

Revisão Literária sobre Previsão de Trajetórias

Francisco Dantas Nobre Neto^[1], Cláudio de Souza Baptista^[2] Cláudio Elizio Calazans Campelo^[3], Niara Fernandes Barbosa Formiga Dantas^[4]

[1] dantas.nobre@ifpb.edu.br, Instituto Federal de Educação, Ciências e Tecnologia da Paraíba – Campus Campina Grande [2] baptista@dsc.ufcg.edu.br, Universidade Federal de Campina Grande – Campus Campina Grande. [3] campelo@dsc.ufcg.edu.br, Universidade Federal de Campina Grande – Campus Campina Grande. [4] niara.formiga@ifpb.edu.br, Instituto Federal de Educação, Ciências e Tecnologia da Paraíba – Campus Picuí

RESUMO

Em virtude da difusão de smartphones equipados com recursos GPS, a possibilidade real de rastrear objetos tem impulsionado estudos em diferentes ramos de pesquisa, incluindo os da área de previsão de trajetórias. Entre várias definições de previsão de trajetória, uma definição coerente é a de prever um destino e, em alguns casos, a rota até o destino, dada a ciência de um conjunto de informações, como o local da partida e o trecho já percorrido. Com relação à utilidade, um sistema de previsão de trajetórias pode prever o restante de trajetória em curso e alertar o usuário a respeito de um congestionamento, sugerindo um desvio. Este artigo realiza uma revisão literária da temática de previsão de trajetórias, apresentando os conceitos básicos e os desafios importantes que precisam ser considerados.

Palavras-chave: Trajetória. Previsão de trajetória. Reconhecimento de padrões. Revisão literária.

ABSTRACT

Due to the diffusion of smartphones equipped with GPS devices, the possibility of tracking objects has fostered studies in many fields, including trajectory prediction. Among many possible definitions for trajectory prediction, an appropriate one is the capability of predicting a destination, and, in some cases, the route to the destination, given a set of information as input, such as departure and the route already taken. Regarding the usefulness, a trajectory prediction system can predict the remaining trajectory to the destination, warning the user to take a deviation, in case of traffic jam. This paper reviews the literature on trajectory prediction, providing the basic concepts, as well as the challenges that need to be addressed.

Keywords: Trajectory. Trajectory prediction. Pattern recognition. Literature review.

1 Introdução

Em virtude da larga difusão de *smartphones* capazes de capturar seus próprios posicionamentos, por meio do auxílio de dispositivos GPS embutidos, a possibilidade de rastrear locais que os usuários visitam é muito mais viável atualmente do que antigamente.

Um sistema para rastrear localizações visitadas pelos usuários tem variadas aplicações práticas e importantes, tais como: obter um perfil de deslocamentos do usuário, com a possibilidade de deixá-lo ciente a respeito de uma rota congestionada; e ou lembrá-lo de visitar algum local, com base em seu perfil de deslocamento. Em ambos os casos, o rastreamento de localizações permite identificar quais são as rotas que o usuário costuma percorrer, quais são os lugares visitados (e, até mesmo, quais as características destes lugares) e quanto tempo é demandado para a realização de uma trajetória.

De posse de informações a respeito das trajetórias realizadas, uma aplicação prática está na possibilidade de previsão de trajetórias. Prever uma trajetória significa prever um destino e ou uma rota que um usuário deverá alcançar, dado um percurso realizado até o momento.

Com relação à previsão de trajetórias, é recomendado que um conjunto de rotas, previamente realizadas, seja utilizado como entrada para elaboração do perfil de deslocamentos de um usuário. Este perfil de deslocamento é importante para sugerir, por exemplo, que, quando o usuário origina um deslocamento do *lugar X*, ele provavelmente alcança o *lugar Y*, percorrendo o *caminho P*.

O propósito deste artigo é apresentar os trabalhos mais importantes na temática de previsão de trajetórias, informar quais são as abordagens de previsão mais utilizadas, além de apresentar quais são os desafios e oportunidades relacionados a este tema.

2 Conceitos básicos

Esta seção apresenta os conceitos básicos que servirão para melhor entendimento do artigo. Serão apresentadas as definições sobre trajetória, segmentos e sobre o funcionamento de um modelo de previsão.

Definição 1: Trajetória – Uma trajetória representa a movimentação de um objeto entre dois

lugares visitados e pode ser representada pela tupla $T = \langle o, d, S \rangle$, em que:

- o é uma região geográfica que representa a origem da trajetória;
- d é uma região geográfica que representa o destino da trajetória;
- S representa a rota que une a origem e o destino da trajetória.

A rota S pode ser representada de três maneiras principais: (1) como um conjunto de coordenadas espaciais (latitude, longitude); (2) como um conjunto de segmentos; ou (3) como um conjunto de células (ou grades), em que o mapa é segmentado em células de tamanho $n \times m$. Para todos os casos, o conjunto deve estar ordenado de forma temporal, para que seja possível identificar o progresso da trajetória.

Com relação à diferença entre as representações, ao se usar um conjunto de coordenadas espaciais para representar um deslocamento, obtém-se maior precisão espacial a respeito do caminho percorrido, porém, o tamanho do conjunto de coordenadas espaciais pode ser extenso. Ao se usar um conjunto de segmentos ou um conjunto de células, estes podem substituir várias coordenadas espaciais por um único segmento, embora se diminua a precisão espacial.

A Figura 1 ilustra a diferença de formas de representação de uma rota. Na Figura 1 (A), a rota é representada como um conjunto de coordenadas espaciais. Na Figura 1 (B), a rota é representada por um conjunto de segmentos. Já na Figura 1 (C), a mesma rota da Figura 1 (A) é representada por um conjunto de células. Entre as três formas, a de representação por coordenadas espaciais é a mais precisa, e a de representação por um conjunto de células é a que mais perde precisão.

Definição 2: Segmento – Um segmento é representado pela tupla $\langle id, P_{início}, P_{fim} \rangle$, em que id é um identificador único do segmento, $P_{início}$ é a coordenada de início do segmento e P_{fim} é a coordenada final do segmento.

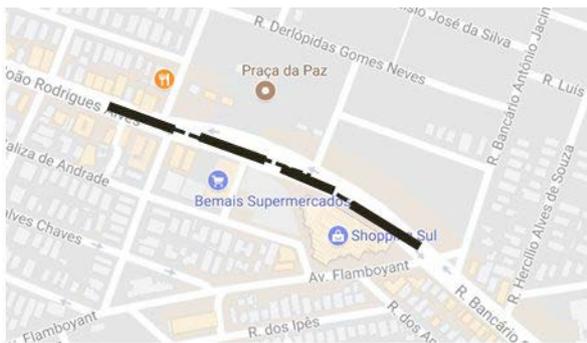
Um preditor que manipula um conjunto de segmentos para representar uma trajetória pode resultar em um ganho de performance, quando comparado a um preditor que manipula um conjunto de coordenadas. Isto ocorre pelo fato de que um único segmento é capaz de representar várias coordenadas espaciais, diminuindo, assim, a quantidade de informação a ser

manipulada. Para que seja possível associar coordenadas espaciais a segmentos, deve-se utilizar alguma técnica de *map-matching* (QUDDUS; NOLAND, 2006). Uma técnica de *map-matching* é responsável por associar coordenadas espaciais a segmentos.

Figura 1 – Formas de representação de uma rota (em (A), a rota é representada por um conjunto de coordenadas espaciais; em (B), é utilizado um conjunto de segmentos; e em (C), um conjunto de célula



(A) Rota representada por um conjunto de 10 coordenadas espaciais.



(B) Rota representada por um conjunto de quatro segmentos.



(C) Rota representada por um conjunto de seis células.

Fonte: Autoria própria.

Definição 3: Previsão de trajetórias – Para a previsão de trajetórias, um modelo de previsão recebe a origem do deslocamento, o conjunto de dados RP que representa a Rota Parcial percorrida até o momento e alguma informação contextual (por exemplo, dia da semana da trajetória). Como saída, o modelo informa o destino e pode informar, também, a rota restante a ser percorrida para alcançar o destino.

Quanto à Definição 3, esta pode ter pequenas variações, dependendo do modelo de previsão. Por exemplo, há modelos que recebem como entrada a origem da trajetória e a rota parcial percorrida, resultando, apenas, no destino da trajetória. Há modelos, ainda, que utilizam apenas a origem do deslocamento e algumas informações contextuais, como o dia da semana e o horário da partida, para prever a região que o usuário deverá alcançar. Outra classe de modelos prevê, além do destino, a rota restante que deverá ser percorrida até alcançar o destino.

Quanto à previsão de trajetórias, esta pode ser ainda agrupada em (1) previsão personalizada e (2) previsão coletiva. Na previsão personalizada, são utilizadas informações de deslocamento, apenas sobre o perfil do usuário, para prever uma trajetória, sem sofrer interferência de deslocamentos de outros usuários. Na previsão coletiva, o preditor é capaz de acessar todas as trajetórias já realizadas, sem filtrar pelo usuário, e prever uma trajetória de um usuário com base em uma trajetória realizada por outros usuários.

Outra questão importante a respeito de previsão de trajetórias é o tipo do deslocamento. Há pessoas que utilizam seu veículo de forma privada, cujas rotas são realizadas para atender a suas necessidades. Há, também, pessoas que utilizam o carro para o trabalho, como é o caso dos taxistas, cujas trajetórias são determinadas conforme a necessidade do cliente. Identificar o escopo de previsão com que se quer trabalhar é importante, uma vez que um preditor dificilmente funcionará para ambos os casos.

3 Revisão da literatura

Esta seção apresenta os trabalhos investigados na temática de previsão de trajetórias. Os trabalhos foram agrupados em dois grupos principais: um que trata de modelos de previsão que utilizam informações espaciais e temporais a respeito da trajetória, e outro que, além dessas informações, também utiliza conteúdo semântico, como, por exemplo, o papel que um lugar representa para o usuário, ou seja, é

a identificação da representatividade do lugar para o usuário, se é sua casa, seu trabalho ou seu lazer. Antes de investigar os trabalhos, porém, foi especificado um planejamento sobre os itens a serem analisados no conteúdo dos artigos.

3.1 Planejamento da revisão

Para a investigação dos trabalhos relacionados à temática de previsão de trajetórias, alguns questionamentos foram definidos para melhor entendimento a respeito do modelo de previsão proposto. Os questionamentos são apresentados a seguir:

Qual o objetivo da pesquisa? Qual a principal contribuição do artigo? Previsão em tempo real (online)? O modelo de previsão proposto é coletivo ou individual? Qual a granularidade da composição de uma rota (células, segmentos, pontos GPS)? Qual a técnica principal utilizada pelo modelo preditor? Preveem-se lugares não visitados anteriormente? Valida-se o modelo com dados reais? Se dados reais são utilizados, a base de rotas é própria ou pública? Se pública, qual? Utilizam-se somente dados históricos para previsão futura? Utilizam-se informações contextuais? Utilizam-se informações de tipo de lugar? Quais são as métricas utilizadas para avaliação (precisão, cobertura e medida-F)? Há validação cruzada? O modelo é eficaz e eficiente?

3.2 Previsão de trajetórias sem semântica

Para previsão de trajetórias, o modelo criado por Simmons et al. (2006) usa a técnica de Cadeias de Markov Oculta (HMM) e informação contextual (dia da semana, horário e velocidade do veículo), em um corpus formado por 46 rotas na região de Michigan, nos Estados Unidos. A taxa de acerto do modelo deles é de 98%, embora seja necessário fazer uma observação: apenas 5% das transições entre os segmentos permitem seguir por mais de um caminho. Isto é, 95% dos segmentos têm apenas um único caminho à frente, o que pode facilitar a tarefa de previsão do próximo segmento. Para o caso dos 5% das transições que permitem seguir por mais de um segmento, a taxa de acerto do algoritmo fica entre 70% e 80%. O modelo de previsão desenvolvido por Krumm (2008) é capaz de prever apenas os próximos segmentos (até os 10 seguintes), mas não prevê o destino de um trajeto. Seu modelo utiliza Cadeias de Markov para previsão, analisa os 10 últimos segmentos percorridos e, com uma taxa de acerto de 90%,

consegue prever o próximo segmento. Quando é feita a previsão para os próximos 10 segmentos, a taxa de acurácia diminui para 50%.

O modelo de previsão proposto por Froehlich e Krumm (2008) usa a técnica de Algoritmo de Casamento de Padrão (do inglês *Closest Match Algorithm*), baseado no algoritmo de distância de Hausdorff (HUTTENLOCHER; RUCKLIDGE, 1992), que tenta identificar a similaridade entre uma rota em curso (rota parcial) e uma rota realizada anteriormente. Se houver uma rota similar (interseção e proximidade entre as coordenadas geográficas do deslocamento), recupera-se a rota restante, baseada em uma rota já realizada. O modelo proposto por eles alcança uma precisão de 40%, quando 50% da rota é percorrida. Uma limitação para a aplicabilidade prática do trabalho deles, porém, é a ausência da técnica de *map-matching*. Por outro lado, Tiwiri, Arya e Chatuverdi (2013) apresentam uma metodologia similar àquela de Froehlich e Krumm (2008), no entanto, eles utilizam em seu modelo a técnica de *map-matching* para reduzir o tamanho dos dados. No trabalho publicado por eles, a taxa de acurácia foi semelhante à de Froehlich e Krumm (2008), porém, o quantitativo de dados manuseados foi reduzido e a eficiência do modelo foi melhor.

O algoritmo *Prediction by Partial Matching* (PPM) foi proposto para o problema de previsão de destinos no trabalho de Burbey e Martin (2008). O modelo de previsão proposto pelos autores considera informações contextuais referentes à chegada do usuário aos lugares que costuma visitar, a duração temporal em que o usuário permanece nestes lugares e o lugar atual onde ele se encontra. Com base nessas informações, é previsto o próximo destino do usuário, mas sem considerar os segmentos a serem percorridos. Uma vez que o preditor desenvolvido por Burbey e Martin (2008) considera apenas a origem do usuário, não há mudança no destino a ser previsto, mesmo quando o usuário segue por outro caminho.

Xue et al. (2013) desenvolveram um modelo que usa cadeias de Markov e inferência Bayesiana para previsão de rotas. A abordagem permite prever, inclusive, lugares que um usuário não visitou ainda, ou seja, superando a limitação de prever passeios apenas baseado em dados históricos. Isto é possível pelo fato de que uma rota realizada gera uma ou mais rotas novas (não realizadas), para contemplar regiões circunvizinhas à rota que foi percorrida. O corpus utilizado no trabalho deles foi de motoristas de taxi,

cujos deslocamentos são definidos pelos passageiros. Embora o modelo de previsão proposto por Xue et al. (2013) consiga prever lugares nunca visitados, ele faz uso apenas de informações geográficas e não considera o papel que um lugar representa para o usuário (como, por exemplo, casa ou trabalho).

Lee et al. (2016) desenvolveram um preditor que utiliza Casamento de Padrões para prever o próximo destino, dada uma origem. Uma trajetória é representada por uma tupla que contém informações sobre a localização da origem, localização do destino, o dia da semana e o tempo durante o qual o usuário permaneceu lá. Esta trajetória é chamada de *Spatiotemporal-periodic* (STP). Em seguida, o preditor segmenta uma trajetória em subtrajetórias, para poder identificar trechos esporádicos e rotineiros. Assim, uma trajetória será composta por um conjunto de subtrajetórias, que pode possuir lacunas. No trabalho desses autores, esse conjunto de subtrajetórias é chamado de GSTP. Com isso, mesmo que o usuário percorra uma rota esporádica, o algoritmo de Casamento de Padrões, por meio das trajetórias GSTP, associa o trajeto corrente com algum já realizado.

Herder Siehndel e Kawase (2014) apresentaram várias abordagens de previsão de destino, realizando uma comparação entre elas. Foram implementados preditores baseados nas técnicas:

- Top-N: em que o preditor tenta prever o destino baseado nos N lugares mais visitados;
- Last-N: em que o modelo tenta prever o destino baseado nos últimos N lugares que o usuário visitou;
- Hora: em que o modelo seleciona os N lugares de um mesmo horário em que a previsão de destino está sendo feita, e tenta prever conforme esse horário;
- Distância: em que o modelo seleciona os N lugares mais próximos da posição corrente do usuário para prever o destino;
- Cadeia Simples de Markov: este modelo utiliza uma cadeia simples de Markov, em que observa o estado (lugar) atual para prever o destino.

Para avaliação das abordagens implementadas, foram utilizadas as bases de deslocamentos MSR GPS *Privacy* (BRUSH; KRUMM; SCOTT, 2010), e a base do projeto do *GeoLife* (ZHENG et al., 2009).

Conforme os experimentos apresentados pelos autores, as cadeias de Markov obtiveram os melhores resultados, alcançando uma precisão de 62.9%.

Trasarti et al. (2015) propuseram um preditor que mescla as informações individuais e coletivas dos usuários, para realizar previsão de destino. Assim, uma vez que são usadas informações acerca de deslocamentos de todos os usuários para predição, esse preditor é capaz de prever lugares ainda não visitados. É importante destacar que esse modelo manipula coordenadas geotemporais ao invés de realizar qualquer discretização (por exemplo, uso de *map-matching*), o que pode gerar uma grande quantidade de dados para manuseio. Os autores alegam que o uso de coordenadas geotemporais resultará em uma previsão de destino (para eles, o destino é uma coordenada geotemporal) espacialmente mais próxima do real. Para prever uma rota, eles utilizam um algoritmo que representa uma função de distância da rota que está em curso com as rotas históricas dos usuários. A rota histórica que obtiver o menor valor, resultante do cálculo desta função, será considerada para a previsão.

Liu, Liu e Li (2016) desenvolveram um sistema individual de recomendação que realiza a previsão dos N (top-N) destinos mais prováveis para onde o usuário irá se deslocar. A metodologia do preditor se baseia no método *Skip-gram*, em que ocorre a associação de uma palavra diferente para cada lugar diferente visitado pelo usuário, resultando em uma sentença. A sentença representa todos os lugares visitados em um trajeto, e ela é informada na etapa de treinamento do preditor. Para a sugestão dos destinos mais prováveis a serem visitados pelo usuário, é utilizada uma extensão da técnica *Weighted Approximately Ranked Pairwise* (WARP).

Técnicas de descoberta de conhecimento (KDD), como regras de associação, já foram usadas para criação de modelos preditores. Quando um veículo inicia um deslocamento, uma regra de associação é gerada para a rota em curso (conforme os segmentos percorridos). Em seguida, uma função de identificação de padrões é ativada, para verificar se o conjunto de segmentos percorrido, até o momento, está em uma árvore de segmentos gerada anteriormente. O modelo proposto por Morzy utiliza uma versão do algoritmo *Apriori* para geração das regras de associação. Tanaka et al. (2009) apresentam um procedimento de previsão baseado na via onde o usuário está (como, por exemplo, se rodovia ou via arterial). Além disso, o

processo considera também as informações contextuais de dia da semana, horário do deslocamento, número de passageiros, condições climáticas e peso da bagagem. Em Monreale et al. (2009), uma trajetória é representada por um conjunto de células, em que cada célula representa as localizações visitadas pelo usuário. Associada a cada localização está o horário em que o usuário a visitou. Para previsão de destino, o modelo utiliza regras de associação.

Figueiredo et al. (2016) propuseram um modelo denominado de *TriberFlow*, que utiliza uma Cadeia Semi Markoviana, juntamente com a técnica Passeio Aleatório (do inglês *Random Walk*). Na etapa de criação do modelo preditor, os trajetos são categorizados em dois grupos: (1) aqueles realizados frequentemente (também chamados de trajetos estacionários), e que, segundo os autores, são mais fáceis de prever; e (2) aqueles realizados esporadicamente (também chamados de trajetos transientes), isto é, trajetos que não são realizados rotineiramente. Com relação à informação contextual, o modelo faz uso da informação de data e hora de realização dos trajetos.

Para previsão de destino, também já foi usada a técnica de Árvore de Sufixo Probabilística, ou somente PST, do inglês *Probabilistic Suffix Trees* (RON, SINGER; TISHBY, 1996), que cria uma árvore de contexto para representar uma cadeia de Markov de ordem variável, semelhante ao que ocorre na técnica de compressão de dados PPM. No âmbito do trabalho de Lei, Li e Peng (2013), uma trajetória é representada por um conjunto de células, em que cada célula pode representar vários segmentos. O modelo de previsão de Lei, Li e Peng (2013) funciona com informações individuais e considera, como entrada, a informação geográfica referente ao deslocamento, além do horário do deslocamento. De posse dessas informações, o modelo de previsão constrói as árvores de sufixo. Rocha et al. (2016) desenvolveram um sistema para previsão de destinos chamado TPRED, que também utiliza um mecanismo de previsão de destino baseado no PST. De maneira semelhante ao que ocorre no trabalho de Lei, Li e Peng (2013), uma trajetória também é representada por um conjunto de células. O sistema descrito pelos autores extrai e utiliza, além da informação geográfica, as informações contextuais: dia da semana e hora. O modelo de previsão proposto por Rocha et al. (2016), além de prever para onde o objeto em movimento se desloca (isto é, destino do objeto), também prediz quando o objeto deve alcançar o destino.

3.3 Previsão de trajetórias com semântica

Em sistemas baseados em localização, a informação semântica representa a associação de dados espaciais sobre a localização de uma informação que agrega valor característico sobre o local (SPACCA-PIETRA et al., 2008). Isto é, busca-se associar dados espaciais, referentes às regiões visitadas, ao papel que determinada região possui para um usuário (se casa ou trabalho, por exemplo). Com isso, o uso da semântica busca superar a limitação de identificar apenas as regiões visitadas, mas entender a razão pela qual o usuário visitou tal região, e realizou determinado deslocamento (se de casa para o trabalho).

O trabalho de Ying, Lee e Weng está entre os pioneiros na utilização de semântica para melhoria da previsão de destinos. O corpus de deslocamento utilizado para testes foi criado por captura de dispositivos GPS e por sinais das torres de celulares. Foi criada uma base de anotações semânticas, conhecida como *Geographic Semantic Information Database* (GSID), que contém informações oriundas do *Google Maps*. O modelo de previsão apresentado se divide em dois módulos principais: um *offline*, responsável por anotar semanticamente as regiões geográficas de paradas; e outro *online*, que prevê em tempo real o destino que um usuário deve alcançar. A abordagem obteve uma taxa de acurácia entre 53% e 68% na previsão do destino. Ying et al. (2014) apresentaram um aperfeiçoamento, em que o novo modelo é capaz de sugerir itens a serem comprados, caso o usuário esteja em um estabelecimento.

Lung, Chung e Dai (2014) desenvolveram um modelo para previsão de destinos e descoberta do meio de transporte utilizado. Seu modelo usa HMM para a tarefa de previsão e consulta à biblioteca do *Google Maps*, para enriquecer semanticamente uma trajetória. Utilizando dados reais, o modelo de previsão obteve uma taxa de acurácia de 68.3% na previsão do próximo lugar a ser visitado. Uma lacuna identificada no trabalho publicado por Lung, Chung e Dai (2014), no entanto, é a dificuldade de identificar a base de dados que foi utilizada nos testes (não estão claros, por exemplo, a duração da coleta, nem o meio de transporte utilizado). Além disso, não foi apresentado um detalhamento maior sobre o procedimento de como a etapa de previsão é realizada. Embora Lung, Chung e Dai (2014) apresente uma abordagem que possibilite a previsão de lugares que o usuário nunca visitou antes, este detalhamento fica vago. O que fica claro, apenas, é que eles usam semântica

para tipos de lugares e realizam detecção do meio de transporte utilizado.

Huang, Ying e Tseng (2012) desenvolveram um modelo capaz de observar quatro aspectos contextuais referentes à localização onde um usuário está, como informação temporal, espacial, comportamental e ambiental. Este modelo utiliza variáveis contextuais e semânticas para realizar previsão do próximo tipo de lugar onde um usuário vai estar e foi elaborado para lidar com os dados disponíveis pelo desafio da Nokia, conhecido como *Mobile Data Challenge* (MDC) (LAURILA et al., 2012). O modelo de previsão desenvolvido por Huang, Ying e Tseng (2012) utilizou um extrator de características dos dados. Depois, foi utilizada a análise *Chi-quadrado* para identificar as características que possuem maior influência na taxa de acurácia (mais representativas). Para os testes, foram utilizados os algoritmos de classificação SMO (PLATT, 1998), J48 (QUILAN, 1993) e SimpleLogistic (PENG; LEE; INGERSOLL, 2002). Ocorreram testes tanto com os algoritmos de forma isolada como também com a combinação entre eles. A melhor taxa de acurácia foi obtida combinando os algoritmos SMO e SimpleLogistic, em que foi obtido 65,77% de acurácia. Zhu et al. (2013) também desenvolveram um modelo para o desafio MDC, destinado para a inferência do tipo de lugar onde um usuário está. Para elaboração desse modelo de inferência, foram usados os algoritmos de Lógica de Regressão (LR), *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995), *Random Forests* (RF) (BREIMAN, 2001) e *Gradient Boost Trees* (GBT) (NATEKIN; KNOLL, 2013).. A melhor taxa de acurácia obtida foi com o algoritmo GBT, com 75,1% de acerto.

O modelo de previsão desenvolvido por Nademgeba, Taleb e Hafid (2012), conhecido como *Destination Prediction Model* (DPM), utiliza informações contextuais para aperfeiçoamento da previsão, incluindo o papel que um lugar representa para o usuário. As informações contextuais, no entanto, devem ser reportadas pelo usuário por meio de um questionário, tarefa esta que não é prática de uso rotineiro em modelos preditivos. Para a previsão, este modelo utiliza um algoritmo que representa uma função de distância, para obter a trajetória histórica mais próxima da trajetória atual. Além disto, dada a posição atual, o modelo realiza uma análise das informações contextuais do usuário (como horário e dia da semana) e observa os últimos lugares visitados pelo usuário (inclusive, o contexto destes lugares).

Nobre Neto, Baptista e Campelo (2016) propuseram um modelo de previsão capaz de prever o destino e a rota até o destino. Seu modelo de previsão utiliza a técnica de compressão de dados PPM, como núcleo para previsão, e contempla informações espaço-temporais e semânticas. Eles utilizam informação semântica na técnica de previsão e a adequam para permitir prever, inclusive, destinos não visitados.

3.4 Análise dos trabalhos relacionados

Nos trabalhos descritos na Subseção 3.2, os modelos de previsão propostos realizam a previsão de destino e ou rota, considerando apenas informações espaciais e temporais, além de algum dado contextual, como dia da semana e horário da partida de um deslocamento. Outro aspecto a ser observado é que, na maior parte dos trabalhos, a trajetória é representada por um conjunto de células ao invés de um conjunto de segmentos. Proceder desta maneira diminui o quantitativo de dados a ser manipulado por um modelo, porém diminui a precisão espacial da previsão de rotas e de destino, especialmente quando o tamanho da célula é muito grande.

Nos trabalhos apresentados na Subseção 3.3, os modelos de previsão consideram o papel que um lugar representa para o usuário apenas como uma informação complementar para a previsão de destino, mas não a utiliza na tentativa de predizer um destino nunca visitado. O Quadro 1 apresenta os principais trabalhos relacionados. Na primeira coluna do Quadro 1 estão os trabalhos sob análise. A partir da segunda coluna, estão as características para comparação.

4 Abordagens de previsão de trajetórias utilizadas

Conforme literatura de previsão de trajetórias, é possível perceber que há modelos de previsão que preveem apenas o destino de uma rota e outros que informam o destino e o caminho até a rota.

Os modelos de previsão que preveem o destino são responsáveis por indicar apenas o destino que o usuário deverá alcançar. O percurso a ser realizado até o destino não é informado. Quanto a este tipo de abordagem, há uma desvantagem aparente: ao não prever a rota que o usuário irá seguir, não é possível informar a respeito de imprevistos no percurso, como, por exemplo, trechos interditados ou com trânsito congestionado.

Por outro lado, os modelos de previsão que indicam o destino e a rota até este destino têm a capacidade de informar sobre imprevistos ao longo da rota. Este tipo de abordagem, no entanto, deve ter cuidado ao realizar a previsão, pois o usuário pode seguir para o mesmo destino, percorrendo uma nova rota, e o modelo de previsão não conseguir compreender que a trajetória resultará em um destino já conhecido.

Comparando estes dois tipos de abordagem, os modelos de previsão que preveem apenas o destino de uma trajetória costuma ser mais performático, uma vez que não há a necessidade de processar o caminho até o destino, porém pode ser menos útil do que um modelo de previsão que prevê também a rota até o destino.

Com relação às informações manuseadas, os modelos podem utilizar apenas informações espaciais (deslocamentos já realizados), espaço-temporais (deslocamentos e dia da semana/horário da realização) e espaço-temporais com semântica (identificar

a importância dos lugares visitados ao usuário, por exemplo, se casa ou trabalho).

Modelos de previsão que utilizam informações espaciais tiveram, como base para uma previsão, trajetórias realizadas previamente, independentemente de quando esta foi realizada. Esta abordagem tem o problema de ter que utilizar todas as trajetórias para realizar uma previsão e pode representar alto custo computacional.

Os modelos de previsão que utilizam informações espaço-temporais conseguem identificar similaridades de uma trajetória a ser prevista com uma trajetória já feita, e, além disso, consegue identificar se o horário entre estas são similares. Esta abordagem é útil para diminuir a quantidade de trajetórias a ser utilizada para comparação. Por exemplo, uma trajetória que o usuário só realiza em um domingo pela manhã pode ser desconsiderada numa possível previsão de uma trajetória realizada de segunda a sexta, das 07h às 08h.

Quadro 1 – Principais trabalhos relacionados à temática de previsão de trajetórias

Trabalho	Item A	Item B	Item C	Item D	Item E	Item F	Item G	Item H
Ying et al. (2011)	N	C	C	N	A	T	T	NA
Froehlich e Krumm (2008)	S	I	G	N	P	G	A	NA
Tiwiri et al. (2013)	S	I	S	N	P	G	A	NA
Burbey e Martin (2008)	N	I	SD	N	P	G	A	NA
Xue et al. (2013)	S	C	C	S	A	G	C/O	NA
Herder et al. (2014)	N	I	SD	N	A	G	P/O	NA
Simmons et al. (2006)	S	C	S	N	P	G	A	S
Krumm (2008)	S	I	S	N	P	G	A	NA
Huang, Ying e Tseng (2012)	NA	I	C	S	A	T	A	S
Zhu et al. (2012)	NA	NA	SD	S	A	T	A	S
Figueiredo et al. (2016)	NA	I	SD	N	A	G	A	NA
Nademgeba et al. (2012)	S	I	C	N	A	T	A	NA
Lee et al. (2016)	N	I	C	N	A	G	A	NA
Lie et al. (2016)	NA	I	SD	N	A	G	P/O	S
Trasarti et al. (2015)	S	C	G	S	P	G	P/O	NA
Lung, Chung e Dai (2014)	S	I	NA	NA	NA	T	A	NA
Rocha et al. (2016)	NA	I	C	N	P	G	P/O	NA
Nobre Neto, Baptista e Campelo (2016)	S	I	S	S	A	T	T	S

Fonte: Autoria própria

Item A: Previsão em tempo real? Valores possíveis: Sim (S), Não (N) ou Não Informado (NA)

Item B: Modelo de previsão proposto é coletivo ou individual? Valores possíveis: Coletivo (C), Individual (I) ou Não Informado (NA)

Item C: Qual a granularidade da composição de uma rota? Valores possíveis: células/grade (C), segmentos (S), Coordenadas Geo-Temporais (G), Somente Destino (SD) ou Não Informado (NA)

Item D: Prevê lugares não visitados anteriormente? Valores possíveis: Sim (S), Não (N) ou Não Informado (NA)

Item E: Se dados de usuários reais, a base de rotas é própria ou pública? Valores possíveis: Própria (P) ou Pública/Aberta (A)

Item F: Utiliza que tipo de informação contextual? Geográfica, Temporal e Semântica (T); ou somente Geográfica e Temporal (G);

Item G: Quais são as métricas utilizadas para avaliação (precisão, cobertura e medida-F) do preditor? Valores possíveis: Todas as métricas, isto é, precisão, cobertura e medida-F, (T); Precisão (P); Cobertura (C); Medida-F (M); Acurácia (A); Outra (O); e Não Informado (NA);

Item H: Há validação cruzada? Sim (S), Não (N) ou Não Informado (NA)

Os modelos de previsão que contemplam, além de informações espaço-temporais, conteúdo semântico, têm a capacidade de distinguir o papel que os lugares representam ao usuário. Com isso, é possível filtrar as previsões a serem feitas também com base na semântica. Por exemplo, em um domingo à noite, dificilmente um usuário estará se deslocando para o trabalho, ao sair de um restaurante. A trajetória até o trabalho pode ser desconsiderada, portanto, para fins de previsão.

A Figura 2 apresenta um exemplo de um preditor que prevê o destino e a rota, dado um percurso parcial e informações espaço-temporais e semânticas sobre o deslocamento. Na figura, o usuário está se deslocando de uma região geográfica que representa sua casa e percorreu cinco segmentos (representado pela linha sólida, na cor azul). Após percorrer o quinto segmento, foi realizada uma previsão de destino e de rota. Foram utilizadas, ainda, as informações a respeito do dia da semana e horário de realização da rota. O destino previsto é uma região geográfica que representa o trabalho, e a rota prevista até este destino é representada pela linha tracejada, na cor laranja.

5 Desafios

Para trabalhar com a temática de previsão de trajetórias, alguns desafios precisam ser delineados. A seguir, são apresentadas as principais questões para se trabalhar com previsão de trajetórias:

- **Uso de base de dados:** a maioria dos trabalhos nesta área costuma utilizar base de deslocamentos reais. Essas bases, no entanto, podem não estar disponíveis, uma vez que questões de privacidade impedem a divulgação destes dados. Além disso, algumas bases de dados que estão disponíveis para acesso não possuem um rastreamento preciso (por coordenadas espaciais, por exemplo), dificultando a construção de um modelo que manipula um conjunto de coordenadas;

Figura 2 – Exemplo de um preditor, que prevê o destino e a rota



Fonte: Autoria própria.

- **Definição da representação da rota:** uma rota pode ser representada de variadas maneiras (por um conjunto de coordenadas espaciais ou por um conjunto de segmentos). Utilizar um conjunto de segmentos fará com que o modelo seja mais rápido e tenha uma razoável precisão. Utilizar um conjunto de células diminui, ainda mais, a precisão, comparado ao uso de segmentos;
- **Abordagem para previsão:** boa parte das abordagens de previsão analisam trajetórias históricas, previamente realizadas,

para identificação de similaridades com uma trajetória em curso. Caso o usuário esteja, entretanto, realizando uma trajetória nova, o preditor pode sugerir um lugar que resultará em erro. Identificar uma trajetória nova, para evitar uma previsão errada, pode ser considerado como um desafio;

- Contexto de aplicação do preditor: o preditor deverá ser elaborado conforme seu contexto de atuação: se de forma personalizada ou de forma coletiva.

6 Conclusão

Este artigo apresentou trabalhos relacionados à temática de previsão de trajetórias. Foram definidos os conceitos sobre o que é uma trajetória e sobre o papel de um modelo de previsão. Ainda sobre os modelos de previsão, estes foram agrupados conforme as informações que esses modelos manuseiam: se utilizam informações espaciais, espaço-temporais ou, ainda, se adicionam semântica. Detalhou-se, também, que os preditores podem prever não somente o destino de uma trajetória como também a rota até o alcance do destino previsto. A utilidade de prever toda a rota está na possibilidade de o preditor ser capaz de alertar sobre algum imprevisto ao longo da rota, como um tráfego congestionado. Com relação às técnicas utilizadas, muitos dos modelos de previsão sugeridos analisam trajetórias já realizadas e aplicam alguma técnica que identifica similaridade da trajetória a ser prevista com alguma trajetória já realizada.

Embora a temática de previsão de trajetórias venha recebendo atenção e possua uma considerável literatura, ainda há lacunas a serem preenchidas. Por exemplo, uma lacuna em aberto está relacionada à previsão de trajetórias que estão sendo realizadas pela primeira vez. A possibilidade de prever um destino nunca realizado anteriormente não é trivial, já que a previsão não possuirá dados históricos para previsão. Um caminho de investigação viável, no entanto, está na identificação do perfil de trajetos já realizados pelo usuário. Saber quais tipos de lugares são mais visitados, e em quais horários, pode resultar em uma previsão (ou sugestão) de destino ainda não visitado.

REFERÊNCIAS

- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.
- BRUSH, A.; KRUMM, J.; SCOTT, J. Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location. **UbiComp'10**, p. 95-104, 2010.
- BURBEY, I.; MARTIN, T. L. Predicting Future Locations Using Prediction-by-Partial-Match. In: **Proceedings of the first ACM international Workshop on Mobile Entity Localization and Tracking in GPS-less Environments**, 2008. p. 1-6.
- CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, p. 273-297, 1995.
- FIGUEIREDO, F. et al. TribeFlow: Mining & Predicting User Trajectories. In: **Proceedings of the 25th International Conference on World Wide Web**, Montreal, abr. 2016. p. 695-706.
- FROEHLICH, J.; KRUMM, J. Route Prediction from Trip Observations. **Society of Automotive Engineers (SAE)**, abr. 2008.
- HERDER, E.; SIEHNDEL, P.; KAWASE, R. **Predicting User Locations and Trajectories**. In: 22nd International Conference, UMAP 2014. Aalborg, Denmark: Springer. 2014. p. 86-97.
- HUANG, C. M.; YING, J. J. C.; TSENG, V. S. **Mining Users' Behaviors and Environments for Semantic Place Prediction**. Mobile Data Challenge by Nokia Workshop. Newcastle, 2012. p. 6.
- HUTTENLOCHER, D. P.; RUCKLIDGE, W. J. **A Multi-Resolution Technique for Comparing Images using the Hausdorff Distance**. Cornell University. Ithaca, p. 23, 1992.
- KRUMM, J. A Markov Model for Driver Turn Prediction. **Society of Automotive Engineers (SAE)**, abr. 2008.
- LAURILA, J. K. et al. **The Mobile Data Challenge: Big Data for Mobile Computing Research**. Proc. Mobile Data Challenge Workshop (MDC) In: conjunction with Pervasive. Newcastle, 2012.
- LEE, S. et al. Next Place Prediction Based on Spatiotemporal Pattern Mining of Mobile Device Logs. **Sensors**, p.1-19, jan. 2016.

LEI, P.; LI, S.; PENG, W. QS-STT: QuadSection clustering and spatial-temporal trajectory model for location prediction. **Distributed and Parallel Databases**, p. 231-258, 2013.

LIU, X.; LIU, Y.; LI, X. Exploring the context of locations for personalized location recommendations. **Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence**, New York, p. 1188-1194, jul. 2016.

LUNG, H. Y.; CHUNG, C. H.; DAI, B. R. Predicting Locations of Mobile Users Based on Behavior Semantic Mining. **Trends and Applications in Knowledge Discovery and Data Mining**, v. 8643, p. 168-180, 2014.

MONREALE, A. et al. **WhereNext**: a Location Predictor on Trajectory Pattern Mining. **Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)**. Paris: ACM. p. 637-646. 2009.

MORZY, M. Prediction of moving object location based on frequent trajectories. **ISICIS**, p. 583-592, 2006.

NADEMBEGA, A.; TALEB, T.; HAFID, A. A Destination Prediction Model based on historical data, contextual knowledge and spatial conceptual maps. In: **Communications (ICC), 2012 IEEE International Conference on IEEE**, p. 1416-1420, jun. 2012.

NATEKIN, A.; KNOLL, A. Gradient boosting machines, a tutorial. **Frontiers in Neurorobotics**, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885826/>>. Acesso em: fev. 2016.

NOBRE NETO, F. D.; BAPTISTA, C. S.; CAMPELO, C. E. C. A user-personalized model for real time destination and route prediction. In: **Proceedings of the 19th IEEE International Conference on Intelligent Transportation System (ITSC)**. Rio de Janeiro, p. 401-407. 2016.

PENG, C.Y. J.; LEE, K. L.; INGERSOLL, G. M. An Introduction to Logistic Regression Analysis and Reporting. **The Journal of Education Research**, v. 96, n.1, p. 3-14, 2002.

PLATT, J. C. **Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines**. MIT Press. , p. 21. 1998. (MSR-TR-98-14).

QUDDUS, M. A.; NOLAND, R. B. A high accuracy fuzzy logic based map matching algorithm for road transport. **Journal of Intelligent Transportation Systems**, v. 10, n. 3, p. 103-115, 2006.

QUILAN, J. R. **C4.5**: Programs for Machine Learning. San Francisco: Morgan Kaufmann, 1993.

ROCHA, C. L. et al. **TPRED**: a Spatio-Temporal Location Predictor Framework. In: **Proceedings of the 20th International Database Engineering & Applications (IDEAS)**. Montreal: ACM. p. 34-42, 2016.

RON, D.; SINGER, Y.; TISHBY, N. The power of amnesia: Learning probabilistic automata with variable memory length. **Machine Learning**, p. 117-149, 1996.

SPACCAPIETRA, S. et al. A conceptual view on trajectories. **Data & Knowledge Engineering**, v. 65, n. 1, p. 126-146, 2008.

SIMMONS, R. et al. Learning to Predict Driver Route and Destination Intent. In: **Intelligent Transportation Systems Conference**, p.127-132, 2006.

TANAKA, K. et al. A Destination Prediction Method Using Driving Contexts and Trajectory for Car Navigation Systems. In: **Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09**. p.190-195, 2009.

TIWIRI, V. S.; ARYA, A.; CHATUVERDI, S. Route prediction using trip observations and map matching. In: **IEEE 3rd International Advance Computing Conference (IACC)**, p. 583-587, feb. 2013.

TRASARTI, R. et al. MyWay: Location prediction via mobility profilin. **Information Systems**, v. 64, p. 1-18, nov. 2015.

XUE, A. Y. et al. Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection

Against Such Prediction. In: **Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE**, 2013, p. 254-265.

YING, J. J. C.; LEE, W. C.; WENG, T. C. Semantic Trajectory Mining for Location Prediction. In: **Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**, 2011. p. 34-43.

YING, J. C. et al. Semantic trajectory-based high utility item recommendation system. **Expert Systems with Applications**, v. 41, n. 10, p. 4762-4775, 2014.

ZHENG, Y. et al. Mining interesting locations and travel sequences from GPS trajectories. In: **Proceedings of International conference on World Wild Web (WWW 2009)**, 2009. p. 791-800.

ZHU, Y. et al. Feature engineering for semantic place prediction. **Pervasive and Mobile Computing**, Amsterdam, p. 772-783, dec. 2013.