

Análisis del comportamiento contextual del usuario y su relación con el consumo de aplicaciones móviles

Luis Antonio Rojas-Potosi*

Leandro Krug Wives**

Alejandro Fernández***

Juan Carlos Corrales****

Recibido: 18/09/2013 • Aceptado: 17/01/2014

Resumen

El mercado de aplicaciones móviles crece aceleradamente, captando la atención sobre el análisis del consumo de las mismas para mejorar su recomendación y la satisfacción del usuario, soportado en la búsqueda de información que permita explicar la ocurrencia de eventos del ciclo de vida de las aplicaciones. En este sentido algunos estudios muestran una fuerte relación entre dichos eventos e información sobre la situación contextual actual del usuario y su registro de consumo de aplicaciones. Sin embargo, dichos trabajos no consideran la relación existente entre el registro de estados de las dimensiones contextuales (denominado como CUBe) y la ocurrencia de eventos del ciclo de vida de las aplicaciones, tal como lo hace este estudio. En este sentido, se presenta un análisis estadístico sobre un dataset real, evidenciando una fuerte relación entre el CUBe y el lanzamiento de un conjunto de aplicaciones. Finalmente, se describe un escenario donde estas relaciones pueden ser utilizadas para enriquecer sistemas de recomendación para aplicaciones móviles.

Palabras clave: análisis contextual de comportamiento, usuario, aplicaciones móviles.

<? * Magister(C) en Ingeniería Telemática, investigador del Grupo de Ingeniería Telemática de la Universidad del Cauca, Colombia, Popayán – Cauca, Carrera 17 # 19N-166B, Tel: (+57-2) 8230443, email: jarojas@unicauca.edu.co

<? ** PhD en Ciencias de la Computación, profesor asociado en el Departamento de Informática Aplicada del Instituto de informática de la Universidad Federal de Rio Grande del Sur (UFRGS), Brasil, Porto Alegre – Rio Grande do Sul, UFRGS Sala: 233 - Predio 43424 (67) email: wives@inf.ufrgs.br

<? *** , PhD en Ciencias de la Computación, Profesor Titular de la Facultad de Informática en la Universidad Nacional de la Plata (UNLP), Argentina, Buenos Aires, Lfia, calle 50 y 115 La Plata, Prov. Buenos Aires, email: alejandro.fernandez@lfia.info.unlp.edu.ar

<? **** PhD en Ciencias de la Computación, Profesor Titular del Departamento de Telemática de la Universidad del Cauca, Colombia, Popayán – Cauca, Edificio de Ingenierías sector Tulcán, Oficina 403, Tel: (+57-2) 8209800, ext. 2129, email: jcorral@unicauca.edu.co

Analysis of the contextual behavior of the user and its relation with the consumption of mobile applications

Abstract

The mobile applications market is growing rapidly, capturing the attention on the analysis of the consumption of applications to improve their recommendation and the satisfaction of the user. The analysis is based on the search for information that allow explaining the occurrence of events of the life cycle of the applications. In this sense, several studies show a strong relation between such events and information about the current contextual situation of the users and their application consumption record. Nevertheless, the mentioned studies do not take into account the relation existing between the contextual dimensions status record (known as CUBe), and the occurrence of application life cycle events, but it was taken into account for this research work. Therefore, a statistical analysis about an actual dataset is presented, evidencing a strong relation between the CUBe and the launch of a set of applications. Finally, a scenario where these relations can be used for enriching recommendation systems for mobile applications is described.

Key words: contextual behavior analysis, user, mobile applications.

INTRODUCCIÓN

Los dispositivos móviles han evolucionado desde aquellos con funcionalidades básicas de comunicación, hasta los dispositivos multitarea actuales, capaces de soportar la mayoría de actividades diarias del usuario con aplicaciones software. El mercado de estas aplicaciones crece constantemente, captando la atención en el análisis de su consumo para mejorar procesos de recomendación y experiencia de uso. En este sentido, uno de los principales temas de investigación es la búsqueda de información que permita explicar la ocurrencia de eventos del ciclo de vida de las aplicaciones [1]; siendo el teléfono móvil la principal interfaz para la captura de diferentes tipos de datos que podrían explicar dichos eventos [2].

En este sentido, algunos estudios han mostrado una fuerte relación entre los eventos del ciclo de vida de las aplicaciones y la situación contextual actual del usuario [1, 3, 4], la cual es el conjunto de los estados actuales en las diferentes dimensiones contextuales del usuario. Por otro lado, otros trabajos han utilizado las secuencias de consumo al margen de la información contextual, obteniendo una mejora al explicar el registro de consumo de aplicaciones [5]. Sin embargo, los trabajos mencionados no han explorado la relación entre las secuencias de estados en las dimensiones contextuales y la ocurrencia de eventos del ciclo de vida de las aplicaciones, tal como lo hace la presente propuesta. En adelante, la secuencia de estados en dimensiones contextuales del usuario será referida como comportamiento contextual del usuario (CUBE) y su análisis será guiado por las siguientes preguntas:

- ¿Existe alguna relación entre el CUBE y el consumo de aplicaciones?
- ¿Cómo el resultado del análisis entre el CUBE y el consumo de aplicaciones puede ser aplicado a dominios de investigación relacionados?

Para responder estas preguntas, se presenta un análisis estadístico orientado a describir el CUBE y su relación con el consumo de aplicaciones, utilizando los datos recolectados por Nokia para el “MDC Challenge” [6, 7]. Los resultados muestran que dependiendo de la aplicación móvil se presentan diferentes niveles de relación entre su consumo y el CUBE, destacando características comunes entre las aplicaciones con los mayores niveles de relación. En este sentido, el análisis también presenta que dichos niveles son influenciados por la ventana de observación de la muestra, siendo incrementada al tomar grupos de días agregados de acuerdo con características del CUBE. Como complemento, este artículo describe un potencial escenario donde pueden ser aplicados los resultados del análisis.

Así, antes de presentar el análisis estadístico en la sección 3, y la discusión de los resultados en la sección 4, primero se expondrá el planteamiento del problema en la sección 2. La sección 5 es dedicada a describir la diferencia entre este estudio y otros

trabajos enfocados al análisis del consumo de aplicaciones. En la última sección el lector encontrará las conclusiones y los trabajos futuros.

1. PLANTEAMIENTO DEL PROBLEMA

En trabajos orientados al análisis contextual del consumo, contexto es generalmente definido como toda información que puede ser utilizada para caracterizar la situación de una entidad, donde una entidad es una persona, lugar u objeto que es considerado relevante para la interacción entre el usuario y una aplicación [8, 9]. Los tipos de información contenidos en la descripción del contexto son conocidos como dimensiones contextuales $D = \{d_i\}_{1 \leq i \leq n}$, por ejemplo $D = \{ubicación, intensidad\ de\ luz, temperatura, etc.\}$. Así, el comportamiento contextual del usuario (CUBE) es definido como el conjunto $B = \{R_i\}_{1 \leq i \leq n}$, de registros históricos $R_i = \{t_h, s_{i,h}\}_{1 \leq h \leq m}$ para cada una de las n dimensiones contextuales, donde s_h es el estado de la dimensión i en un tiempo t_h . De acuerdo con la anterior definición, cada dimensión contextual considerada en el CUBE debe ser descrita por estados discretos. Por otro lado, la información sobre consumo de aplicaciones está representada por un registro $AR_i = \{t_j, a_j, \Delta t_j\}_{1 \leq j \leq p}$ que contiene tuplas de la forma {tiempo, aplicación consumida, duración de consumo}. Una interpretación gráfica de los datos es presentada en la figura 1.

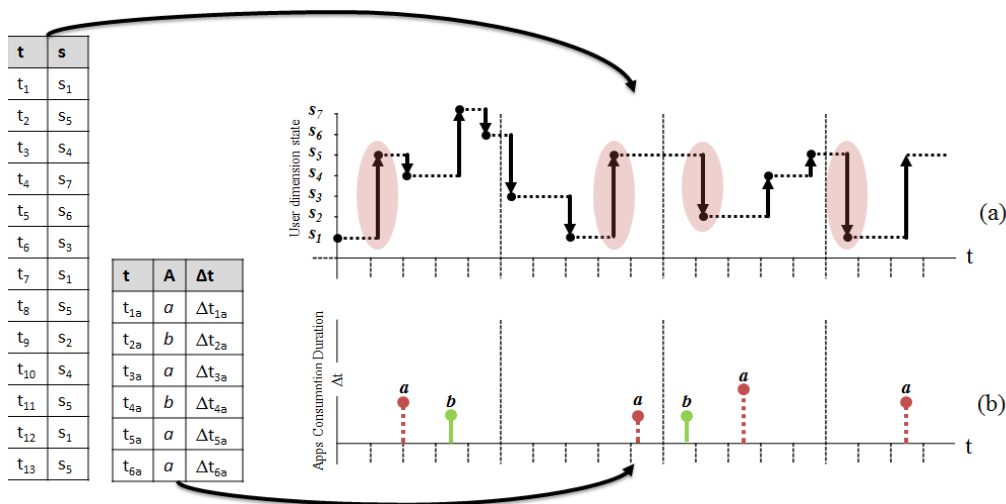


Figura 1: representación gráfica de (a) el registro histórico de una dimensión contextual (b) un registro de consumo de aplicaciones.

Fuente: elaboración propia

La figura 1a representa de manera gráfica el registro de cambios de estado de una variable contextual, que hace parte del CUBE. En dicha figura las flechas indican el

cambio entre dos estados consecutivos, en tanto que las líneas punteadas representan la permanencia en un estado. El problema del presente artículo es identificar si existe una relación entre los cambios de estado (flechas) y el consumo de aplicaciones. Por ejemplo: el problema para la aplicación a en el caso de la figura 1 consiste en identificar una eventual relación entre el consumo de dicha aplicación y el cambio de estado inmediatamente anteriores a su consumo (sombreados con rojo).

1.1. Colección de datos

Para poder explorar el anterior problema, se utilizó el dataset del MDC challenge [6, 7]. La versión disponible del dataset analizado contiene 184 participantes los cuales registran el uso 646 aplicaciones. Todos los datos fueron recolectados en smart-phones durante más de un año. Para la exploración de la relación entre el CUBE y el consumo de aplicaciones, se redujo el espacio de observación del CUBE a la dimensión contextual de “afiliación a celdas GSM”, la cual es comparada con el registro del consumo de aplicaciones.

2. ANÁLISIS DE LA RELACIÓN ENTRE CONSUMO Y CUBE

Este análisis contempla dos aproximaciones: una basada en el análisis estadístico básico de las secuencias de celdas y la ocurrencia del consumo de aplicaciones para todo el período observado, y otra enfocada en el análisis estadístico específico de grupos de días agregados de acuerdo con características del CUBE.

2.1. Análisis estadístico general

Este análisis busca determinar dos tipos de relaciones: *i)* la probabilidad de que una secuencia $ss = \{s_{i-2}, s_{i-1}\}$ esté presente como antecedente al consumo de una aplicación a , con respecto al número de apariciones de a ($p(ss|a)$), y *ii)* la probabilidad que la aplicación a sea el consecuente de la aparición de la secuencia ss , con respecto al número de apariciones de ss ($p(a|ss)$). Una representación gráfica de la aproximación propuesta se presenta en la figura 2.

Estos dos tipos de probabilidades ($p(App|ss)$ y $p(ss|App)$) pueden ser deducidos del análisis de la secuencia de estados contextuales y consumo de aplicaciones, producida al mezclar los registros de consumo y cambio del estado de variables contextuales. En este sentido, para cada aplicación a_i consumida por el usuario u , se tendrá un conjunto de secuencias de la forma $\mathbf{A} = \{s_{i-2}, s_{i-1}, a_i\}_{1 \leq i \leq n}$, las cuales contienen como tercer elemento a la aplicación a_i a analizar. A su vez, para cada ss única del conjunto \mathbf{A} , se analiza el conjunto de secuencias que en el registro del usuario contienen el fragmento ss como precedente $\mathbf{B} = \{s_{i-2}, s_{i-1}, z_*\}_{1 \leq i \leq n}$, siendo z_i un registro correspondiente a una aplicación o a un estado. Con base en los anteriores conjuntos se define el conjunto

C_j como la intersección que contiene todas las secuencias ssa que tienen a ss_j como antecedente y a a_i como precedente. A partir de A, B y C se obtiene las probabilidades $p(ss|a)$ y $p(a|ss)$ tal como lo presenta la figura 2.

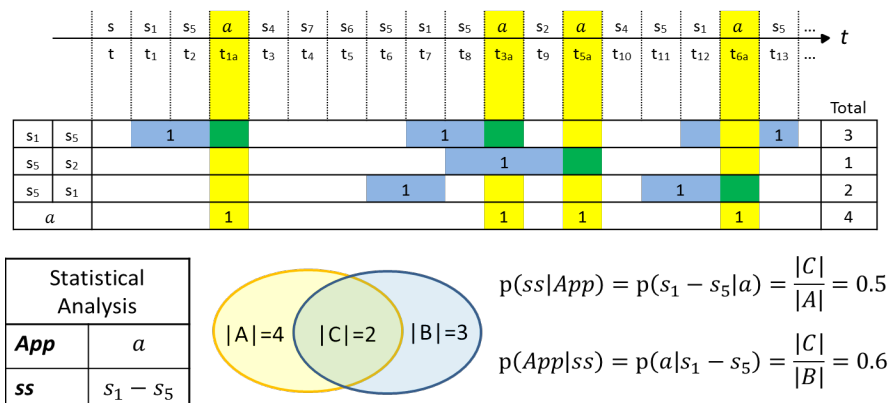


Figura 2. Análisis estadístico de secuencias.

Fuente: elaboración propia

Como resultado, para cada aplicación a_i y cada secuencia ss_j se tiene los valores de $|A|$, $|B|$, $|C|$, $p(ss_j|a_i)$ y $p(a_i|ss_j)$; de gran importancia es el valor de $p(a_i|ss_j)$, ya que indica el grado de relación entre la aparición de la secuencia ss_j y el consumo de la aplicación a_i . Sin embargo, $p(a_i|ss_j)$ es sumamente elevado en secuencias ss_j que solamente suceden una vez, generando valores atípicos en dichas situaciones, los cuales, al depender solamente de un valor fortuito, no representan una relación clara entre la aparición de la secuencia ss y el consumo de la aplicación a_i . Por otro lado, cada uno de los anteriores valores se encuentra asociado a una secuencia ssa , y no permite identificar el grado de relación general entre el consumo de la aplicación a_i y el cambio de estados de la dimensión contextual. Por tal razón, se propone *car* (CUBe-App-relationship), como una medida que agrega los anteriores valores describiendo un nivel de relación entre el cambio de estados (en general) y el consumo de una aplicación específica, tal como lo presenta la siguiente ecuación:

$$car_i(th_1, th_2) = \frac{\sum_{ssa \in A_i} |C_{ssa}|}{|A_i|} \tag{1}$$

Donde, A_i corresponde al conjunto de todas las secuencias únicas que cumplen con las siguientes condiciones:

- $|A| > 1$, descarta aquellas aplicaciones que solamente se consumen una vez en el periodo observado. El valor de *car* para dichas aplicaciones es cero.

- $|B| > th_1$, esta condición excluye a aquellas secuencias ss que se presentan de manera única y no explican una relación entre el CUBE y el consumo de aplicaciones. Para el caso del presente artículo el valor de th_1 es 1.
- $p(a|ss) > th_2$, esta condición es utilizada para descartar a aquellas secuencias ssa que presentan una baja relación entre su ocurrencia y el lanzamiento de la aplicación a_i

Los límites th_1 y th_2 son utilizados para regular las condiciones anteriores. El límite th_2 se encuentra en un rango de 0 (cero) a 1(unos) y al igual que th_1 guarda una relación inversa con el valor del car , tal como lo presentada la figura 3.

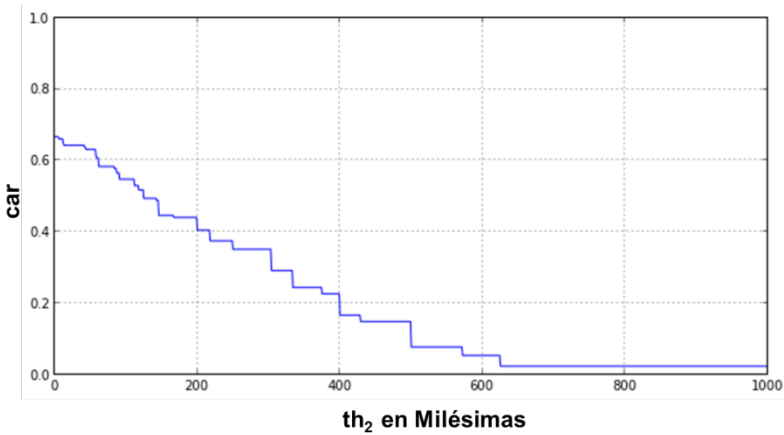


Figura 3. Relación entre el valor del car (eje y) y el valor de th_2 (eje x dado en milésimas).

Fuente: elaboración propia

Sin embargo, a pesar del grado de generalidad alcanzada por el valor de car , aún no es posible deducir un nivel de relación entre el CUBE y el consumo de la aplicación analizada. Por tal motivo, y a partir del análisis de la gráfica car versus th_2 , se propone una aproximación a dicho nivel de relación, mediante el cálculo del área bajo la curva del car , denominada $carl$ (CUBE-App-Relationship level), dependiendo de th_2 , como lo presenta la siguiente ecuación:

$$carl_i(th_1) = \int_0^1 car_i(th_1, th_2) dth_2 \tag{2}$$

De esta manera, se genera una matriz de la forma $M_{m \times n}$ donde cada celda almacena los valores de $carl_{j,i}$ para cada usuario u_j y aplicación a_i . El valor del $carl$ representa el valor de la relación entre el cambio de estados contextuales de la dimensión analizada y el lanzamiento de una aplicación móvil. Las columnas de la matriz corresponden a vectores y_i (valores del $carl$ para una misma aplicación a_i dada por distintos usuarios)

de longitud m , y las filas a vectores \mathbf{x}_j (valores del *carl* asociado a un mismo usuario u_j y distintas aplicaciones) de longitud n . Al calcular la media de los vectores y_i y organizarla de mayor a menor se obtienen las figuras 4a y 4b:

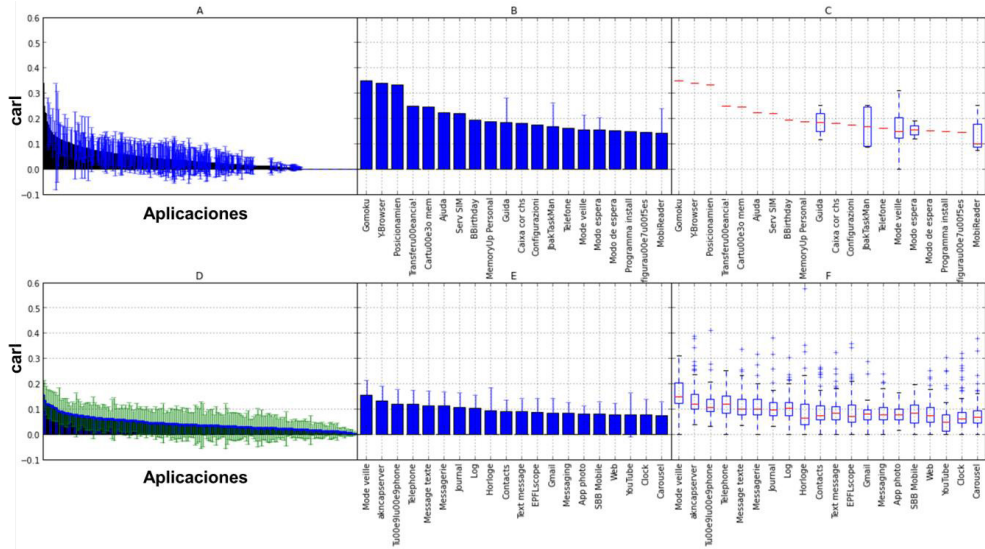


Figura 4. Carl por Aplicaciones organizadas de acuerdo al valor medio del *carl*: media y desviación estándar (a,b,d,e), y mediana y cuartiles (c,f).

Fuente: elaboración propia

Como se puede apreciar en las figuras 4a y 4b, la máxima media alcanzada por una aplicación corresponde a 0.35. Sin embargo, dicha media se presenta en una aplicación que ha sido consumida solamente por un usuario durante todo el período de observación del dataset (figura 4c). Así, debido a su bajo aporte en la descripción estadística, se descartaron aquellas aplicaciones que han sido consumidas por menos de 10 usuarios de los 184, las figuras 4d, 4e y 4f. En la figura 4f se puede observar que las aplicaciones que presentan mayor media en su valor de *carl* tienen varios valores extremos altos, muy superiores a la máxima mediana, lo cual no explica el bajo valor de la media en la figura 4e. Sin embargo, la gráfica de los cuartiles en la figura 4f expone una gran concentración de valores levemente inferiores a la media, repercutiendo en un bajo nivel de desviación estándar para aplicaciones con altos valores de media, figura 4e. Por otro lado, para aquellas aplicaciones con bajo valor medio del *carl* la desviación estándar tiene valores elevados con relación a su media, figura 4d. En este sentido, la desviación estándar no presenta una relación significativa directa con el valor de la media para el caso de las aplicaciones con alto valor medio de *carl*. En dicho caso, la correlación entre el valor de desviación y media de las aplicaciones con mayor media, presenta incluso valores negativos (-0.27), como lo sugiere la figura 5, en la cual se

presenta el coeficiente de correlación de Pearson al comparar los vectores de media y desviación estándar de las top- k aplicaciones según la media, variando el k .

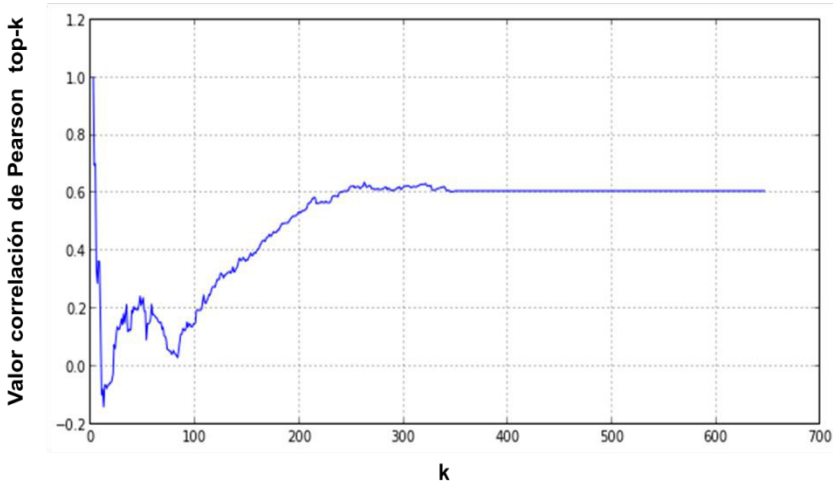


Figura 5. Correlación de Pearson entre la desviación estándar y el valor medio del *carl* para las top- k aplicaciones por valor medio de *carl* ordenadas de acuerdo al valor medio de *carl*.

Fuente: elaboración propia

Lo anterior indica que estas top-20 aplicaciones tienden a disminuir su valor de desviación estándar a medida que el valor medio del *carl* aumenta y sugieren que en múltiples usuarios existe una relación *estable* entre el cambio de estados de la variable contextual y el consumo de dicha aplicación.

A su vez, con el fin de encontrar la relación existente entre las *carl-top-k* aplicaciones, se analizó su nivel de correlación (entre aplicaciones) utilizando los valores del *carl*, a partir de la figura 6.

La figura 6 presenta una leve tendencia a tener altos valores de correlación entre las aplicaciones que presentan mayor valor medio de *carl* (esquina superior izquierda figura 6). Lo anterior sugiere que este tipo de aplicaciones son influenciadas de manera similar dependiendo del comportamiento contextual individual del usuario, en este caso enfocado en la dimensión de celdas GSM. En otras palabras, si un usuario presenta un alto nivel de *carl* con alguna aplicación que tiene un alto valor medio de *carl*, es muy probable que las demás aplicaciones con alto nivel medio de *carl* tengan un valor de *carl* alto para ese usuario. Lo anterior sugiere la existencia de un conjunto de aplicaciones cuyo nivel de *carl* se encuentra correlacionado, presentando la posibilidad de utilizar este hecho en sistemas de filtrado colaborativo, en el sentido de recomendar aplicaciones con un alto valor medio de *carl*, si el usuario presenta un alto nivel de *carl* para una de ellas.

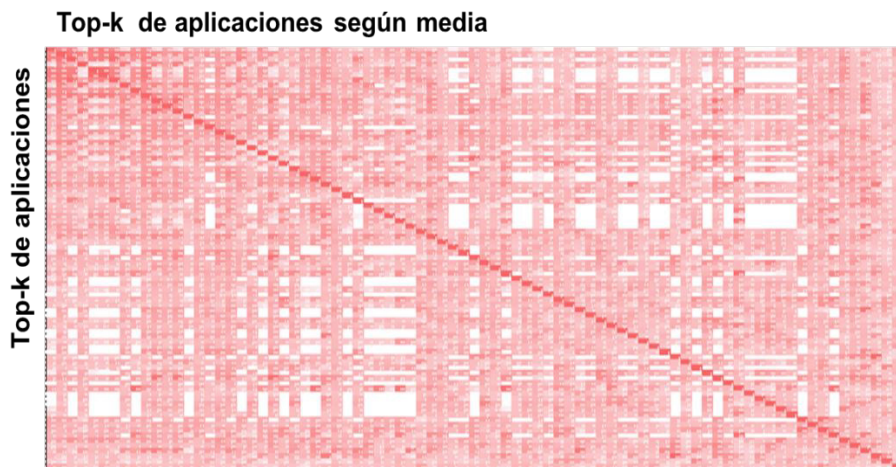


Figura 6. Correlación de Pearson entre aplicaciones, a partir del análisis de su vector de *carl* por usuario, organizadas de acuerdo a su valor medio de *carl* de manera descendente. Entre mayor correlación mayor concentración del color rojo.

Fuente: elaboración propia

2.2. Análisis sobre agrupaciones de días de acuerdo con la dimensión contextual

En el análisis estadístico se pueden observar valores de *carl* relativamente bajos. Sin embargo, ese hecho no indica de manera absoluta que la relación entre el CUBE y el consumo de aplicaciones sea baja. En este sentido, buscando determinar factores de observación, que influyan sobre el *carl*, se analizó el impacto del período de tiempo de la muestra observada, con base en diversos estudios que han demostrado que las características contextuales del usuario presentan diferencias dependiendo del periodo de tiempo observado; la agrupación por días fue el factor más común en el análisis [1, 10, 11, 3]. A partir de esa premisa, se decidió analizar el CUBE por día, utilizando la aproximación presentada por Huang en [11], de la cual se adaptó su algoritmo para la detección de lugares significativos (algoritmo 1 [11]) en la detección de grupos de días destacados según características del CUBE. Dicho algoritmo genera un conjunto de grupos destacados de días, para los cuales se analiza el *carl* por aplicación, dependiendo del usuario. Un ejemplo de los resultados obtenidos es presentado en la figura 7.

En la figura 7 los grupos destacados tienen identificadores numéricos, obtenidos por el algoritmo. Los identificadores “0” y “-1” corresponden al grupo que contiene a todos los días y al grupo que contiene a los días que no pertenecen a un grupo destacado, respectivamente. En dicha figura se puede observar que el valor del *carl* es superior para el grupo destacado “11”; el caso contrario es presentado por el grupo “39”. De este ejemplo se resalta que el valor del *carl* puede ser afectado al especificar los días

de observación. Sin embargo, este resultado específico no nos permite generalizar el hecho de que el valor del *carl* puede incrementarse al ser observado en un conjunto específico de días.

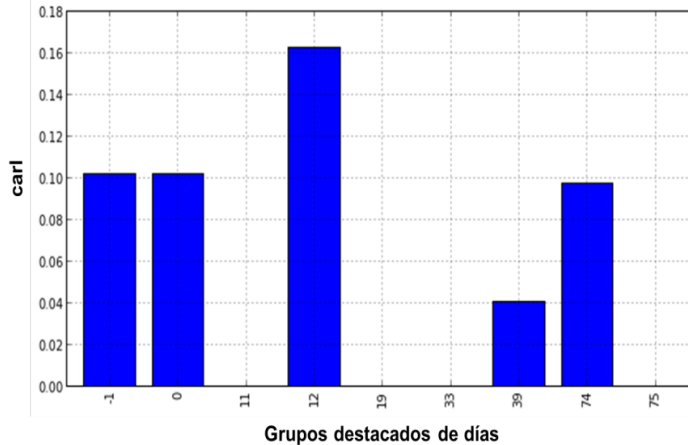


Figura 7. Ejemplo de valores del *carl* para distintos grupos de días destacados.

Fuente: elaboración propia

Con el fin de generalizar dichos resultados se realizó una comparación entre el máximo valor de *carl* obtenido en los grupos destacados y el valor observado en el grupo de “todos los días” (“0”). Dicha comparación fue realizada para cada una de las 7255 tuplas de la forma $\{usuario, aplicación\}$. De dicha comparación se observó que para el 80 % de las tuplas (5.825), el valor del *carl* de alguno de los grupos destacados supera al valor del *carl* calculado para el grupo de todos los días. A su vez, para cada dupla se analizó la relación entre la diferencia del valor máximo observado en uno de los grupos destacados y el observado en grupo de todos los días, obteniendo los valores de la tabla 1:

Tabla 1. Descripción estadística de la relación entre la diferencia entre el valor de *carl* máximo en los grupos de días destacados y el presentado en la observación general.

| | |
|---------------------|---------------------|
| Media | 0.6740482754027893 |
| Desviación estándar | 1.5062226545287989 |
| Valor máximo | 45.285714285714285 |
| Mediana | 0.34498177053175 |
| Valor mínimo | -0.9761904761904763 |

Fuente: elaboración propia

De los anteriores valores estadísticos se puede apreciar que el *carl* máximo de los grupos de días destacados es superior al *carl* de todos los días, reforzando la premisa

de que al enfocar el análisis en un grupo de días similares (de acuerdo a sus estados contextuales) se puede incidir en el aumento del nivel del *carl*. Lo anterior sugiere que el nivel de relación entre el consumo de aplicaciones y el CUBE depende del tiempo de observación; a su vez, si el tiempo de observación es seleccionado de acuerdo con características del CUBE (como lo realiza la adaptación del algoritmo de Huang) el valor del *carl* tiende a aumentar.

3. DISCUSIÓN

Después del análisis del dataset MDC, utilizando una medida que permite cuantificar el nivel de relación entre el CUBE del usuario y su registro de consumo de aplicaciones (*carl*) se observaron los siguientes hechos: *i)* la relación entre el CUBE y el consumo de aplicaciones no se presenta con el mismo nivel para todas las aplicaciones móviles. *ii)* Aquellas aplicaciones con un elevado valor medio de *carl*, comparten algunas características en común, como la alta correlación de su valor de *carl* dependiendo de sus usuarios (figura 6). *iii)* Existe una fuerte relación entre el CUBE y el consumo de las top-k aplicaciones de acuerdo con su valor medio de *Carl*, lo cual es soportado por el hecho de que los valores del *carl* para las aplicaciones que presentan altos valores medios son estables respecto a su desviación estándar, presentan en algunos casos una relación inversa entre la desviación estándar y la media, e indican que un alto valor medio del *carl* no se explica por una distribución aleatoria de sus componentes. *iv)* El *carl* es dependiente del periodo de observación y su valor puede incrementarse si se focaliza en días similares de acuerdo con el CUBE del usuario. Este último hecho sugiere que existen días en los cuales el comportamiento del usuario puede explicar en mayor medida el consumo de ciertas aplicaciones, hecho que puede ser utilizado para mejorar sistemas de recomendación y la usabilidad de las aplicaciones.

3.1 Descripción de un escenario de aplicación

Un ejemplo de la utilización de los anteriores hallazgos es la construcción de nuevos sistemas de recomendación de aplicaciones móviles. Un sistema de recomendación de aplicaciones móviles busca generar sugerencias de aplicaciones útiles al usuario, evitándole la complejidad relacionada con el análisis de la información descriptiva de todas las miles de aplicaciones existentes en las tiendas [12, 13]. Así, en este dominio, se podrían generar nuevos sistemas de recomendación que podrían tener las siguientes capacidades, entre otras:

- A partir de la detección del tipo de día con respecto a una dimensión del CUBE del usuario podrían recomendar nuevas aplicaciones, altamente utilizadas en días similares, y relacionar el cambio de estado que presente mayor probabilidad para la aceptación de la recomendación.

- Recomendar aplicaciones con niveles de *carl* elevados de acuerdo con el *carl* presentado en usuarios con comportamiento contextual similar. Este tipo de recomendaciones podría soportarse en: *i)* el análisis de matrices “*usuario frente a aplicaciones*”, donde las celdas pueden ser valores del *carl*, o *ii)* el análisis de la similitud del usuario mediante la representación del CUBe del usuario con otros modelos probabilísticos que permitan su comparación.

Las anteriores aproximaciones podrían entregar recomendaciones de aplicaciones frecuentemente utilizadas y relevantes para momentos específicos del día, caracterizados por el cambio de estados en las distintas dimensiones contextuales. El éxito de este tipo de sistemas podría ser medido a varios niveles relacionados con el estado de consumo conseguido por las aplicaciones recomendadas (descartada, vista, utilizada solamente una vez, utilizada frecuentemente), tal como lo describe el trabajo de Böhmer [5].

4. TRABAJOS RELACIONADOS

Este trabajo se enmarca dentro del área de análisis de consumo de aplicaciones móviles. Dicha área aborda la explicación de los eventos del ciclo de vida de las aplicaciones, a partir de la observación de diferentes tipos de información. Entre los primeros trabajos se destacan los de Bohmer [12, 4, 14], los cuales exploraron la relación existente entre el consumo de aplicaciones y distintas dimensiones contextuales y demográficas, encontrando varios indicios de la dependencia del consumo de ciertos grupos de aplicaciones de acuerdo con las características del contexto actual del usuario. En esta línea, trabajos posteriores lograron resultados relevantes al analizar la probabilidad de la co-ocurrencia entre estados actuales de dimensiones contextuales y su influencia en el lanzamiento de aplicaciones, resaltando la utilización de información sobre el tiempo, la localización, el perfil demográfico del usuario y la última aplicación utilizada [11, 5, 15]. Sin embargo, estos trabajos, a diferencia del trabajo de Natarajan [16], no han considerado el concepto de CUBe, relacionado con el historial de los estados de las dimensiones contextuales. En el trabajo de Natarajan se utiliza el concepto de contexto interaccional, el cual es análogo al CUBe, relacionándolo con el registro/secuencia de aplicaciones utilizadas en la sesión actual del usuario. En dicho trabajo se destaca que la naturaleza repetitiva del consumo de aplicaciones hace posible explorar la relación entre su consumo y distintos registros históricos de información del usuario. Sin embargo, dicho trabajo no explora modelos que permitan relacionar registros de diferentes dimensiones contextuales y el consumo de aplicaciones, como es realizado en el presente trabajo, en el cual se expone un análisis estadístico sobre esta relación, modelos que la describen y factores que la influyen.

5. CONCLUSIONES

Este análisis exploratorio brindó herramientas estadísticas que permitieron determinar la existencia de una clara relación entre el CUBE del usuario y el consumo de aplicaciones. Cabe resaltar que, como era esperado, no todas las aplicaciones tienden a presentar una estrecha relación entre su consumo y el CUBE del usuario. Además, este estudio permitió determinar que el tiempo de observación del CUBE y el consumo puede incidir positivamente en el valor de dicha relación.

Los resultados de este estudio exploratorio pueden ser potencialmente utilizados en diferentes áreas de aplicación, resaltando especialmente la generación de nuevas técnicas de agregación de información para la recomendación de aplicaciones. En dicha área, se espera que la utilización de un enfoque centrado en el CUBE del usuario conduzca a recomendaciones novedosas, altamente personalizadas y con características proactivas [17].

Como trabajos futuros se destaca la implementación del escenario propuesto para sistemas de recomendación de aplicaciones móviles, la aplicación de este análisis exploratorio en distintos datasets, involucrando nuevas dimensiones contextuales, y la extensión de este análisis enfocado en la predicción del CUBE del usuario.

AGRADECIMIENTOS

Este artículo utilizó fragmentos del dataset MDC, disponible por el instituto de Investigación Idiap, Suiza y propiedad de Nokia. El trabajo del MSc(c) Luis Antonio Rojas-Potosí es financiado por COLCIENCIAS y el departamento de Telemática de la Universidad del Cauca por medio del proyecto TelComp2.0, convenio 521.

REFERENCIA

- [1] M. Böhmer, B. Hecht, J. Schöning, A. Krüger and G. Bauer, "Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage," New York, NY, USA, 2011.
- [2] N. Aharony, "Social fMRI : measuring and designing social mechanisms using mobile phones," Boston, MA, 2012.
- [3] W. Pan, N. Aharony and A. Pentland, "Composite Social Network for Predicting Mobile Apps Installation," *CoRR*, vol. abs/1106.0359, 2011.
- [4] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan and D. Estrin, "Diversity in smartphone usage," New York, NY, USA, 2010.
- [5] M. Böhmer and L. Ganev, "AppFunnel: A Framework for Usage-centric Evaluation of Recommender Systems that Suggest Mobile Applications," 2013.

- [6] J. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. Do, O. Dousse, J. Eberle and M. Miettinen, "The mobile data challenge: Big data for mobile computing research," 2012.
- [7] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," Berlin, 2010.
- [8] A. K. Dey, "Understanding and Using Context," *Personal Ubiquitous Comput.*, vol. 5, pp. 4-7, January 2001.
- [9] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira and P. B. Kantor, Eds., Springer US, 2011, pp. 217-253.
- [10] L. Baltrunas, B. Ludwig, S. Peer and F. Ricci, "Context relevance assessment and exploitation in mobile recommender systems," *Personal and Ubiquitous Computing*, vol. 16, pp. 507-526, 2012.
- [11] K. Huang, C. Zhang, X. Ma and G. Chen, "Predicting mobile application usage using contextual information," New York, NY, USA, 2012.
- [12] M. Böhmer, G. Bauer and A. Krüger, "Exploring the Design Space of Context-aware Recommender Systems that Suggest Mobile Applications," Barcelona, Spain, 2010.
- [13] F. Ricci, L. Rokach and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira and P. B. Kantor, Eds., Springer US, 2011, pp. 1-35.
- [14] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang and S. Venkataraman, "Identifying diverse usage behaviors of smartphone apps," New York, NY, USA, 2011.
- [15] T. Yan, D. Chu, D. Ganesan, A. Kansal and J. Liu, "Fast app launching for mobile devices using predictive user context," New York, NY, USA, 2012.
- [16] N. Natarajan, D. Shin and I. S. Dhillon, "Which App Will You Use Next?: Collaborative Filtering with Interactional Context," New York, NY, USA, 2013.
- [17] L. Rojas-Potosi, L. Suarez-Meza, L. Ordoñez-Ante and J. Corrales, "Web Resources Recommendation based on Dynamic Prediction of User Consumption on the Social Web," 2012.
- [18] J. Zhuang, T. Mei, S. C. Hoi, Y.-Q. Xu and S. Li, "When recommendation meets mobile: contextual and personalized recommendation on the go," New York, NY, USA, 2011.