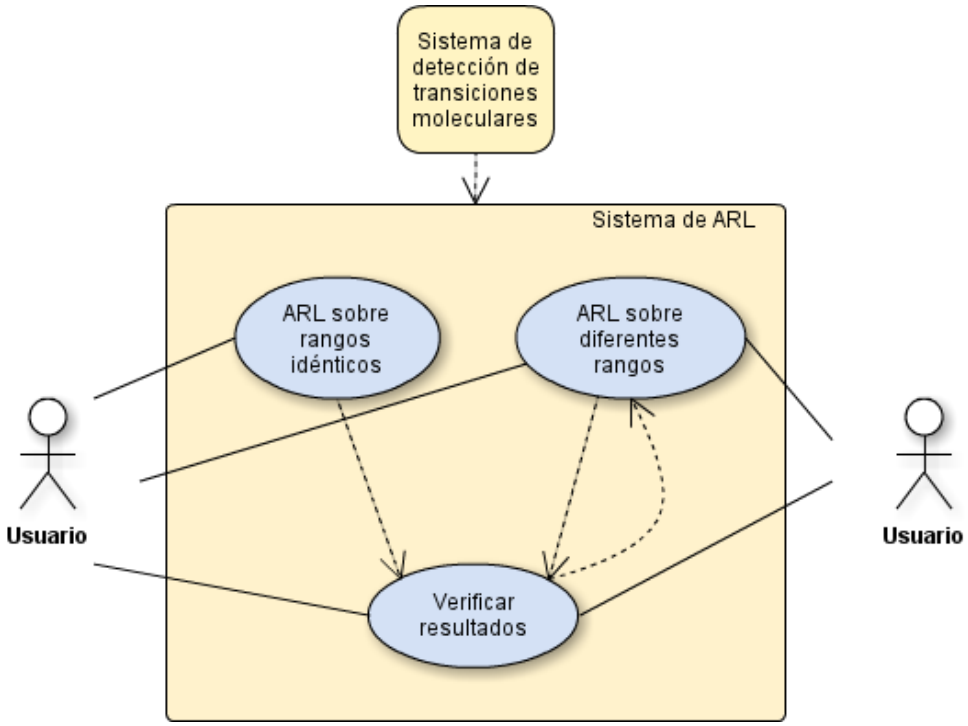

Figura 3: Diagrama de casos de uso del sistema.

correspondientes, el usuario podrá observar las reglas generadas y, si así lo desea, ajustar nuevamente los parámetros para obtener mejores resultados sobre el mismo conjunto de datos.

Posteriormente, el mismo u otro usuario podrá verificar los resultados obtenidos en una sesión de ARL anterior y ajustar los parámetros de búsqueda a su agrado para luego volver a correr los algoritmos sobre los mismos iniciales.

5.2.2. Casos

En la Tabla 3 se muestra una descripción detallada de los casos de uso y se indica, de ser así, a qué requerimiento está asociado.

ID	Caso de uso	Descripción	Tipo	Ref.
1	Ejecutar ARL sobre datos con rangos espectrales idénticos	El usuario selecciona un conjunto de espectros extraídos a partir de un sólo cubo de datos o de más de uno, pero siempre en rangos de frecuencia idénticos, selecciona los parámetros adecuados y ejecuta los algoritmos de ARL sobre ellos.	Esencial	1,2,4
2	Ejecutar ARL sobre datos con distintos rangos espectrales	El usuario selecciona un conjunto de espectros extraídos a partir de dos o más cubos de datos con distintos rangos de frecuencia, selecciona los parámetros adecuados y ejecuta los algoritmos de ARL sobre ellos.	Esencial	1,2,3
3	Verificar resultados	El usuario examina las reglas generadas por los algoritmos determinando si le dan información valiosa o si, en su defecto, necesita volver a ejecutarlos con distintos parámetros	Esencial	3

Cuadro 1: Diagrama de casos de uso del sistema.

Referencias

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*, volume 22, pages 207–216. ACM, 1993.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, pages 3–14. IEEE, 1995.

[3] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499, 1994.

[4] Ferenc Bodon. A fast apriori implementation. In *Proceedings of the IEEE ICDM workshop on frequent itemset mining implementations (FIMI'03)*, volume 90, 2010.

[5] Sergey Brin, Rajeev Motwani, and Craig Silverstein. Beyond market baskets: Generalizing association rules to correlations. In *ACM SIGMOD Record*, volume 26, pages 265–276. ACM, 1997.

[6] Sergey Brin, Rajeev Motwani, Jeffrey D Ullman, and Shalom Tsur. Dynamic itemset counting and implication rules for market basket data. In *ACM SIGMOD Record*, volume 26, pages 255–264. ACM, 1997.

[7] Chun Hing Cai, Ada Wai-Chee Fu, CH Cheng, and WW Kwong. Mining association rules with weighted items. In *Database Engineering and Applications Symposium, 1998. Proceedings. IDEAS'98. International*, pages 68–77. IEEE, 1998.

[8] Pedro Carmona-Saez, Monica Chagoyen, Francisco Tirado, Jose M Carazo, and Alberto Pascual-Montano. Genecodis: a web-based tool for finding significant concurrent annotations in gene lists. *Genome biology*, 8(1):R3, 2007.

[9] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

[10] R Chaves, JM Górriz, J Ramírez, IA Illán, D Salas-Gonzalez, and M Gómez-Río. Efficient mining of association rules for the early diagnosis of alzheimer's disease. *Physics in medicine and biology*, 56(18):6047, 2011.

[11] Edith Cohen, Amos Fiat, and Haim Kaplan. Associative search in peer to peer networks: Harnessing latent semantics. *Computer Networks*, 51(8):1861–1881, 2007.

[12] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[13] Luc Dehaspe, Hannu Toivonen, and Ross D King. Finding frequent substructures in chemical compounds. In *KDD*, volume 98, page 1998, 1998.

[14] Cristian Estan, Stefan Savage, and George Varghese. Automatically inferring patterns of resource consumption in network traffic. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 137–148. ACM, 2003.

[15] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke. Privacy preserving mining of association rules. *Information Systems*, 29(4):343–364, 2004.

[16] Paolo Ferragina and Antonio Gulli. A personalized search engine based on web-snippet hierarchical clustering. *Software: Practice and Experience*, 38(2):189–225, 2008.

[17] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 121–132. ACM, 2008.

[18] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1):55–86, 2007.

[19] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data mining and knowledge discovery*, 8(1):53–87, 2004.

[20] Farkhund Iqbal, Hamad Binsalleeh, Benjamin Fung, and Mourad Debbabi. A unified data mining solution for authorship analysis in anonymous textual communications. *Information Sciences*, 231:98–112, 2013.

[21] Murat Karabatak and M Cevdet Ince. An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, 36(2):3465–3469, 2009.

[22] Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. Media meets semantic web–how the bbc uses dbpedia and linked data to make connections. In *The semantic web: research and applications*, pages 723–737. Springer, 2009.

[23] Stefan Kramer, Luc De Raedt, and Christoph Helma. Molecular feature mining in hiv data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 136–143. ACM, 2001.

[24] Wenke Lee and Salvatore J Stolfo. A framework for constructing features and models for intrusion detection systems. *ACM transactions on Information and system security (TiSSEC)*, 3(4):227–261, 2000.

[25] Xin Li, Lei Guo, and Yihong Eric Zhao. Tag-based social interest discovery. In *Proceedings of the 17th international conference on World Wide Web*, pages 675–684. ACM, 2008.

[26] Jong Soo Park, Ming-Syan Chen, and Philip S Yu. *An effective hash-based algorithm for mining association rules*, volume 24. ACM, 1995.

[27] Animesh Patcha and Jung-Min Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12):3448–3470, 2007.

[28] Anthony J Remijan and Andrew J Markwick-Kemper. Splatalogue: Database for astronomical spectroscopy. 2008.

[29] Cristóbal Romero and Sebastian Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, 2007.

[30] Cristóbal Romero, Sebastián Ventura, and Enrique García. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1):368–384, 2008.

[31] Ashok Savasere, Edward Robert Omiecinski, and Shamkant B Navathe. An efficient algorithm for mining association rules in large databases. 1995.

[32] Petr Škoda and Jaroslav Vážnỳ. Searching of new emission-line stars using the astroinformatics approach. *arXiv preprint arXiv:1112.2775*, 2011.

[33] Ramakrishnan Srikant and Rakesh Agrawal. Mining quantitative association rules in large relational tables. In *ACM SIGMOD Record*, volume 25, pages 1–12. ACM, 1996.

[34] Wei Wang, Jiong Yang, and Philip S Yu. Efficient mining of weighted association rules (war). In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 270–274. ACM, 2000.

[35] Mohammed Javeed Zaki and Mitsunori Ogihara. Theoretical foundations of association rules. In *3rd ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pages 71–78. Citeseer, 1998.