

Perfil de Uso de Aplicativos Móveis: Caracterização e Aplicações

Augusto C. S. A. Domingues¹, Fabrício A. Silva², Thais Regina M. B. Silva², Antonio A. F. Loureiro¹

¹Departamento de Ciência da Computação - Universidade Federal de Minas Gerais

²Universidade Federal de Viçosa - *Campus Florestal*

{augusto.souza, loureiro}@dcc.ufmg.br, {fabricio.asilva, thais.braga}@ufv.br

Resumo. *O grande volume de dados gerados pelos avanços das tecnologias móveis tem feito com que os provedores de serviço se interessem cada vez mais pela sua coleta e análise. Neste trabalho, um conjunto de dados real e de larga escala relacionado ao uso detalhado de aplicativos móveis é investigado pela primeira vez com o objetivo de se identificar padrões de acesso, de tráfego de dados, de tempo de uso e de transição entre serviços. Além da caracterização, este trabalho contribui com a comunidade científica em outras duas frentes. Primeiramente, foi proposto e validado um modelo para geração de dados sintéticos, que poderá ser adotado por outros pesquisadores. Além disso, foi possível avaliar as métricas relevantes para a predição do próximo aplicativo a ser utilizado, o que permite a prévia preparação de algum aplicativo para reduzir o tempo de espera dos usuários, aumentando a sua satisfação.*

Abstract. *The large amount of data generated by the advances of mobile technology makes the service providers get more and more interested in collecting and analyzing it. In this study, we investigate for the first time a real and large-scale dataset related to the use of mobile applications to identify patterns of access, data traffic, period of use, and transition among services. The characterization results reveal that different services are consumed differently by their users. In addition to the characterization, the contributions of this work are two-fold. First, we propose and validate a synthetic data generator that can be used by other researchers. Also, we evaluate which metrics affect the prediction of the next application to be launched, reducing then the users' waiting time.*

1. Introdução

O avanço das tecnologias móveis e sem fio dos últimos anos fez surgir uma nova frente de pesquisa que foca na análise de grandes volumes de dados [Laurila et al. 2012]. Por um lado, a redução no custo dos dispositivos móveis, principalmente *smartphones*, possibilitou um aumento significativo no número de usuários desses equipamentos. Por outro lado, as empresas que oferecem serviços móveis perceberam a importância em se conhecer os seus usuários, e passaram então a coletar dados de seus clientes.

Um dos grandes interesses na análise de dados está relacionado ao conhecimento de padrões de uso dos serviços móveis, dada a sua importância em diferentes aspectos

[Yang et al. 2015]. Por exemplo, operadoras de redes celulares podem tomar melhores decisões de investimento em infraestrutura de acordo com a demanda. Além disso, as operadoras podem oferecer planos de dados específicos para atender cada perfil de uso, aumentando a retenção de clientes. Empresas de propaganda podem direcionar suas divulgações para usuários com algum perfil específico, aumentando assim a efetividade da propaganda. Esse tipo de empresa pode ainda utilizar os estudos para identificar os melhores dias e horários para divulgações, diminuindo a repulsão dos usuários por propagandas em momentos inadequados. Por fim, diferentes segmentos de empresas podem utilizar as análises para conhecer melhor o perfil de seus usuários e, assim, prover melhores serviços.

O principal objetivo deste trabalho é caracterizar o perfil de uso de serviços móveis - com ênfase nas aplicações - a partir de grandes volumes de dados reais coletados diretamente de *smartphones* de mais de 5000 usuários durante o período de Janeiro a Dezembro de 2014. Como resultados, espera-se responder às seguintes perguntas:

1. Qual a frequência e duração de uso de alguns dos principais aplicativos móveis?
2. Existe algum padrão de navegação entre os aplicativos?
3. Qual o impacto de cada aplicativo na rede em termos de dados trafegados?
4. Como o padrão de uso está relacionado com o dia da semana e hora do dia?

Com base nos resultados de caracterização, são apresentadas duas contribuições relevantes para a comunidade científica. Em primeiro lugar, é proposto e validado um modelo que representa o uso detalhado de aplicativos móveis, e pode ser usado para gerar dados sintéticos. Além disso, diferentes métricas são avaliadas para a predição de qual o próximo aplicativo a ser utilizado. Esse resultado ajuda na preparação de aplicativos antes de seu uso, reduzindo o tempo de espera do usuário e aumentando a sua satisfação.

Este texto está organizado da seguinte forma. A Seção 2 apresenta os principais trabalhos relacionados. Na Seção 3 são descritos os detalhes do conjunto de dados analisados. A Seção 4 faz a análise desses dados com base em métricas de uso de aplicações móveis. A Seção 5 apresenta as contribuições práticas da caracterização. Finalmente, as conclusões do trabalho são apresentadas na Seção 6.

2. Trabalhos Relacionados

As redes sociais são fontes interessantes para o estudo de perfil de usuário, e são exploradas em alguns trabalhos [Xu et al. 2015, Fiadino et al. 2015, Fernandez et al. 2014, Jin et al. 2013]. O objetivo desses trabalhos é entender como o comportamento dos usuários em termos de redes sociais pode ser utilizado para melhoria dos serviços providos. Diferentemente dos estudos citados, o presente trabalho visa caracterizar e identificar perfis de usuários considerando não somente as redes sociais, mas uma lista mais ampla de aplicativos. Além disso, este trabalho utiliza dados de duração de um ano coletados por um agente instalado no *smartphone* do usuário, gerando assim informações mais precisas do que os trabalhos citados, que utilizam dados coletados do tráfego de rede.

Outros estudos exploram não somente as redes sociais para identificar perfis de usuários, mas também outros aspectos como o tráfego HTTP [Yang et al. 2015], a troca de mensagens SMS [de Almeida Oliveira et al. 2015], o padrão de tráfego [Wang et al. 2015, Lim et al. 2016], o comportamento do usuário [Chittaranjan et al. 2013], padrões de mobilidade [Leo et al. 2016, Pavan et al. 2015, Hong et al. 2015] e ligações feitas nos arredores de um estádio de futebol [Xavier et al. 2012]. Apesar desses trabalhos realizarem

Tabela 1. Características dos principais trabalhos relacionados

Trabalho	Local da coleta	Critério de análise	Escala	Duração
[Yang et al. 2015]	Provedor	Tráfego de dados	≈ 4.500.000	1 semana
[de Almeida Oliveira et al. 2015]	Provedor	Padrões de uso de SMS	≈ 20.000	1 semana
[Wang et al. 2015]	Provedor	Tráfego de dados	≈ 150.000	1 mês
[Malmi e Weber 2016]	Provedor	Uso de aplicações	≈ 3.500	1 mês
[Xavier et al. 2012]	Provedor	Padrões de mobilidade	≈ 30.000	3 dias
[Leo et al. 2016]	Provedor	Padrões de mobilidade	≈ 7.000.000	1 ano
[Lim et al. 2016]	WiFi AP	Padrões Uso	≈ 3.500	4 meses
[Pavan et al. 2015]	GPS	Padrões de mobilidade	13	4 dias
[Hong et al. 2015]	GPS	Padrões de mobilidade	≈ 1.000.000	30 dias
[Xu et al. 2015]	<i>Smartphone</i>	Uso de aplicações	≈ 25	6 semanas
[Chittaranjan et al. 2013]	<i>Smartphone</i>	Características pessoais	≈ 120	17 meses
[Li et al. 2015]	<i>Smartphone</i>	Tráfego de dados	≈ 2.000.000	1 mês
[Do et al. 2011]	<i>Smartphone</i>	Uso de aplicativos	77	8 meses
Trabalho Atual	<i>Smartphone</i>	Uso detalhado de aplicativos	5.342	1 ano

a caracterização de perfis de usuários, os dados usados no presente estudo permitem uma elaboração mais precisa de perfis sob diferentes perspectivas, com informações detalhadas de uso coletadas diretamente do *smartphone* de um grande número de usuários por um longo período.

Existem também trabalhos que tentam identificar perfis de usuários de acordo com os aplicativos instalados em seus dispositivos [Malmi e Weber 2016, Li et al. 2015, Seneviratne et al. 2014, Do et al. 2011]. Esses trabalhos tentam inferir características dos usuários, como gênero, idade e religião considerando os aplicativos instalados. Diferentemente, o presente trabalho não analisa somente a existência ou não de aplicativos, mas também detalhes da interação do usuário com os mesmos.

A Tabela 1 apresenta uma comparação entre os principais trabalhos relacionados. O trabalho atual possui como diferenciais a coleta de dados mais detalhados do uso de aplicativos, feita diretamente dos dispositivos móveis dos usuários. Além disso, os dados utilizados possuem uma duração de um ano, o que diminui algum viés dependendo da época da coleta, e considera 5,342 usuários, que é uma amostra significativa.

3. Conjunto de dados

O conjunto de dados usado contém informações anônimas relacionadas ao uso de aplicativos móveis em *smartphones*. A coleta foi feita por uma empresa da área de telefonia móvel por meio de um agente instalado em dispositivos *Android* de usuários voluntários. Essa coleta ocorreu de Janeiro a Dezembro de 2014, em algumas regiões do Brasil. O conjunto de dados foi coletado de 5.342 usuários de planos pós-pago. Cada registro do conjunto de dados representa um evento realizado pelo usuário utilizando a rede móvel e contém os seguintes campos:

- *Usuário*: um identificador único e anônimo para cada usuário;
- *Aplicação*: nome da aplicação usada;
- *Bytes Recebidos*: total de *bytes* recebidos ao longo da duração do evento;
- *Bytes Transmitidos*: total de *bytes* transmitidos ao longo da duração do evento;
- *Data de início*: Data e hora de início do evento;

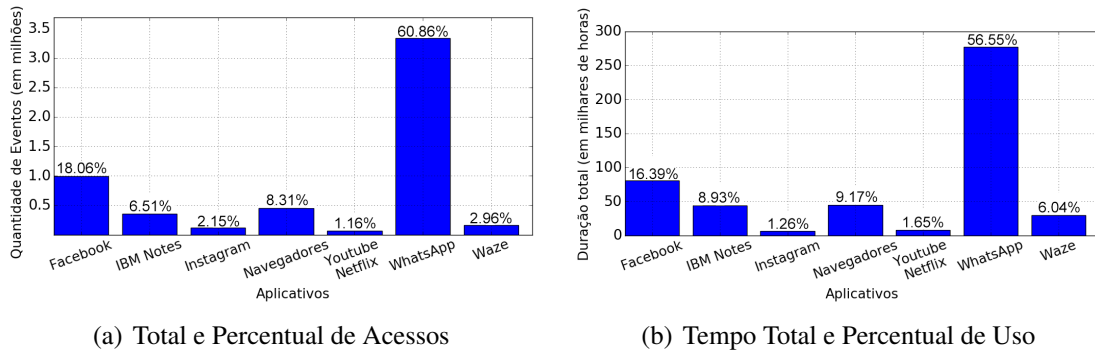


Figura 1. Acessos e tempo de uso dos aplicativos

- *Data de término*: Data e hora de término do evento.

Os aplicativos monitorados foram *Facebook*, *Instagram*, *WhatsApp*, *IBM Notes/Verse*, *Waze*, *Youtube*, *Netflix* e *Navegadores* (*Chrome*, *Firefox*, *Opera*, *Built-in Android Browser*). Durante o período de monitoração, foi gerado um evento com os campos descritos todas as vezes que o usuário acessou (colocou em primeiro plano) um dos aplicativos monitorados. Com isso, é possível conhecer detalhes do uso dos aplicativos pelos usuários.

Os dados utilizados neste trabalho foram coletados de usuários de planos pós-pagos de uma operadora de celular. Com isso, a representatividade das análises não engloba os usuários de planos pré-pagos, que podem apresentar um perfil de uso diferente. No Brasil, os usuários de planos pós-pagos representam aproximadamente 33% dos contratos das operadoras [TEL 2017]. Porém, em termos de valores de receita, representam um percentual significativo para as operadoras. Isso mostra que, apesar de englobar um percentual menor da população, os dados analisados são relevantes para as operadoras e provedores de serviços móveis, e portanto, importantes de serem analisados.

Em relação à distribuição espacial, a maioria dos usuários monitorados se concentram na região Sudeste do Brasil, principalmente nas cidades de São Paulo e Rio de Janeiro. Apesar de representarem uma pequena parte geográfica do país, essa região é responsável por 60% do PIB, sendo que as duas cidades de maior abrangência nos dados são responsáveis por 14,45% do PIB e 8,99% da população do país.

Em resumo, os resultados apresentados neste trabalho podem não representar toda a população brasileira em todas as suas regiões. Por se tratar de um território extenso, sabe-se das particularidades de cada região, estado e até mesmo cidade. Assim sendo, as análises representam uma parcela da população que utiliza planos pós-pagos na região Sudeste do país, parcela essa que tem uma significativa importância para a economia.

4. Caracterização do Uso de Aplicativos Móveis

4.1. Perspectiva de Acessos

Identificar a aplicação que os usuários estão mais interessados é útil para prever o comportamento dos mesmos [Yang et al. 2015]. Nesta subseção analisamos a quantidade de eventos e a duração dos eventos por aplicativo.

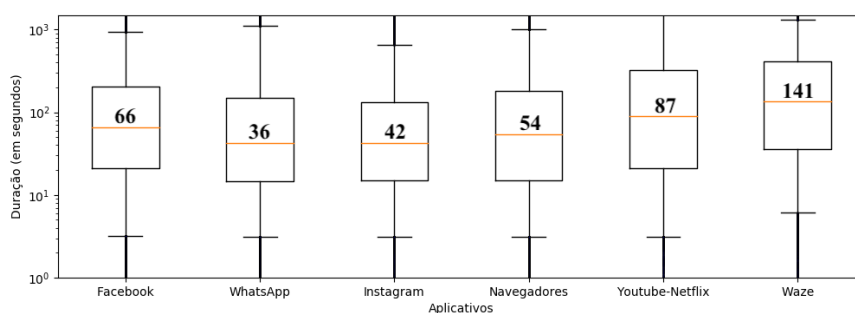


Figura 2. Duração dos eventos por aplicativo

4.1.1. Quantidade de Acessos

Observando a Figura 1(a), é possível notar que as aplicações com o maior número de acessos são *WhatsApp*, *Facebook*, e os Navegadores, nessa ordem. De todas as aplicações estudadas, os Navegadores são os únicos que se encaixam no perfil de qualquer usuário, dada a sua natureza de uso geral, o que justifica sua colocação em terceiro lugar em termos de número de acessos. Em termos percentuais, a Figura 1(a) ilustra que *WhatsApp* representa mais de 60% dos eventos e o *Facebook* aproximadamente 18%. Isso demonstra a popularidade desses dois aplicativos, que vem sendo cada vez mais utilizados no Brasil.

4.1.2. Duração dos Acessos

A análise do tempo de uso por aplicativo é apresentada na Figura 1(b). Novamente, os aplicativos com maior tempo total de uso são *WhatsApp* (56,55%), *Facebook* (16,39%) e os Navegadores (9,17%). Vale observar que, apesar da quantidade de eventos do *WhatsApp* representar mais de 60% do total, a duração de uso representa 56,55%. Ou seja, os eventos do *WhatsApp* tendem a ser menores que de outros aplicativos.

A Figura 2 detalha melhor a distribuição da duração de uso por evento por aplicativo. Apesar de todas possuírem comportamento semelhante, podemos destacar algumas particularidades. *Youtube/Netflix* e *Waze* apresentam uma mediana de duração de acessos maior quando comparados aos outros aplicativos, de 87 s e 141 s, respectivamente. Isso mostra que essas aplicações possuem um padrão de acesso prolongado. Usuários do *Waze* tendem a acessá-lo por longos períodos de tempo durante um trajeto. Já os usuários do *Youtube* e *Netflix* os utilizam para atividades de entretenimento, como assistir vídeos e filmes. Na Figura 1(a), podemos observar que os eventos de *Waze* são responsáveis por 2,96% dos eventos analisados, mas a Figura 1(b) mostra que os mesmos possuem uma fatia maior (6,04%) quanto ao tempo total de uso dos eventos analisados. Por outro lado, *WhatsApp* e *Instagram* apresentam medianas de aproximadamente 42 s.

4.2. Perspectiva de Navegação entre Aplicativos

A análise da navegação entre aplicativos é importante para que se conheça quais são os fluxos de uso dos usuários. Para os provedores de serviços, tais informações podem ajudar a prover planos mais personalizados que favoreçam aplicativos correlacionados.

A Figura 3 mostra um mapa de nível gerado a partir das transições entre aplicativos de todos os usuários. Cada região do mapa representa a probabilidade de transição

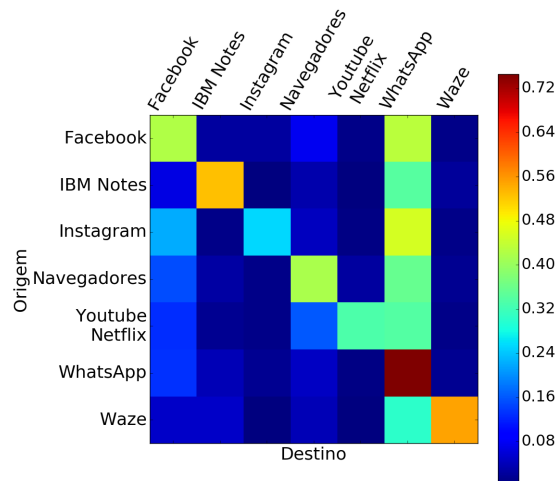


Figura 3. Mapa de nível das transições entre aplicativos

entre dois aplicativos (o eixo Y representa o último aplicativo usado e o eixo X o próximo a ser acessado). Nota-se que, ao sair do *WhatsApp*, o usuário tem cerca de 72% de probabilidade de acessá-lo novamente. Isso demonstra que a aplicação é de uso recorrente, ou seja, o usuário realiza múltiplos acessos seguidos (neste caso, para visualizar novas mensagens e respondê-las). Outras transições com alta recorrência são *IBM Notes* para *IBM Notes* (uso frequente no trabalho), e *Waze* para *Waze* (uso frequente no trânsito).

Outra análise que pode ser feita com base na Figura 3 são as transições que ocorrem para os aplicativos *WhatsApp* (vistas na coluna *WhatsApp*) e *Facebook* (vistas na coluna *Facebook*). Pode-se destacar que grande parte dos usuários tendem a realizar acessos alternados entre suas aplicações preferidas e o *Facebook* e *WhatsApp*. Isto se explica pela popularidade dessas aplicações, que são da categoria dominante na Internet móvel [Yang et al. 2015].

4.3. Perspectiva de Dados Trafegados

Conhecer o comportamento dos usuários quanto ao uso de dados é uma necessidade urgente para os provedores [Yang et al. 2015]. Da mesma forma, [Wang et al. 2015] afirmam que entender os padrões de tráfego de dados em centros urbanos de larga escala é extremamente valioso para os provedores de Internet e usuários móveis. Assim, esta subseção analisa os eventos em relação aos dados trafegados, para que se conheça as aplicações que causam um maior impacto nas redes móveis.

As Figuras 4(a) e 4(b) apresentam a quantidade total de dados enviados e recebidos para cada aplicativo, respectivamente. Com base nelas, é possível observar que as aplicações com a maior quantidade de *upload* e *download* são *Facebook* e *WhatsApp*. Esses aplicativos permitem que os usuários enviem e recebam arquivos multimídia, como fotos, vídeos e músicas, o que justifica o uso elevado de dados.

Considerando cada acesso individualmente, os aplicativos de entretenimento (*Youtube* e *Netflix*) recebem mais dados que os demais, apresentando uma mediana de 48 KBytes em comparação com 37 KBytes do *Facebook* e somente 0,8 KBytes do *WhatsApp*. Porém, a frequência de uso significativamente maior desses dois últimos faz com que o tráfego total deles seja maior. O tráfego relativamente baixo para os eventos individuais de entre-

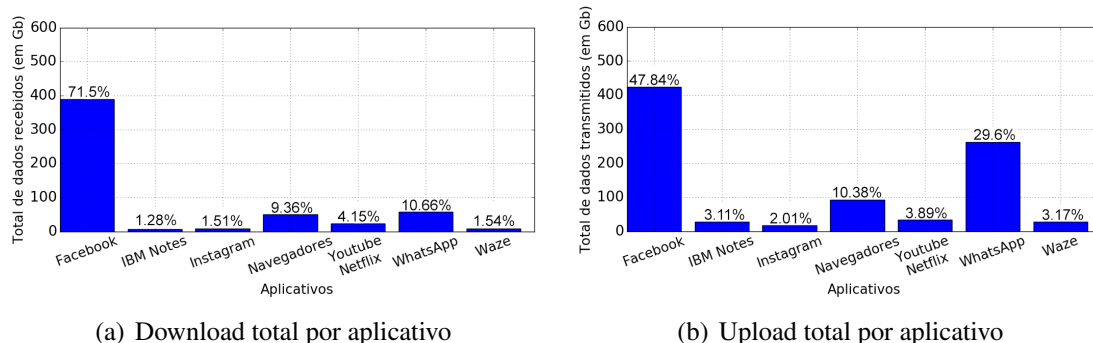


Figura 4. Análise dos dados trafegados

tenimento contrasta com a alta demanda de dados por parte das aplicações de *streaming* de vídeo, sugerindo que muitos usuários utilizam a rede móvel para entretenimento de maneira controlada, deixando o consumo mais pesado para momentos em que estão conectados ao WiFi, por exemplo.

Outra observação interessante é que, mesmo o *WhatsApp* sendo responsável por 60,86% do número de eventos e 56,55% da duração total (ver Figuras 1(a) e 1(b)), ele representa apenas 10,66% do *download* e 29,6% do *upload* total. Por outro lado, o *Facebook* é responsável por 18,06% do número de eventos e 16,39% da duração total, mas representa mais de 71% do *download* e mais de 47% do *upload*. Esses resultados revelam que os usuários, em geral, dedicam mais tempo ao *WhatsApp*, mas demandam mais recursos da rede para acessarem o *Facebook*.

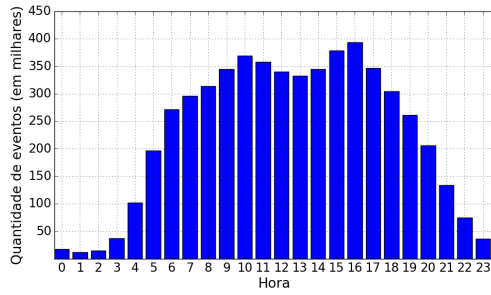
4.4. Perspectiva de Dia da Semana e Hora do Dia

A perspectiva temporal também é um fator importante para a análise do comportamento dos usuários. Com base nela, é possível identificar e modelar padrões de acesso quanto a hora do dia e o dia da semana. Esses padrões podem ser usados por provedores para otimizar os recursos da rede [Fiadino et al. 2015].

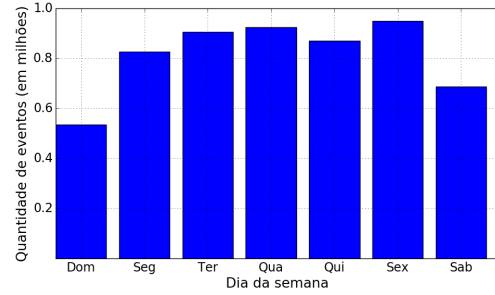
A Figura 5(a) mostra a distribuição temporal dos eventos em relação às horas do dia. É possível observar que eles acontecem em maior quantidade das 06:00h (horário que muitos usuários saem para o trabalho) até às 19:00h (horário que os mesmos retornam para casa). Nesse intervalo, temos um pico às 10:00h, uma ligeira queda nos acessos de 11:00h até às 13:00h, e novamente um pico às 16:00h. Ou seja, aparentemente os usuários tendem a diminuir o fluxo de acessos durante o horário de almoço.

Similarmente, a Figura 5(b) apresenta a distribuição temporal dos eventos em relação aos dias da semana. Nela, nota-se que nos finais de semana o número de acessos é menor do que nos dias úteis. Isso implica que os usuários diminuem o uso de aplicativos móveis em seu tempo livre (o que também pode ser visto na Figura 5(a)).

A Figura 6 apresenta uma análise individual por aplicativo. Pode-se observar, na Figura 6(a), que os usuários do *Facebook* realizam muitos acessos durante todos os dias da semana, com um crescimento durante as tardes. Para o *WhatsApp* (Figura 6(b)), temos um padrão de acesso semelhante, mas com uma queda nos acessos nos finais de semana. Inversamente, os usuários do *Instagram* tendem a acessar a aplicação com maior frequência nos finais de semana (Figura 6(c)).

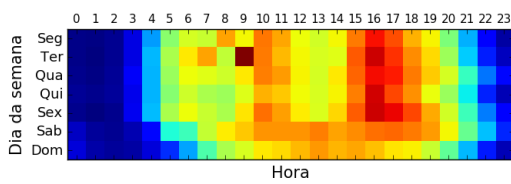


(a) Eventos por hora do dia

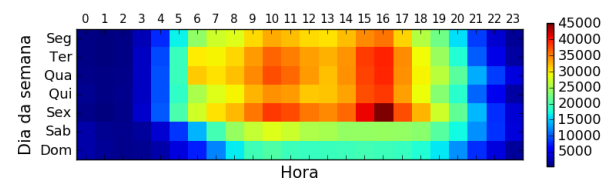


(b) Eventos por dia da semana

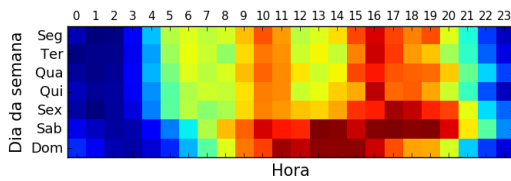
Figura 5. Análise temporal dos dados



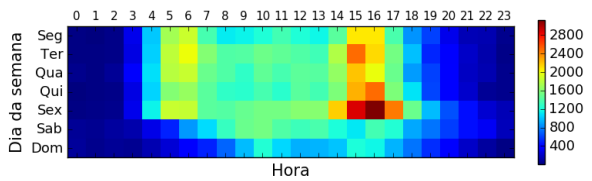
(a) Facebook



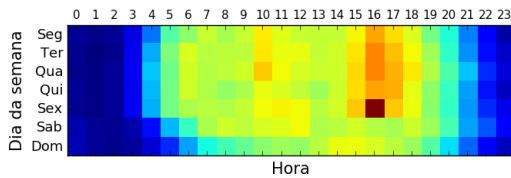
(b) WhatsApp



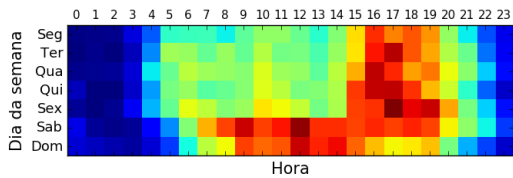
(c) Instagram



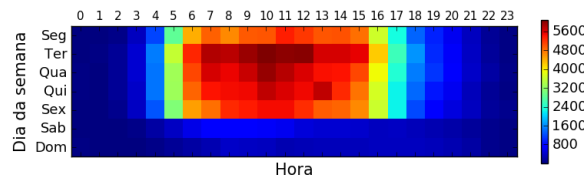
(d) Waze



(e) Navegadores



(f) Youtube/Netflix



(g) IBM Notes

Figura 6. Acessos agrupados por dia e hora

Para o *Waze*, a Figura 6(d) mostra que seus acessos ocorrem com muita frequência no fim da tarde, principalmente às sextas-feiras. Esse comportamento incomum é explicado pelo fato do aplicativo ser usado para a mobilidade no trânsito. Com isso, conclui-se que a frequência de acesso está relacionada à necessidade do usuário de saber informações sobre o trânsito. Mais além, os horários de pico de acesso podem revelar eventos

atípicos, como congestionamentos e acidentes.

Os aplicativos de navegação na *Web* (Figura 6(e)) possuem uma distribuição homogênea de acessos, abrangendo grande parte do dia e da semana. Este padrão pode ser explicado pela generalidade desses aplicativos, já que a necessidade do uso pode ocorrer em diversos contextos.

No caso dos aplicativos *Youtube* e *Netflix* (Figura 6(f)), é visível o maior número de acessos durante o fim da tarde/noite e principalmente durante o fim de semana. Esse padrão compreende o período de descanso dos usuários. Já para o *IBM Notes* (Figura 6(g)), que contém funcionalidades que são usadas durante o horário de trabalho, o padrão é ainda mais claro: muitos acessos de 05:00h até às 17:00h de Segunda à Sexta-Feira, sem nenhum outro pico, e poucos acessos durante o fim de semana.

5. Aplicações

Esta seção apresenta duas contribuições oriundas da análise e caracterização dos dados apresentados. Primeiramente, é apresentado um modelo para gerar dados sintéticos. Em seguida, são avaliadas métricas para prever o próximo aplicativo a ser usado.

5.1. Gerador de Dados Sintéticos

A maioria das soluções inovadoras propostas para redes móveis demandam uma validação antes de entrarem na fase de produção. Usualmente, simulações são usadas para tais validações, que requerem dados de entrada realísticos para representar a demanda esperada. Uma das questões mais desafiadoras neste processo de validação baseia-se na geração de tais dados realísticos. Para ajudar nessa tarefa, foi criado um modelo para construir dados sintéticos realistas baseados nos resultados apresentados neste trabalho.

5.1.1. Geradores Existentes

Nos últimos anos, alguns estudos propuseram geradores de dados para diversos contextos. *Big Data Generator Suite* [Wang et al. 2014] é uma ferramenta que produz enormes conjuntos de dados específicos para uso em *benchmarking* de aplicações de mineração de dados. [Anderson et al. 2014] propõem um *framework* para gerar dados sensoriais sintéticos em larga escala que possam ser usados pela Internet das Coisas (IoT). Finalmente, [Iqbal et al. 2014] usam detalhes de registros de chamadas para produzir matrizes de origem-destino, que simulam onde os usuários de telefonia móvel acessam serviços móveis, e como eles se movem.

As soluções existentes demonstram a importância dos geradores de dados sintéticos em diferentes áreas de pesquisa. No entanto, não existem geradores de dados sintéticos na literatura que produzem dados realistas representando o uso detalhado de aplicações móveis. Isso reforça a ideia de que o modelo proposto neste trabalho é uma contribuição significativa para a comunidade.

5.1.2. O modelo

O objetivo do modelo é gerar registros sintéticos representando os acessos dos usuários às aplicações. As métricas cobertas pelo modelo são: duração de cada registro, demanda

de *download* para cada registro, demanda de *upload* para cada registro, e tempo de inter-
acesso para cada aplicação. O tempo de inter-acesso representa o intervalo entre dois
acessos de uma mesma aplicação em um mesmo dia.

A Tabela 2 sumariza o modelo para todas as métricas. Para cada uma dessas métri-
cas, foi feita uma estimativa de parâmetros de distribuição, e foi identificado que a distri-
buição *log-normal*(μ, σ) representa bem todas elas, com diferente parâmetros. Os valores
dos parâmetros para cada métrica e aplicação são apresentados na Tabela 2. Ressalta-
se que para dados de *download* e *upload*, os valores zero são tratados separadamente e
também é apresentada a probabilidade de haver esses valores ($P(\text{zeros})$).

5.1.3. Validação

O modelo foi validado com base no procedimento estatístico *teste-t* [Jain 1991], que com-
para se dois conjuntos de dados são oriundos de uma mesma população. A hipótese do
teste é que os dois conjuntos (real e gerado) foram gerados com base na mesma distribui-
ção, e calcula o intervalo de confiança e o *p-value* da diferença das médias. Em resumo,
os dois conjuntos de dados podem ser considerados da mesma distribuição se o intervalo
de confiança incluir o zero, com um valor alto de *p-value*.

Os resultados do intervalo de confiança (I.C.) podem ser vistos na Tabela 2. Pode-
se perceber que, para todas as métricas e aplicações, o intervalo de confiança de 95%
inclui o zero. Isso significa que o modelo gerado representa bem os dados reais, e pode ser
utilizado para a geração de dados sintéticos. Além disso, o *p-value* de todas as métricas
e aplicações foi maior que 0,74, o que significa que é muito improvável que os dois
conjuntos, real e gerado, tenham sido gerados a partir de distribuições diferentes.

5.2. Predição de próximo aplicativo

A predição do próximo aplicativo que os usuários vão abrir oferece implicações impor-
tantes para a qualidade de experiência dos mesmos. Primeiramente, é possível mostrar
os aplicativos mais prováveis na tela inicial do dispositivo para um acesso agilizado, re-
duzindo o tempo que os usuários gastam para procurá-los. Adicionalmente, as aplica-
ções podem realizar um pré carregamento de dados antecedendo sua abertura para uma
execução mais rápida. Finalmente, há uma economia de energia quando o tempo gasto
procurando por um aplicativo é reduzido e o *download* de dados é feito em segundo plano.

Alguns estudos na literatura propõem algoritmos para prever o próximo aplica-
tivo que um usuário irá abrir. [Srinivasan et al. 2014] apresentam um framework que exe-
cuta no dispositivo móvel e minera padrões de comportamento de usuários móveis usando
regras de associação. O framework usa regras de associação históricas para prever o pró-
ximo aplicativo a ser usado. [Zou et al. 2013] e [Shin et al. 2012] propõem o uso de pro-
babilidades condicionais para prever o próximo aplicativo a ser usado. Esses modelos
se baseiam nos dados históricos de uso das aplicações. O objetivo de [Huang et al. 2012]
é prever o próximo aplicativo que um usuário irá abrir baseado em informações contex-
tuais como tempo, localização e perfil.

No geral, a principal conclusão obtida desses estudos é que a adoção de probabi-
lidades condicionais dado o último ou os dois últimos aplicativos usados leva a uma boa

Tabela 2. Sumário do modelo para gerar dados sintéticos. O Intervalo de Confiança (I.C.) valida o modelo.

Métrica	Duração (s)	Download (Kbytes)	Upload (Kbytes)	Tempo de inter-acesso (s)
	log-normal	log-normal	log-normal	log-normal
IBM Notes/Verse	$\mu=3,83, \sigma=1,91$ I.C. [-0,006, 0,010]	$\mu=1,98, \sigma=1,79$; P(zeros)=0,46 I.C. [-0,011, 0,010]	$\mu=2,28, \sigma=2,01$ P(zeros)=0,08 I.C. [-0,010, 0,008]	$\mu=6,36, \sigma=2,30$ I.C. [-0,013, 0,009]
Instagram	$\mu=3,78, \sigma=1,62$ I.C. [-0,009, 0,016]	$\mu=2,96, \sigma=0,13$ P(zeros)=0,13 I.C. [-0,017, 0,015]	$\mu=3,54, \sigma=1,82$ P(zeros)=0,01 I.C. [-0,013, 0,016]	$\mu=6,31, \sigma=2,59$ I.C. [-0,023, 0,027]
Facebook	$\mu=4,14, \sigma=1,68$ I.C. [-0,003, 0,006]	$\mu=3,55, \sigma=2,11$ P(zeros)=0,07 I.C. [-0,007, 0,005]	$\mu=3,96, \sigma=1,88$ P(zeros)=0,01 I.C. [-0,004, 0,005]	$\mu=5,98, \sigma=2,59$ I.C. [-0,007, 0,008]
Browsers	$\mu=3,99, \sigma=1,76$ I.C. [-0,003, 0,010]	$\mu=3,83, \sigma=1,91$ P(zeros)=0,14 I.C. [-0,007, 0,010]	$\mu=3,81, \sigma=1,93$ P(zeros)=0,01 I.C. [-0,007, 0,008]	$\mu=5,65, \sigma=2,28$ I.C. [-0,012, 0,0136]
Waze	$\mu=4,76, \sigma=1,70$ I.C. [-0,015, 0,008]	$\mu=2,98, \sigma=1,76$ P(zeros)=0,09 I.C. [-0,014, 0,011]	$\mu=3,94, \sigma=1,67$ P(zeros)=0,01 I.C. [-0,010, 0,012]	$\mu=4,83, \sigma=2,55$ I.C. [-0,019, 0,020]
WhatsApp	$\mu=3,89, \sigma=1,75$ I.C. [-0,002, 0,002]	$\mu=0,28, \sigma=1,89$ P(zeros)=0,15 I.C. [-0,003, 0,003]	$\mu=2,16, \sigma=1,95$ P(zeros)=0,03 I.C. [-0,003, 0,002]	$\mu=5,71, \sigma=2,24$ I.C. [-0,003, 0,003]
Youtube	$\mu=4,38, \sigma=1,87$ I.C. [-0,021, 0,021]	$\mu=4,03, \sigma=2,30$ P(zeros)=0,11 I.C. [-0,023, 0,032]	$\mu=4,49, \sigma=2,09$ P(zeros)=0,01 I.C. [-0,025, 0,021]	$\mu=4,93, \sigma=2,68$ I.C. [-0,038, 0,042]

precisão com baixos custos computacionais. Em outras palavras, o histórico recente de uso de aplicativos é mais relevante que outros contextos, como localização, para prever o próximo. Até o momento, nenhum dos estudos considera o intervalo de inter-acesso em seus algoritmos. Adicionalmente, o número de usuários avaliados é pequeno, variando de 34 a 106.

Os objetivos desta seção estão relacionados aos pontos fracos dos estudos existentes. Primeiro, são avaliados os algoritmos existentes com os melhores resultados usando um conjunto de dados em larga escala. Este é o primeiro trabalho que avalia soluções para a predição do próximo aplicativo usando dados em larga escala. O segundo objetivo é avaliar como o tempo de inter-acesso afeta a predição.

Para isso, foram implementadas e avaliadas três soluções:

- Último Usado (*UU*): o próximo aplicativo é predito baseado no último aplicativo usado, ou seja, é calculado $P(app_i | app_j)$ para todo aplicativo app_i , onde app_j representa o último aplicativo usado;
- Últimos Dois Usados (*U2U*): o próximo aplicativo é predito baseado nos dois últimos aplicativos usados, ou seja, é calculado $P(app_i | app_j \wedge app_k)$ para todo aplicativo app_i , onde app_j e app_k são os dois últimos aplicativos usados, nesta ordem;
- Tempo de inter-acesso (*TIA*): o próximo aplicativo é predito baseado na Função de Distribuição Cumulativa (CDF) do tempo de inter-acesso dos aplicativos, como $P(app_i) = F(app_i, t)$, onde $F(app, t)$ é a CDF de app dado que se passaram t unidades de tempo desde a última vez em que o mesmo aplicativo foi acessado.

A Tabela 3 apresenta os resultados obtidos. Primeiro, foram usados 80% dos

Tabela 3. Resultados da predição de próximo aplicativo. De fato, os aplicativos usados mais recentemente afetaram mais a predição. O tempo de inter-acesso não possui grande impacto.

	Precisão	
	Próximo app	Próximos dois apps
UU	67,03%	89,82%
U2U	67,80%	90,28%
TIA	60,76%	82,26%

dados para treinar o modelo probabilístico, e computar a precisão nos 20% restantes. Quando o primeiro aplicativo predito está correto, considera-se um acerto na precisão da predição do próximo *app* (primeira coluna). A segunda coluna (Próximos dois apps) da tabela indica a precisão quando um dentre os dois aplicativos mais prováveis está entre o correto. Ou seja, dentre duas indicações, uma está correta.

Duas principais conclusões podem ser levantadas desses resultados. Primeiro, os aplicativos usados recentemente são de fato uma maneira simples porém eficiente para prever o próximo aplicativo a ser usado. Segundo, o tempo de inter-acesso de um aplicativo não possui efeito relevante na predição.

6. Conclusões e Trabalhos Futuros

Neste trabalho, são apresentadas análises e caracterizações do comportamento de mais de 5000 usuários de *smartphones* do Brasil, monitorados durante um ano. Foi possível observar que aplicações de redes sociais são responsáveis por grande parte da quantidade total de eventos gerados e do tempo total de uso. Foi observado também que os usuários tendem a realizar acessos alternados entre suas aplicações prediletas e as redes sociais, reforçando a ideia de que esta categoria é a dominante na *Internet* móvel. Adicionalmente, foi apresentado que, apesar de ser responsável pela maior parte dos acessos, o *WhatsApp* representa somente uma pequena fatia do tráfego total de dados, indicando que a popularidade não é um fator de alta relevância na demanda de dados de um aplicativo. Por fim, a análise temporal demonstrou que as aplicações seguem padrões bem definidos de hora e dia de uso, principalmente as pertencentes às categorias *Entretenimento* e *Escritório*.

Os resultados alcançados podem ser explorados de diferentes formas pelos provedores de serviços móveis. Neste trabalho, são apresentadas duas aplicações. Primeiro, a caracterização do comportamento dos usuários foi usada para definir um modelo de geração de dados sintéticos realistas de acessos a aplicações móveis. Esses dados podem ser usados pela comunidade científica em simulações de redes móveis. Segundo, foram avaliados quais são os fatores que influenciam na escolha do próximo aplicativo a ser acessado pelo usuário. Esse conhecimento permite que as aplicações sejam preparadas em segundo plano antes do acesso, resultando em um menor tempo de espera e uma melhor experiência de usuário.

Trabalhos futuros incluem a análise do conjunto de dados quanto aos padrões de chamadas de voz realizadas, uma vez que este artigo focou no comportamento dos usuários quanto aos acessos a aplicativos. Os conhecimentos obtidos neste trabalho também serão utilizados como entrada para um modelo de aprendizado que visa classificar usuários de acordo com o perfil de uso.

7. Agradecimento

Este trabalho contou com o apoio da Fapemig, CNPq e FUNARBE.

Referências

- (2017). Teleco: Intelligence in telecommunications. <http://www.teleco.com.br/en/>. Accessed at 2017-06-02.
- Anderson, J. W., Kennedy, K., Ngo, L. B., Luckow, A., e Apon, A. W. (2014). Synthetic data generation for the internet of things. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 171–176. IEEE.
- Chittaranjan, G., Blom, J., e Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3):433–450.
- de Almeida Oliveira, R., Brandão, W. C., e Marques-Neto, H. T. (2015). Characterizing user behavior on a mobile sms-based chat service. In *Computer Networks and Distributed Systems (SBRC), 2015 XXXIII Brazilian Symposium on*, pages 130–139. IEEE.
- Do, T. M. T., Blom, J., e Gatica-Perez, D. (2011). Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 353–360. ACM.
- Fernandez, M., Scharl, A., Bontcheva, K., e Alani, H. (2014). User profile modelling in online communities. In *Proceedings of the Third International Conference on Semantic Web Collaborative Spaces-Volume 1275*, pages 1–15. CEUR-WS. org.
- Fiadino, P., Casas, P., Schiavone, M., e D’Alconzo, A. (2015). Online social networks anatomy: On the analysis of facebook and whatsapp in cellular networks. In *IFIP Networking Conference (IFIP Networking), 2015*, pages 1–9. IEEE.
- Hong, L., Zheng, Y., Yung, D., Shang, J., e Zou, L. (2015). Detecting urban black holes based on human mobility data. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 35. ACM.
- Huang, K., Zhang, C., Ma, X., e Chen, G. (2012). Predicting mobile application usage using contextual information. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1059–1065. ACM.
- Iqbal, M. S., Choudhury, C. F., Wang, P., e González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74.
- Jain, R. (1991). *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling*. Wiley professional computing. Wiley.
- Jin, L., Chen, Y., Wang, T., Hui, P., e Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine*, 51(9):144–150.
- Laurila, J. K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T.-M.-T., Dousse, O., Eberle, J., Miettinen, M., et al. (2012). The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing*, number EPFL-CONF-192489.

- Leo, Y., Busson, A., Sarraute, C., e Fleury, E. (2016). Call detail records to characterize usages and mobility events of phone users. *Computer Communications*, 95:43–53.
- Li, H., Lu, X., Liu, X., Xie, T., Bian, K., Lin, F. X., Mei, Q., e Feng, F. (2015). Characterizing smartphone usage patterns from millions of android users. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 459–472. ACM.
- Lim, K.-W., Secci, S., Tabourier, L., e Tebbani, B. (2016). Characterizing and predicting mobile application usage. *Computer Communications*, 95:82–94.
- Malmi, E. e Weber, I. (2016). You are what apps you use: Demographic prediction based on user’s apps. *arXiv preprint arXiv:1603.00059*.
- Pavan, M., Mizzaro, S., e Scagnetto, I. (2015). Mining movement data to extract personal points of interest: A feature based approach.
- Seneviratne, S., Seneviratne, A., Mohapatra, P., e Mahanti, A. (2014). Predicting user traits from a snapshot of apps installed on a smartphone. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):1–8.
- Shin, C., Hong, J.-H., e Dey, A. K. (2012). Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 173–182. ACM.
- Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K. K., Xu, C., e Tapia, E. M. (2014). Mobileminer: Mining your frequent patterns on your phone. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 389–400. ACM.
- Wang, H., Xu, F., Li, Y., Zhang, P., e Jin, D. (2015). Understanding mobile traffic patterns of large scale cellular towers in urban environment. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 225–238. ACM.
- Wang, L., Zhan, J., Luo, C., Zhu, Y., Yang, Q., He, Y., Gao, W., Jia, Z., Shi, Y., Zhang, S., et al. (2014). Bigdatabench: A big data benchmark suite from internet services. In *High Performance Computer Architecture (HPCA), 2014 IEEE 20th International Symposium on*, pages 488–499. IEEE.
- Xavier, F. H. Z., Silveira, L. M., Almeida, J. M. d., Ziviani, A., Malab, C. H. S., e Marques-Neto, H. T. (2012). Analyzing the workload dynamics of a mobile phone network in large scale events. In *Proceedings of the first workshop on Urban networking*, pages 37–42. ACM.
- Xu, R., Frey, R. M., Vuckovac, D., e Ilic, A. (2015). Towards understanding the impact of personality traits on mobile app adoption—a scalable approach. In *23rd European Conference on Information Systems*.
- Yang, J., Qiao, Y., Zhang, X., He, H., Liu, F., e Cheng, G. (2015). Characterizing user behavior in mobile internet. *IEEE Transactions on Emerging Topics in Computing*, 3(1):95–106.
- Zou, X., Zhang, W., Li, S., e Pan, G. (2013). Prophet: What app you wish to use next. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 167–170. ACM.