

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

Reconhecimento de Comportamento de Usuários de Telefone Celular

BOLSISTA: Oswald Mesumbe Ekwoge, FAPEAM

MANAUS

2014

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL

PIB – E – 0208/2013

Reconhecimento de Comportamento de Usuários de Telefone Celular

Bolsista: Oswald Mesumbe Ekwoqe, FAPEAM

Orientadora: Profa. Dra. Eulanda Miranda dos Santos

MANAUS

2014

Resumo

Este projeto investigou a aplicação de técnicas de reconhecimento de padrões ao problema de classificação de usuários de telefones móveis em função da capacidade de disseminação de informação de cada usuário. As seguintes atividades foram realizadas: 1) busca por bases de dados públicas para o problema investigado; 2) levantamento de atributos relevantes; 3) organização de uma base de dados própria; e 4) comparação do desempenho de técnicas de reconhecimento de padrões na base gerada no projeto.

Os resultados obtidos na primeira atividade indicaram que existem poucas bases de dados com ênfase na identificação de comportamento de usuários de telefonia celular disponíveis publicamente. A segunda atividade mostrou que é possível destacar os seguintes comportamentos de usuários de telefonia celular: Extroversão, Socialização, Neuroticismo, Consciência e Abetura. Por outro lado, alguns atributos só podem ser obtidos por meio de sensores incorporados aos telefones celulares dos usuários, especialmente *smartphones*. Devido à dificuldade de obtenção desses equipamentos, na terceira etapa deste projeto, optamos por obter informações por meio do uso de questionários. As questões respondidas pelos usuários foram definidas com base na revisão da literatura. Um questionário *online* foi elaborado e distribuído. A partir dos questionários respondidos, foi feita a coleta dos dados obtidos a fim de gerarmos uma base de dados para ser usada nos experimentos. Porém, apesar da literatura indicar a existência de 05 classes de comportamento de usuário de telefonia móvel, há muita interseção entre essas classes. Diante desse fato, nós reduzimos o problema para identificar apenas dois grupos de comportamento: comportamento recomendado ou comportamento não recomendado. Esses dois grupos de comportamento estão relacionados com o objetivo geral do projeto: encontrar uma rota entre dois dispositivos desconectados por meio da descoberta de hábitos humanos e padrões em sua conectividade. Dois métodos de agrupamento foram utilizados para o teste e avaliação na quarta etapa do projeto: *Expectation Maximization* (EM) e *KMeans*. Os métodos de agrupamento foram utilizados porque não havia dados rotulados. Os resultados obtidos pelo método EM indicaram que mais de 60% dos usuários investigados podem ser recomendados como bons elementos de transmissão de informação.

SUMÁRIO

1. Introdução	6
2. REVISÃO BIBLIOGRÁFICA	7
3. MÉTODOS UTILIZADOS	9
3.1. <i>Pesquisa Bibliográfica</i>	<i>9</i>
3.2. <i>Obtenção da base</i>	<i>9</i>
3.3. <i>Métodos de agrupamento.....</i>	<i>9</i>
4. RESULTADOS E DISCUSSÕES	10
5. CONCLUSÕES.....	14
6. REFERÊNCIAS BIBLIOGRÁFICAS	Erro! Indicador não definido.

1. Introdução

O tipo de comunicação conhecido como Redes Tolerantes a Atrasos e Desconexões (DTN - *Delay and Disruption Tolerant Networking*), tem despertado o interesse da comunidade científica para fornecer conectividade para áreas isoladas. A tecnologia DTN utiliza-se da mobilidade dos usuários para o tráfego assíncrono de mensagens, isto é, usuários móveis podem receber, transportar e entregar mensagens a nós intermediários ou ao nó destino, que se encontram isolados, tudo isso de uma forma automática e transparente.

Dentre as diversas possibilidades de uso da mobilidade de usuários para tráfego de mensagens via DTN, uma opção que se apresenta muito viável é estudar e usar a mobilidade e o comportamento social de moradores da região de interesse para disseminar informações e aumentar o alcance da comunicação via DTN. A comunicação pode ser viabilizada e as informações disseminadas utilizando-se dispositivos móveis como telefones celulares e *tablets*, configurados para operar no modo DTN.

Nesse contexto, o aprendizado e a identificação de padrões de comportamento de usuários de dispositivos móveis tornam-se fundamentais para o aprendizado e a identificação de padrões de mobilidade e de comportamento social de moradores da região de interesse, uma vez que os movimentos dos dispositivos de comunicação devem espelhar os movimentos de seus proprietários. Portanto, encontrar uma rota entre dois dispositivos desconectados implica em descobrir hábitos em movimentos humanos e padrões em sua conectividade tais como frequência de contatos, duração média de um contato, e rede de contatos. Essas informações podem ser usadas para prever futuros encontros e identificar melhores candidatos para a retransmissão dos dados para o destino. Portanto, existe um desafio em estudar comportamento humano, e, especificamente, a sua aplicação para a pesquisa de redes DTN (DMYTRO *et. al.*,2011).

Este projeto é parte componente de um projeto de pesquisa bem mais amplo, cujo título é “Inclusão Social de Comunidades Ribeirinhas da Amazônia Utilizando Redes Tolerantes a Atrasos e Desconexões (*Delay Tolerant Networks - DTN*)”, e tem como objetivo estudar a aplicação de técnicas de reconhecimento de padrões ao problema de reconhecimento de comportamento de usuários de telefonia móvel. O conhecimento e os resultados obtidos com este projeto serão úteis para o desenvolvimento de protocolos de comunicação baseados em modelos de mobilidade humana em comunidades ribeirinhas para possibilitar o uso de redes DTN, a fim de contribuir para a disseminação de informações em áreas isoladas.

A fim de iniciar o estudo no contexto de reconhecimento de comportamento de usuários de telefonia celular, este trabalho apresenta resultados obtidos a partir da aplicação de técnicas de agrupamento em dados coletados de 147 usuários para fornecer indícios de grupos naturais de comportamento. Os resultados, e o protocolo experimental utilizado são descritos neste relatório.

2. Revisão Bibliográfica

Nesta seção serão descritas algumas informações que foram incorporadas através da revisão bibliográfica, tais como bases de dados e o processo de classificação através de aprendizagem de máquina.

A personalidade humana pode ser prevista com segurança a partir de registros de telefones celulares convencionais. Usando um conjunto de indicadores desconhecidos, que são informados por meio de psicologia, pode ser computada, a partir dos dados disponíveis para todas as operadoras, a capacidade de prever a personalidade dos usuários, com uma precisão média em traços de 42% melhor do que o acaso, atingindo até 61% de precisão em um problema de três classes. Dado o número crescente e rápido de assinaturas de telefone móvel e a disponibilidade de registros telefônicos para os pesquisadores, esses novos indicadores de personalidade abrem as portas para caminhos interessantes para futuras pesquisas em ciências sociais. Por exemplo, o padrão do uso de mídias sociais como o Facebook ou o Twitter pode ser usado para prever o comportamento dos usuários, ou informações sobre o uso de vários aplicativos para celulares das pessoas (por exemplo, YouTube, Internet, Calendário, Jogos, etc) ou rede social para fazer inferências sobre humor e traços de personalidade dos proprietários de telefone. Embora estas abordagens sejam interessantes, elas somente são possíveis se acesso a muitas informações sobre toda a rede ou sobre as pessoas for possível, para que seja instalado um aplicativo de rastreamento específico no telefone de uso social das pessoas. Estes constrangimentos prejudicam grandemente o uso de tais métodos de classificação para as investigações em larga escala.

Uma série de novos indicadores foi desenvolvida para permitir a previsão da personalidade dos usuários. Para criar uma lista de indicadores, foram examinadas teorias e pesquisas em psicologia da personalidade e, mais especificamente, a literatura chamada "*Five Factor Model*", ou o modelo de cinco fatores de personalidade, o paradigma dominante na pesquisa da personalidade. O modelo de cinco fatores é uma organização hierárquica de traços de personalidade em termos de cinco dimensões básicas: Extroversão (isto é, a tendência de buscar estimulação na companhia de outros, ser extrovertido e energético), Socialização (isto é, a tendência a ser carinhoso, compassivo e cooperativa), consciência (ou seja, a tendência de mostrar a auto-disciplina, ser organizado e objetivo para a realização), neuroticismo (isto é, a tendência de experimentar emoções desagradáveis facilmente) e abertura (ou seja, a tendência de ser intelectualmente curioso, criativo e aberto a sentimentos).

A partir dessa revisão da literatura, foram gerados novos indicadores que podem ser facilmente computados a partir de registros de operadoras e que acreditava-se que iriam significativamente ser responsáveis por eventuais diferenças de personalidade. Estes indicadores caem sob cinco grandes categorias: uso básico do telefone (por exemplo, número de chamadas, número de textos ou SMS), o

comportamento dos usuários ativos (por exemplo, número de chamadas iniciadas, o tempo para responder a um SMS), localização (raio de giro, o número de lugares a partir do qual foram feitas as chamadas), regularidade (por exemplo, chamadas temporais de rotina, chamada e SMS inter-tempo) e diversidade (chamada entropia, número de interações dividido pelo número de contatos).

Outra investigação da característica mais importante para prever cada traço revelou associações interessantes. Indicadores relacionados com a mobilidade dos usuários (ou seja, a distância percorrida e a entropia de lugares) foram úteis para prever Extroversão e Socialização. Estes resultados estão em linha com pesquisas anteriores. Além disso, essas características se relacionam com os diferentes aspectos da diversidade da rede social de um usuário: os extrovertidos tendem a procurar mais amigos do que os introvertidos; os indivíduos agradáveis tendem a ser selecionados mais como amigos por outras pessoas; indivíduos conscientes tendem a gostar de organização, precisão e pontualidade. Verificou-se também que o melhor preditor de Consciência foi a variação do tempo entre chamadas telefônicas. Por último, o mais forte preditor de abertura foi o tempo médio entre as interações.

A literatura também mostra que extroversão e neuroticismo são as duas dimensões da personalidade que são mais diretamente associadas com a emoção. Em particular, a extroversão é um forte preditor de emoções positivas e neuroticismo é um forte preditor de emoção negativa.

Diante desse contexto, é possível destacar os seguintes atributos como informações relevantes para representar os dados utilizados no processo de reconhecimento de comportamento de usuário de telefonia celular: quantidade de chamadas recebidas/realizadas, quantidade de mensagens recebidas/realizadas, tempo para responder uma mensagem recebida, locais a partir dos quais são feitas as chamadas, dentre outros. Por outro lado, muitos desses atributos só podem ser obtidos por meio de sensores incorporados aos telefones celulares dos usuários, especialmente *smartphones*. Devido à dificuldade de obtenção desses equipamentos, neste trabalho, optamos por obter informações de usuários de telefones celulares por meio do uso de questionários, conforme descrição apresentada na próxima seção.

3. Metodologia

Visando alcançar as metas estabelecidas por este projeto, os seguintes passos foram realizados:

4.1 Pesquisa Bibliográfica

A primeira etapa da pesquisa bibliográfica envolveu a busca por bases públicas para iniciar os experimentos práticos. Os resultados obtidos indicam que existem poucas bases de dados com ênfase na classificação de comportamento de usuários de telefonia celular disponíveis publicamente. Diante dessa dificuldade, a próxima etapa do trabalho envolveu o levantamento de atributos utilizados para representar dados de comportamento de usuários de telefone celular nos vários artigos estudados e classes existentes no problema, mais especificamente considerando usuários de *smartphones*. Essas informações foram utilizadas na geração de uma nova base de dados, a qual foi usada nos experimentos práticos do trabalho. Além disso, foram feitas pesquisas bibliográficas no contexto de técnicas e software necessários para rodar os experimentos.

4.2 Obtenção da Base

A coleta das informações de usuários de *smartphones* foi feita por meio de um questionário *online*. O questionário, composto de 20 questões (definidas com base na revisão da literatura), foi divulgado na lista de alunos dos cursos de graduação em Ciência da Computação e Sistemas de Informação da UFAM, assim como entre alunos de um curso de nivelamento do Programa de Pós-Graduação em Informática da UFAM realizado em Rio Branco-Acre.

Um grupo de 150 pessoas respondeu o questionário, sendo que o processo de coleta dos dados foi anonimizado. Porém, a organização da base foi difícil porque alguns usuários forneceram informações falsas. A fim de reduzir esses problemas, apenas 147 questionários foram considerados, dentre os 150 preenchidos. Portanto, a base de dados investigada nos experimentos é composta por 147 instâncias, as quais são representadas por 20 atributos (20 questões do questionário). Conforme mencionado anteriormente, a literatura indica a existência de 05 classes de comportamento de usuário de telefonia móvel, porém, há interseção entre essas classes. Diante desse fato, nós reduzimos o problema para identificar apenas dois grupos de comportamento: comportamento indicado ou comportamento não indicado. Esses dois grupos de comportamento estão relacionados com o objetivo geral do projeto: encontrar uma rota entre dois dispositivos desconectados por meio da descoberta de hábitos humanos e padrões em sua conectividade.

4.3 Métodos de Agrupamento

Dois métodos de agrupamento foram utilizados para o teste e avaliação: *Expectation Maximization* (EM) e *KMeans*. Os métodos de agrupamento foram utilizados porque não havia dados rotulados.

5. Resultados e Discussões

O questionário utilizado na coleta de dados para geração da base de dados é composto por questões como: quantidade de chamadas recebidas/realizadas, quantidade de mensagens recebidas/realizadas, tempo para responder uma mensagem recebida, dentre outras. A Tabela 1 mostra as perguntas utilizadas no questionário *online* (https://docs.google.com/a/icompu.ufam.edu.br/forms/d/1Mb6sBV1j-Q13b0quyLa4_MD8QRFeZ-OzKaQOjI6Bjw0/viewform). A maioria das perguntas eram questões de múltipla escolha. A Tabela 2 mostra algumas perguntas com suas respectivas opções.

Tabela 1. Perguntas utilizadas no questionário *online* gerado para coletar dados da base de dados.

No	Pergunta
1	Qual é o seu gênero?
2	Você possui um smartphone?
3	Em média, quantas mensagens SMS você envia por dia?
4	Em média, quantas mensagens SMS você recebe por dia?
5	Para que pessoas você costuma enviar mensagens?
6	Quais são as redes sociais que você usa frequentemente?
7	Quantos amigos você tem na rede social que você usa com mais frequência?
8	Com quem você costuma conversar diariamente na rede social que você usa com mais frequência?
9	Aproximadamente, quantas chamadas você faz diariamente?
10	Aproximadamente, quantas chamadas você recebe diariamente?
11	Para quem você costuma ligar?
12	Quem liga para você regularmente ?
13	Em que horário você costuma fazer/receber ligações?
14	Aproximadamente, quanto tempo você permanece conectado diariamente à rede social que usa com mais frequência?
15	Após ler um SMS, quanto tempo, em média, você leva para enviar a resposta?
16	Após ler uma mensagem enviada por um amigo na rede social que você usa com mais frequência, quanto tempo você leva para respondê-la?
17	Após ler uma mensagem enviada por um conhecido na rede social que você usa com mais frequência, quanto tempo você leva para respondê-la?
18	Você costuma enviar solicitações de amizade para pessoas nas redes sociais?
19	Dentre os seus contatos na rede social que você mais usa, quantos contatos (percentualmente) foram obtidos a partir de solicitações de amizade feita por você?
20	Você participa de quantos grupos na rede social que mais usa?

Tabela 2. Exemplos de perguntas com múltipla escolha utilizadas no questionário *online*.

Aproximadamente, quantas chamadas você recebe diariamente? *

Menos de 5

5-10

10-20

20+

Para quem você costuma ligar? *

Namorado(a)/ Cônjuge

Pessoas do ambiente de trabalho

Amigos

Familiares

Other:

Após a coleta dos dados, técnicas de aprendizagem de máquina para agrupamento de dados foram utilizadas. Conforme mencionado anteriormente, apenas dois grupos de comportamento foram investigados: comportamento indicado ou comportamento não indicado. Esses dois grupos de comportamento estão relacionados com o objetivo geral do projeto: encontrar uma rota entre dois dispositivos desconectados por meio da descoberta de hábitos humanos e padrões em sua conectividade.

A análise de *clusters* (ou agrupamento) é a classificação de objetos em diferentes grupos, ou mais precisamente, o particionamento de um conjunto em subconjuntos (*clusters*) de dados, de modo que os dados de cada subconjunto (idealmente) compartilhem algum traço comum - muitas vezes de proximidade de acordo com alguma medida de distância definida. A tarefa computacional de classificar o conjunto de dados em k *clusters* é muitas vezes referida como *k-clustering*.

Dois diferentes técnicas de agrupamento foram investigadas: *k-means* e *Expectation Maximization* (EM). O primeiro método é considerado o algoritmo clássico de agrupamento baseado na otimização de função custo, enquanto o segundo é um método de agrupamento baseado probabilístico.

EM atribui uma distribuição de probabilidade para cada instância que indica a probabilidade da referida amostra pertencer a cada um dos grupos. O número de grupos pode ser definido usando validação cruzada, ou o pode ser feito *a priori* pelo usuário, conforme foi feito em nossos experimentos. *K-means* é conhecido por ser o algoritmo mais simples e o mais popular para agrupamento. Com base em algum critério de similaridade/dissimilaridade, *k-means* analisa se as instâncias podem ser agrupadas sem ter qualquer conhecimento prévio sobre elas. O critério usado aqui é a distância entre as instâncias. A aglomeração pode ser feita usando as seguintes medidas de distância: a Distância Euclidiana, Distância Manhattan, etc. Distância Euclidiana é medida como uma distância de linha reta entre dois pontos no espaço euclidiano. A Distância Manhattan é a distância entre dois pontos medidos ao longo de eixos em ângulos retos.

Os algoritmos não foram implementados. Foi utilizada a ferramenta Weka (*Waikato Environment for Knowledge Analysis*). Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamados a partir de seu próprio código Java. Weka contém ferramentas para pré-processamento de dados, classificação, regressão, *clustering* (agrupamento), regras de associação e visualização. Além disso, é um ferramenta muito adequada para o desenvolvimento de novos sistemas de aprendizado de máquina.

As figuras a seguir mostram como os dois métodos de agrupamentos investigados foram usados para agrupar as instâncias em duas classes.

```

=== Run information ===
Scheme:      weka.clusterers.EM -I 100 -N 2 -X 10 -max -1 -ll-cv 1.0E-6
-ll-iter 1.0E-6 -M 1.0E-6 -num-slots 1 -S 100
Relation:    Human_Behavior
Instances:    147
Attributes:   20
              Q1
              Q2
              Q3
              Q4
              Q5
              Q6
              Q7
              Q8
              Q9
              Q10
              Q11
              Q12
              Q13
              Q14
              Q15
              Q16
              Q17
              Q18
              Q19
              Q20
Test mode:    evaluate on training data

```

=== Clustering model (full training set) ===

EM

==

Number of clusters: 2
Number of iterations performed: 2

Attribute	Cluster	
	0 (0.65)	1 (0.35)
=====		
Q1		
mean	1.1907	1.2478
std. dev.	0.3928	0.4317
Q2		
mean	1.094	1.0974
std. dev.	0.2919	0.2966
Q3		
mean		1
std. dev.	0.4776	0.1411
Q4		
mean	1.1784	1.9424
std. dev.	0.3829	0.233
Q5		
mean	7.8618	9.7311
std. dev.	4.5528	4.591
Q6		
mean	7.0134	6.6874
std. dev.	4.284	4.3817
Q7		
mean	349.3469	291.5651
std. dev.	374.9755	371.6542
Q8		
mean	3.7256	3.1365
std. dev.	4.6137	3.6933
Q9		
mean	1.307	1.8423
std. dev.	0.5317	1.0397
Q10		
mean	1.2852	1.8053
std. dev.	0.5742	0.9822
Q11		
mean	8.6358	8.3289
std. dev.	4.3522	4.2882
Q12		
mean	8.9609	9.0329
std. dev.	4.3498	4.2209
Q13		
mean	6.9788	7.5189
std. dev.	3.6014	4.0681
Q14		
mean	2.6031	2.6673
std. dev.	1.2836	1.0901
Q15		
mean	1.8258	1.3765
std. dev.	1.1111	1.1111

```

    std. dev.    1.3001    0.6106

Q16
  mean         2.0548    1.6888
  std. dev.    1.1555    0.8188

Q17
  mean         2.5478    1.9614
  std. dev.    1.3594    0.9993

Q18
  mean         1.5695    1.5557
  std. dev.    0.4951    0.4969

Q19
  mean         2.1245    2.0035
  std. dev.    0.9741    0.9843

Q20
  mean         1.79      1.826
  std. dev.    0.7663    0.7266

Time taken to build model (full training data) : 0.2 seconds
=== Model and evaluation on training set ===

Clustered Instances

0          97 ( 66%)
1          50 ( 34%)

Log likelihood: -37.34157

```

Fig1: Resultados obtidos via Weka utilizando o algoritmo *Expectation Maximization*

```

=== Run information ===
Scheme:      weka.clusterers.SimpleKMeans -N 2 -A
"weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation:    Human_Behavior
Instances:   147
Attributes:  20
             Q1
             Q2
             Q3
             Q4
             Q5
             Q6
             Q7
             Q8
             Q9
             Q10
             Q11
             Q12
             Q13
             Q14
             Q15
             Q16
             Q17
             Q18
             Q19
             Q20
Test mode:   evaluate on training data

=== Clustering model (full training set) ===
kMeans
=====

Number of iterations: 5
Within cluster sum of squared errors: 249.2450068005341
Missing values globally replaced with mean/mode

```

```

Cluster centroids:
Attribute      Full Data      Cluster#
              (147)         0           1
              (147)         (68)        (79)
=====
Q1             1.2109         1.25        1.1772
Q2             1.0952         1.1176      1.0759
Q3             1.3469         1.75        1
Q4             1.449          1.9559      1.0127
Q5             8.5238         9.5         7.6835
Q6             6.898          6.6029      7.1519
Q7             328.8844       297.5441    355.8608
Q8             3.517          3.0588      3.9114
Q9             1.4966         1.7794      1.2532
Q10            1.4694         1.7059      1.2658
Q11            8.5306         8.5441      8.519
Q12            8.9864         9.4559      8.5823
Q13            7.1701         7.4706      6.9114
Q14            2.6259         2.6471      2.6076
Q15            1.6667         1.5588      1.7595
Q16            1.9252         1.7794      2.0506
Q17            2.3401         2.0588      2.5823
Q18            1.5646         1.5294      1.5949
Q19            2.0816         2.0294      2.1266
Q20            1.8027         1.7647      1.8354

```

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

```

0          68 ( 46%)
1          79 ( 54%)

```

Fig2: Resultados obtidos via Weka utilizando o algoritmo Kmeans.

A partir dos resultados exibidos nas figuras 1 e 2, foi feita uma análise dos dados componentes de cada grupo, além de ter sido feito um estudo sobre a interseção entre os grupos gerados por cada algoritmo. As seguintes observações podem ser feitas. EM mostrou que 66% dos usuários (97 pessoas dentre 147), são bons candidatos para a retransmissão de informação, enquanto 34% (50 usuários), são candidatos não indicados. Por outro lado, *k-means* mostrou que apenas 54% (79 usuários), são bons candidatos, enquanto 46% (68 usuários) não são. Comparando os dois métodos, a diferença é de 18 usuários. Essa diferença pode ter sido causada por informações erradas fornecidas pelos usuários. Para superar esses problemas, assim como tentar estudar mais profundamente os atributos relevantes dos dois grupos de usuários, a continuação desta pesquisa envolverá o uso de sensores instalados nos telefones celulares, de modo que, no final, os resultados obtidos sejam comparados aos resultados obtidos com dados extraídos dos formulários e sejam utilizados para futuras pesquisas.

6. Conclusões

Este projeto investigou o uso de técnicas de agrupamento de dados aplicadas ao problema de identificação de comportamento de usuários de *smartphones*. Os resultados obtidos não podem ser considerados precisos, devido ao fato da coleta de dados não ter sido muito efetiva. Porém, esta pesquisa forma uma base para futuras pesquisas no contexto de estudos do comportamento humano com escopo mais abrangente. A técnica mais eficiente para obter resultados precisos é o uso de sensores. Quando embutidos em *smartphones*, estes sensores podem determinar a mobilidade humana, e prever seu próximo destino com base em seu histórico de chamadas. Além disso, *smartphones* com sensores embutidos podem gravar as pistas contextuais e atividades dos usuários, incluindo a localização, uso do aplicativo e comportamento de chamadas. Essas informações podem ser consideradas como a entrada e a saída de um quadro de predição, no qual os valores futuros de algumas variáveis (por exemplo, próximo lugar) são previstas com base no contexto atual (local atual, tempo).

Referências Bibliográficas

- Trinh M., Tri D., Daniel G., *Where and What: Using Smartphones to Predict Next Locations and Applications in Daily Life*, 2013.
- Trinh M., Tri D., Daniel G., *Contextual Conditional Models for Smartphone-based Human Mobility Prediction*, UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.
- Yves-Alexandre M., Jordi Q., Florent R., Alex P., *Predicting people personality using novel mobile phone-based metrics*, 2013.
- Davide A., Alessandro G., Luca O., Xavier P., and Jorge L., *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine*, DMYTRO, K., CHIARA, B.C. AND ANDREA, P., *Human Mobility Models for Opportunistic Networks*. IEEE Communications Magazine. 2011. p.157-165.
- EAGLE, N. AND A. PENTLAND, *Reality Mining: Sensing Complex Social Systems*. Personal and Ubiquitous Computing, 2006. 10:p.255-268.
- EAGLE, N., A. PENTLAND, AND D. LAZER, *From the cover: Inferring Friendship Network Structure by using Mobile Phone Data*. Proceedings of The National Academy of Sciences, 2009. 106(36): p. 15374-15278.
- GONZALEZ, M.C., A. HIDALGO, AND A.-L. BARABASI, *Understanding Individual Human Mobility Patterns*. Nature, 2008.
- JAIN, A. K., DUN, R. P.W., MAO, J. *Statistical Pattern Recognition: A Review*. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.22 n.1, 2000.