

Conference Paper

Parallel Generation of Association Rules – Use Case: Psychometric Evaluation

Generación Paralela de Reglas de Asociación – Caso de Uso: Evaluación Psicométrica

Edward Hinojosa C., Hugo C. Rucano A., Rosa Y. G. Paccotacya Y.,
 Roni Apaza A., Israel S. Pancca M., Diego A. Iquira B., and Robin Sánchez Y.

Universidad Nacional de San Agustín, Arequipa, Perú

Abstract

Mining of association rules consists of determining relations among variables in the form of rules in large databases, one of the algorithms most used to perform this task is the Apriori algorithm. This article proposes the parallelization of the Apriori algorithm applied to the Cattell's questionnaire Sixteen Personality Factor, known as the 16PF questionnaire, to generate association rules among different questions. The use case used has a database that contains 49,150 questionnaires answered, with 163 questions per questionnaire. Results obtained show four times less execution time between the proposed parallel algorithm and the serial algorithm, in addition, the results show associations with confidence values greater than 97% (with two variables) and greater than 98% (with three variables), which it would be expected from the point of view of traditional psychology.

Resumen

La minería de reglas de asociación consiste en determinar relaciones entre variables en forma de reglas en grandes bases de datos, uno de los algoritmos más utilizados para realizar esta tarea es el algoritmo Apriori. Este trabajo propone la paralelización del algoritmo Apriori aplicado al cuestionario de personalidad de los 16 factores que ideó Raymond Cattell, conocido como el cuestionario 16PF, para generar reglas de asociación entre las diferentes preguntas. El caso de uso utilizado cuenta con una base de datos que contiene 49150 cuestionarios respondidos, con 163 preguntas por cuestionario. Los resultados obtenidos muestran un tiempo de ejecución 4 veces menor del algoritmo paralelo propuesto al algoritmo serial, además, los resultados muestran asociaciones con valores de confianza mayores a 97% (con 2 variables) y mayores al 98% (con 3 variables), lo que se esperaría desde el punto de vista de la psicología tradicional.

Keywords: rules of association, Apriori algorithm, parallelization, personality questionnaire

Palabras clave: Reglas de Asociación, Algoritmo Apriori, paralelización, cuestionario de personalidad

Corresponding Author:
 Edward Hinojosa C.
 ehinojosa@unsa.edu.pe

Received: 4 December 2018
 Accepted: 5 December 2018
 Published: 27 December 2018

Publishing services provided by
 Knowledge E

© Edward Hinojosa C. et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review under the responsibility of the SIIPRIN-CITEGC Conference Committee.

 OPEN ACCESS

1. Introducción

La minería de reglas de asociación o minería de reglas de clasificación asociativa, consiste en diferentes técnicas y/o métodos para determinar relaciones entre variables en grandes bases de datos, por ejemplo, el algoritmo Apriori [4], que se usa para extraer conjuntos de elementos frecuentes y reglas de asociación relevantes.

La minería de reglas de asociación es utilizada en diferentes áreas, por ejemplo, visión artificial, análisis de flujo de datos, bioinformática, recuperación de información, sistemas de administración de bases de datos, análisis de datos de ventas, análisis de comportamiento del cliente en banca, seguros, atención de la salud, la psicología, entre otras [1].

Los algoritmos de minería de reglas de asociación más conocidos son FP-Growth[2], Eclat[3] y Apriori[4]. El primero usa un FP-Tree (Frequent Pattern Tree o Árbol de Patrones Frecuentes), que genera conjuntos de elementos frecuentes escaneando la base de datos solo dos veces sin ningún proceso de iteración para la generación de candidatos. El primer escaneo es el proceso de construcción de FP-Tree y el siguiente es la generación de patrones frecuentes del FP-Tree [5].

Por otro lado, el proceso de generación de candidatos en Apriori y Eclat es similar, excepto en la forma en que representan los datos de candidatos y transacciones [7].

La mayoría de los algoritmos de minería de reglas de asociación utilizan el procesamiento en serie, debido al alto grado de dependencia de datos [6], por lo tanto, estos algoritmos son potencialmente incapaces de manejar los conjuntos de datos actuales crecientes, así como, tomar una gran cantidad de tiempo de ejecución [8].

Por lo tanto, las comparaciones de rendimiento para procesos seriales generalmente muestran que el algoritmo FP-Growth es más rápido que los algoritmos Apriori y Eclat, sin embargo, cuando el soporte mínimo es alto o en una plataforma multi-core, Apriori supera a FP-Growth [9].

El análisis del cuestionario de personalidad ha sido elegido como caso de estudio, debido a la importancia de saber cómo funciona nuestra mente y predecir el comportamiento de las personas. Por ejemplo, en el campo de la psicología, las personas, sus personalidades y comportamientos se dividen en diferentes tipos, por esa razón, existen diferentes modelos y una variedad de teorías y cuestionarios que contribuyen a entender la psicología humana. Sin embargo, un problema aún en abierto es predecir el comportamiento de una persona en el futuro [13]. Una propuesta para este problema fue desarrollada en [10] donde muestra una forma de obtener estas predicciones sin

necesidad de abstracción alguna sobre la mente o su funcionamiento interno, aplicando minería de reglas de asociación.

Este artículo propone la paralelización del algoritmo Apriori para la generación de reglas de asociación basado en OpenMP (Open Multi-Processing) para obtener asociaciones entre las variables del cuestionario 16PF.

En la siguiente sección se presentan algunos conceptos relacionados a la psicología de la personalidad y las teorías de la personalidad. En la sección 3 se explica el algoritmo paralelo propuesto. En la sección 4 se analizan los resultados obtenidos en función del tiempo de ejecución del algoritmo y las asociaciones descubiertas usando el algoritmo paralelo propuesto sobre la base de datos del caso de uso. Finalmente, en la sección 5 se muestran las conclusiones de artículo, así como los trabajos futuros.

2. Psicología de la Personalidad

2.1. Psicología de la Personalidad

La personalidad se define como un conjunto dinámico y organizado de características que posee una persona, las cuales influyen de manera única en sus cogniciones, emociones, motivaciones y comportamientos en diversas situaciones. La palabra personalidad proviene del latín persona, que significa máscara; en el teatro del mundo latino antiguo, la máscara no se usaba como un dispositivo de la trama para disfrazar la identidad de un personaje, sino era un dispositivo empleado para representar o tipificar su carácter [10].

Allport argumenta que la psicología de la personalidad puede [18]:

- Predecir el comportamiento de una persona en base a sus características individuales;
- Descubrir rasgos comunes;
- Desarrollar leyes generales sobre cómo se produce la singularidad de un individuo;
- Entre otros.

2.2. Teorías de la Personalidad

El estudio de la personalidad se basa en la idea esencial de que todas las personas son similares en algunos aspectos, pero diferentes en otros [14, 15]. La personalidad

es ese patrón de pensamientos, sentimientos y comportamientos característicos que distingue a una persona de otra y que persiste en el tiempo y las situaciones [14]. Existen diferentes teorías relacionadas a la personalidad, a continuación mostramos resumidamente alguna de ellas:

2.2.1. Teorías de rasgos

Según el Manual de Diagnóstico y Estadística de la Asociación Americana de Psiquiatría, los rasgos de personalidad son patrones perdurables de percepción, relación y pensamiento sobre el medio ambiente y sobre uno mismo que se exhiben en una amplia gama de contextos sociales y personales. Los rasgos son estables, únicos e influyen en el comportamiento [15]. La investigación de Raymond Cattell [11, 12] propagaba una estructura de personalidad de dos niveles con dieciséis factores primarios (16 factores de personalidad – 16PF) y cinco factores secundarios. Para Cattell, la personalidad misma se definió en términos de predicción del comportamiento; además, definió la personalidad como aquello que permite una predicción de lo que una persona hará en una situación dada.

2.2.2. Teorías de tipos

El tipo de personalidad se refiere a la clasificación psicológica de diferentes tipos de personas [20]. Por ejemplo, introvertidos y extrovertidos.

2.2.3. Teorías psicoanalíticas

Las teorías psicoanalíticas explican el comportamiento humano en términos de la interacción de varios componentes de la personalidad, liderado por Freud, quien propuso que la energía psíquica podría convertirse en comportamiento [16].

2.2.4. Teorías conductistas

Los conductistas explican la personalidad en términos de los efectos que los estímulos externos tienen sobre el comportamiento [17].

2.2.5. Teorías cognitivas sociales

La teoría cognitiva explica el comportamiento centrado en las cogniciones sobre el mundo, especialmente aquellas sobre otras personas [19].

3. Cuestionario de personalidad de los 16 factores de Raymond Cattell

El cuestionario 16PF fue desarrollado originalmente en la década de 1940 por Raymond Cattell para medir los factores principales de la personalidad normal. El instrumento proporciona puntajes en los 16 rasgos de primer orden, 5 dimensiones globales (extroversión, ansiedad, dureza mental, independencia y autocontrol) y 3 escalas sobre estilos de respuesta (gestión de impresiones, respuestas infrecuentes y consentimiento) [22].

A través del análisis de los factores, Cattell identificó lo que él llamó rasgos superficiales y originales. Los rasgos superficiales representan grupos de variables correlacionadas y los rasgos originales representan la estructura subyacente de la personalidad. Cattell consideraba que los rasgos originales eran mucho más importantes para comprender la personalidad que los rasgos superficiales. Los rasgos originales identificados se convirtieron en la base principal del modelo de 16 PF y su objetivo es medir la personalidad en función de 16 rasgos originales (Tabla 1).

Durante más de medio siglo, el cuestionario 16PF ha demostrado ser útil para comprender y predecir una amplia gama de comportamientos importantes, proporcionando así una gran fuente de información para los que realizaron el cuestionario. Por ejemplo, el instrumento ha sido eficaz para predecir áreas tan diversas como la creatividad, habilidades sociales, empatía, compatibilidad conyugal, potencial de liderazgo, planificación del desarrollo profesional, asesoramiento y coaching, así como más de cien perfiles ocupacionales [21].

Tabla 1: Factores Primarios (FP) y su descripción en el Cuestionario 16 PF.

4. Algoritmo Paralelo Propuesto

El algoritmo propuesto es una extensión paralela del algoritmo propuesto en [10], basado en el algoritmo Apriori, con el objetivo de extraer reglas de asociación, lo que lleva a descubrir un gran número de asociaciones inesperadas.

TABLA 1

FP	Descripción Nivel Inferior	Descripción Nivel Superior
A	Reservado, separado, crítico, frío (Sizotimia)	Extrovertido, afectuoso, sin trabas, participante (Afectotimia)
B	Menos inteligente, pensamiento concreto (Inteligencia baja)	Más inteligente, pensamiento abstracto, brillante (Inteligencia alta)
C	Le afectan los sentimientos, emocionalmente menos estable, fácilmente trastornado (Poca fuerza del yo/inestabilidad emocional)	Emocionalmente estable, enfrente la realidad, calmo (Mucha fuerza del yo/estabilidad emocional)
E	Humilde, suave, obediente, conformista (Sometimiento)	Afirmativo, independiente, agresivo, empecinado (Dominancia)
F	Sobrio, prudente, serio, taciturno (Desurgencia)	Descuidado, atolondrado, alegre, entusiasta (Surgencia)
G	Expeditivo, obra a su antojo, deriva las obligaciones (Fortaleza superyoica inferior)	Escrupuloso, perseverante, sosegado, atado a las reglas (Fortaleza superyoica superior)
H	Arisco, contenido, apocado, tímido (Trectia)	Aventurado, socialmente audaz, no inhibido, espontáneo (Parmia)
I	Obstinado, confiado en sí mismo, realista, enemigo de frivolidades (Harria)	Compasivo, dependiente, sobreprotegido, sensible (Premsia)
L	Confiado, adaptable, sin celos, fácil de llevarse con él (Alaxia)	Suspicaz, porfiado, difícil de engañar (Protensión)
M	Práctico, cuidadoso, convencional, regulado por realidades externas formal (Praxemia)	Imaginativo, preocupado por urgencias internas, descuidado en asuntos prácticos, bohemio (Autia)
N	Franco, natural, sencillo, sentimental (Sencillez)	Astuto, calculador, mundano, penetrante (Astucia)
O	Flácido, seguro de sí mismo, confiado, sereno (Adecuación imperturbable)	Aprensivo, preocupado, depresivo, perturbado (Tendencia a la culpabilidad)
Q1	Conservador, respeta las ideas establecidas, tolera las dificultades tradicionales (Conservadurismo)	Experimentador, crítico, liberal, analítico, librepensador (Radicalismo)
Q2	Dependiente del grupo, un afiliado y seguidor firme (Adhesión al grupo)	Autosuficiente, prefiere tomar decisiones, ingenioso (Autosuficiencia)
Q3	Informal, no respeta el protocolo, desaliñado, obedece sus propios impulsos (Baja integración de pensamientos propios)	Controlado, socialmente, preciso, autodisciplinado, compulsivo (Alta integración de pensamientos propios)
Q4	Relajado, tranquilo, aletargado, no frustrado (Poca tensión érgica)	Tenso, impulsado, sobreexcitado, irritable (Alta tensión érgica)

Para no perder ningún valor de confianza entre los diferentes factores de personalidad y así encontrar correlaciones importantes, no se reduce el tamaño del conjunto de elementos candidatos, lo que conlleva a realizar una considerable cantidad de cálculos que demanda un mayor y notable tiempo de procesamiento, por lo cual, es necesario una paralelización del algoritmo.

Los algoritmos de reglas de asociación intentan encontrar relaciones dentro de un conjunto de transacciones, donde cada transacción puede estar formada por uno o varios ítems. Un conjunto de ítems es llamado como itemset.

Dos conceptos importantes utilizados en el algoritmo Apriori son el soporte y la confianza. El soporte está definido por el porcentaje de transacciones que contiene un itemset:

$$\text{Soporte}(A \rightarrow B) = \frac{\# \text{ de transacciones conteniendo A y B}}{\# \text{ Total de Transacciones}}$$

La medida de certeza o confiabilidad asociada con cada patrón descubierto es definida como confianza, y se calcula con la siguiente ecuación:

$$\text{Confianza}(A \rightarrow B) = \frac{\text{Soporte}(A \cup B)}{\text{Soporte}(A)}$$

La paralelización del algoritmo Apriori se da en la Generación de la Matriz de Asociación y la Matriz de Confianza definidos en los siguientes algoritmos:

Algoritmo 1. Generación de la Matriz de Asociación para dos variables

```
binary=readfile("binary.csv") //Reads the file that contains
    all the scores in binary
comb=combn(163,2) //Matrix storing all possible combinations
for (k in 0 to ncol(comb)-1)
{
    result = conjunction(binary(data.col(comb[0,k]), data.col(comb
        [1,k]))
    AsocMatrix[comb[0,k], comb[1,k]]=count_frequency(result)
}
AsocMatrix=AsocMatrix+transpose(AsocMatrix)
```

Algoritmo 2. Generación de la Matriz de Confianza para dos variables

```
for(k in 1:ncol(comb))
{
    s1=AsocMatrix[comb[0,k], comb[1,k]]/count_frequency(binary.
        col(comb[0,k]))
    ConfMatrix[comb[0,k], comb[1,k]]=s1
    s2=AsocMatrix[comb[1,k], comb[0,k]]/count_frequency(binary.
        col(comb[1,k]))
    ConfMatrix[comb[1,k], comb[0,k]]=s2
}
```

El Algoritmo 1 básicamente cuenta la frecuencia en que ambas variables obtuvieron una puntuación alta en la data binarizada, mientras que, el Algoritmo 2 genera los valores de confianza para estas variables en ambas direcciones, con ayuda de la matriz generada en el Algoritmo 1.

Ambos algoritmos fueron adaptados para usar con OPenMP, una interfaz de programación de aplicaciones utilizada para la programación paralela de memoria compartida, tanto para dos como para tres variables y siendo cuidadosos en aquellas variables que deben ser privadas, obteniendo como resultado la matriz de confianza. La implementación completa se puede encontrar en:

<https://github.com/RosePY/Parallel-Association-16FP>

5. Experimentos Y Resultados

5.1. Análisis de Datos

Los datos para este artículo se obtuvieron de https://openpsychometrics.org/_rawdata/, repositorio que almacena datos de varios cuestionarios de personalidad. El cuestionario de 16FP tiene 163 preguntas divididas en 16 categorías, para cada pregunta cada persona debe elegir qué tan precisa es en la escala de (1) en desacuerdo, (2) ligeramente en desacuerdo, (3) ni de acuerdo ni en desacuerdo, (4) ligeramente de acuerdo, y (5) de acuerdo.

Los datos están en formato.csv y en formato numérico. Los rangos numéricos se convirtieron en datos binarios, las respuestas que iban de 1 a 3 se convirtieron en 0 y las respuestas 4 y 5 se convirtieron en 1. Es importante señalar la presencia de preguntas negativas cuando la polaridad se invirtió, es decir, las respuestas en desacuerdo y ligeramente en desacuerdo se convirtieron en 1 y el resto en 0.

El tamaño total de los datos es 49159 registros por 163 preguntas (columnas). Como se explicó en la Sección 3, existe un problema en el algoritmo en serie que conduce a una gran cantidad de cálculos y por lo tanto mayor tiempo de procesamiento debido al número de asociaciones. Para ilustrar, el número total de asociaciones considerando dos variables tenemos:

$$\binom{163}{2} = 13203 * 2(\text{dado que son bidireccionales}) = 26406$$

Cuando aumentamos el número de variables a tres (por ejemplo, {A, B}→C), el número de asociaciones(subconjuntos) aumenta a:

$$\binom{163}{2} = 13203 * 161 = 2125683$$

Por lo tanto, se realiza una mejora en el rendimiento del algoritmo al paralelizarlo.

5.2. Resultados de Rendimiento

Se realizaron experimentos usando una computadora con 8 núcleos de 2.4GHz y 7.7 GB de memoria. El código fuente se implementó en C++ y OpenMP.

Para el caso de dos variables, los tiempos de ejecución del algoritmo serial y paralelo fueron de 36.72 segundos y 32.24, respectivamente. Para el caso de tres variables, los tiempos de ejecución del algoritmo serial y paralelo fueron de 4345.54 segundos y 893.75 segundos, respectivamente; donde se puede ver que el tiempo de ejecución del algoritmo paralelo se redujo en más de 4 veces en comparación al tiempo de ejecución del algoritmo serial. La Figura 1 muestra los tiempos de ejecución mencionados con más detalle.

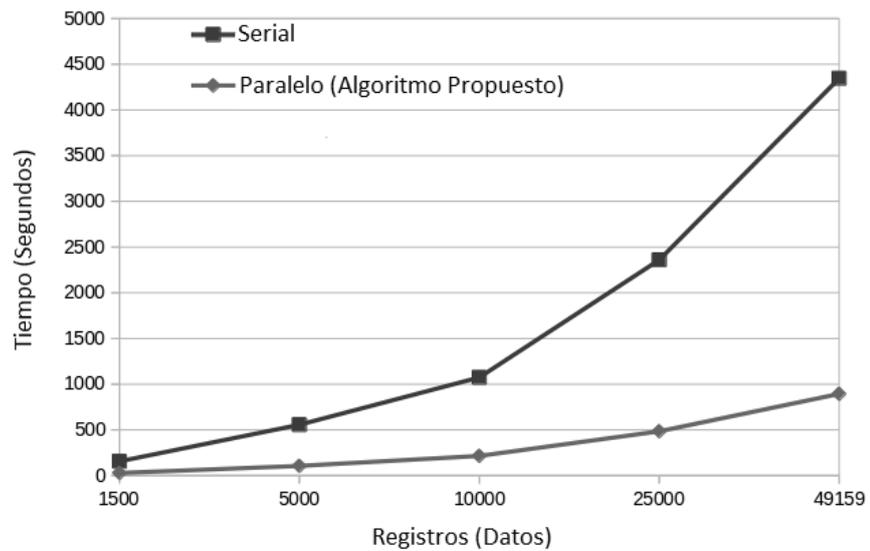


FIGURA 1: Tiempo de Ejecución: Serial y Paralelo para 3 Variables.

5.3. Resultados de Asociación (2 variables)

De las 26406 asociaciones obtenidas entre dos variables, 3618 superaron el 80% en sus valores de confianza, lo que supone el 13,7% de todas las asociaciones, de las cuales 700 (2,65% del total) superaron el 90% en sus valores de confianza. Las asociaciones con valores más altos en confianza descubiertas y obvias son:

- I like to read (H) $\xrightarrow{97.43\%}$ I don't like action movies(H)
- I love to think up new ways of doing things (M) $\xrightarrow{95.11\%}$ I enjoy hearing new ideas(M)
- I want to be left alone (N) $\xrightarrow{95.6\%}$ I enjoy my privacy (N)
- I make insightful remarks (B) $\xrightarrow{95.27\%}$ I use my brain (B)

Todas estas, obtienen el puntaje de confianza alto dentro de su misma categoría. Sin embargo, existen otras asociaciones descubiertas de diferente categoría con valores altos en confianza, por ejemplo:

I am the life of the party	$\xrightarrow{91.1\%}$	I cheer people up (A)
I readily overcome setbacks (C)	$\xrightarrow{95.26\%}$	}
I carry the conversation to a higher level (M)	$\xrightarrow{95.21\%}$	
I am exacting in my work (O)	$\xrightarrow{95.64\%}$	
		I use my brain (B)

En el segundo caso observamos como “I use my brain” de la categoría B está más relacionado con “I am exacting in my work” de la categoría O que con “I make insightful remarks” de su misma categoría, siguiendo “I readily overcome setbacks” y “I carry the conversation to a higher level” de otras categorías.

5.4. Resultados de Asociación (3 variables)

De las 2125683 asociaciones obtenidas entre tres variables, 326825 obtuvieron más del 80% en sus valores de confianza, lo que representa el 15,37% de todas las asociaciones.

Un 3.19% del total, es decir, 67902 asociaciones, superaron el 90% en sus valores de confianza, así como, un 0.1%, es decir, 2673 asociaciones, superaron el 95% en sus valores de confianza.

El valor de confianza más alto obtenido es 98.18%, el cual corresponde a la pregunta “I like to read”. Los siguientes 9 valores de confianza más altos también implican un rasgo similar a “I like to read”, “I read a lot”, los cuales son mostrados en la Figura 2.

Algunas de las asociaciones más interesantes y de diferentes categorías son:

I take an interest in other people’s lives (A)	}	$\xrightarrow{90.64\%}$	I like works of fiction (H)
I seek quiet (N)			
I want to be in charge(D)	}	$\xrightarrow{92.64\%}$	I enjoy hearing new ideas (M)
I take deviant positions (J)			
I know how to comfort others(A)	}	$\xrightarrow{94.19\%}$	I use my brain (B)
I like myself(C)			
I take an interest in other people’s lives(A)	}	$\xrightarrow{94.19\%}$	I challenge things (D)
I don’t put off unpleasant task(O)			
I talk to a lot (G)	}	$\xrightarrow{91.59\%}$	I joke around a lot (E)
I take time out for others(A)			

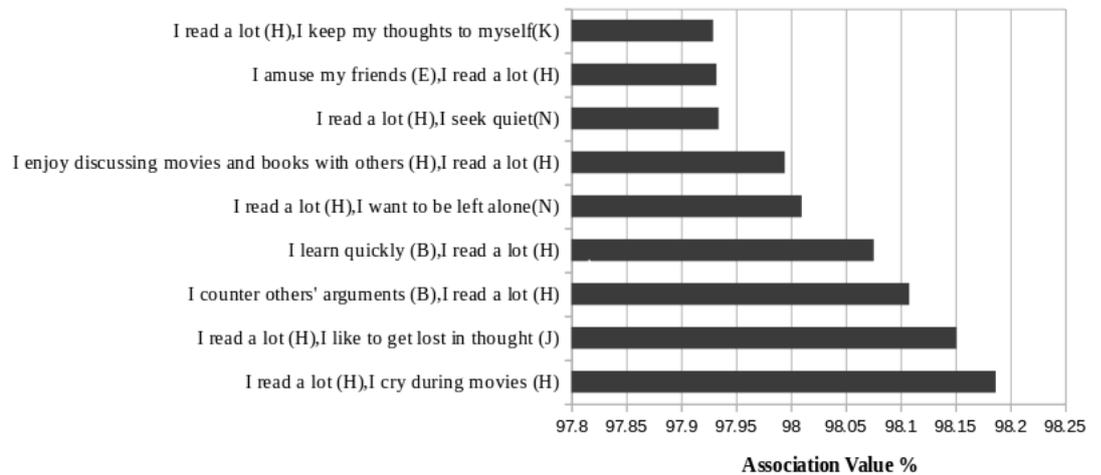


FIGURA 2: Valores de confianza entre la pregunta "I like to read" y otras preguntas.

6. Conclusiones

Este trabajo extiende un trabajo previo de minería de reglas de asociación para un cuestionario de personalidad psicométrica que lleva a descubrir un gran número de relaciones inesperadas entre las preguntas del cuestionario.

La principal diferencia con el trabajo anterior es la paralelización del algoritmo que genera la matriz de valores de confianza de manera que se pueda hallar asociaciones entre dos o más variables con menor tiempo de ejecución.

Los resultados obtenidos sobre el algoritmo propuesto muestran una mejora en el tiempo de ejecución considerable, 4 veces menor en el caso de 3 variables, con respecto al algoritmo en serie; además, los resultados muestran altos valores de confianza entre las preguntas, siendo posible identificar rasgos de personalidad más allá de una categoría, debido a las correlaciones encontradas.

Como trabajos futuros se considera usar el algoritmo paralelo propuesto en otros conjuntos de datos de pruebas de personalidad y considerar condiciones de borde, así como, compararlo con otros algoritmos paralelos propuestos en la literatura para la misma tarea.

Agradecimientos

Este artículo cuenta con el apoyo financiero de CIENCIACTIVA, Becas y Financiamiento de CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica - Perú) y la Universidad Nacional de San Agustín, Arequipa, Perú.

Referencias

- [1] Albert, D. William, K. Fayaz, and D. Veerabhadra Babu. "HSApriori: High Speed Association Rule Mining using Apriori Based Algorithm for GPU." *Int. J. of Multidisciplinary and Current research*. 2014.
- [2] Han, Jiawei, Jian Pei, and Yiwen Yin. "Mining frequent patterns without candidate generation." *ACM sigmod record*. Vol. 29. No. 2. ACM. 2000.
- [3] Zaki, Mohammed J., and Karam Gouda. "Fast vertical mining using diffsets." *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2003.
- [4] Rakesh Agrawal, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen, and A. Inkeri Verkamo. "Fast discovery of association rules." *Advances in knowledge discovery and data mining*. pp. 307-328. 1996.
- [5] Karthikeyan, T., and N. Ravikumar. "A survey on association rule mining." *International Journal of Advanced Research in Computer and Communication Engineering*. 2014.
- [6] Zhang, Fan, Yan Zhang, and Jason Bakos. "Gpapriori: Gpu-accelerated frequent itemset mining." *Cluster Computing (CLUSTER)*, 2011 IEEE International Conference on. IEEE. 2011.
- [7] Albert, D. William, K. Fayaz, and D. Veerabhadra Babu. "HSApriori: High Speed Association Rule Mining using Apriori Based Algorithm for GPU." *Int. J. of Multidisciplinary and Current research*. 2014.
- [8] Spandana, K., D. Sirisha, and S. Shahida. "Parallelizing Apriori Algorithm on GPU." *International Journal of Computer Applications*. 2016.
- [9] Govindaraju, N., and M. Zaki. "Advances in Frequent Itemset Mining Implementations." *FIMI*. 2003.
- [10] Syed Khalid Perwez, Hamza Mohd Zubair, Muhammad Rukunuddin Ghalib, Kauser Ahmed and Mohammed Iftekhar, "Association Rule Mining Technique for Psychometric Personality Testing and Behaviour Prediction." *International Journal of Engineering & Technology*". 2013.
- [11] Cattell, Raymond B., Cattell, A.K., and Cattell, H.E.P. *16PF Fifth Edition Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing. 1993.
- [12] Cattell, Raymond B. "Personality and motivation structure and measurement". 1957.
- [13] Conn, Steven R., and Mark L. Rieke, eds. *16PF fifth edition technical manual*. Institute for Personality & Ability Testing, Incorporated. 1994.

- [14] Phares, E.J.; Chaplin, W.F.. Introduction to personality (Fourth ed.). New York: Longman. pp. 8–9. ISBN 0-673-99456-2. 1997.
- [15] Feist, Jess Feist, Gregory J. Theories of personality (7th ed. ed.). Boston: McGraw Hill Higher Education. ISBN 978- 0073382708. 2009.
- [16] Kahn, Michael. Basic Freud: psychoanalytic thought for the twenty first century (1. paperback ed. ed.). New York: Basic Books. ISBN 9780465037162.. 2002.
- [17] Cheney, W. David Pierce, Carl D. Behavior analysis and learning (4th ed. ed.). New York, NY: Psychology Press. ISBN 9780805862607. 2008.
- [18] Allport, G. W. Personality: A psychological interpretation. New York, NY: Henry Holt & Company. 1937.
- [19] Vernon, P. E., and Vernon, P. E. Personality assessment: A critical survey. London: Methuen. 1964.
- [20] Baumeister, Roy F., Kathleen D. Vohs, and David C. Funder. "Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?." Perspectives on Psychological Science. 2007.
- [21] Cattell, Heather EP, and Alan D. Mead. "The sixteen personality factor questionnaire (16PF)." The SAGE handbook of personality theory and assessment. 2008.
- [22] Butcher, James N., Jeanette Taylor, and G. Cynthia Fekken. "Objective Personality Assessment with Adults" Comprehensive Clinical Psychology: Assessment. 1998.
- [23] Fehriinger, Heather M. "Contributions and limitations of Cattell's sixteen personality factor model." 2004.