ANGÉLICA DE SOUSA PINTO

UM SISTEMA PARA PREDIÇÃO DE TEMPO DE PERCURSO DE ROTA

Proposta de Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas Computacionais. Área de Concentração: Sistemas Computacionais.

Orientador: Prof. Dra. Flávia Bernardini

Coorientador: Prof. Dr. Patrick Barbosa Moratori

RIO DAS OSTRAS, RJ

2016

**RESUMO**

O tempo de viagem de uma cidade de origem para o destino não é um fator constante. Variáveis como condições climáticas, dia da semana e horário influenciam no bom andamento do fluxo de transportes e consequentemente na hora da chegada do destino desejado. O congestionamento causa perda de tempo, prejudica a saúde e atrapalha o crescimento do país. Conseguir prever o tempo que é gasto em uma viagem, em diferentes dias da semana, no feriado, véspera de feriado calculando o tempo necessário gastando o mínimo de tempo possível, aumenta a qualidade de vida das pessoas tornando o tempo mais útil. As tecnologias existentes não apresentam funcionalidade que faça a previsão do tempo gasto em uma viagem baseado em dados coletados da própria estrada. Com algoritmos online, é possível aplicado em uma base de dados coletados de vários pontos da estrada, fazer uma predição do tempo gasto de uma viagem. Algoritmos de regressão online são aplicados na base de dados e os resultados avaliados para descobrir o mais adequado para o problema proposto.

**SUMÁRIO**

Sumário

[**1** **INTRODUÇÃO** 3](#_Toc458254202)

[1.1 OBJETIVOS 3](#_Toc458254203)

[1.2 ORGANIZAÇÃO DA PROPOSTA 3](#_Toc458254204)

[**2** **ALGORITMOS DE APRENDIZADO DE MÁQUINA ONLINE PARA PROBLEMAS DE REGRESSÃO** 4](#_Toc458254205)

[2.1 REGRESSÃO LINEAR online 5](#_Toc458254206)

[2.2 Rede Neural Artificial ONLINE 6](#_Toc458254207)

[**3** **TECNOLOGIAS E TRABALHOS RELACIONADOS** 1](#_Toc458254210)1

[3.1 TECNOLOGIAS EXISTENTES 1](#_Toc458254211)1

[3.2 TRABALHOS RELACIONADOS 1](#_Toc458254212)3

[**4** **METODOLOGIA DA PESQUISA** 1](#_Toc458254213)4

[**5** **ATIVIDADES E CRONOGRAMA** 1](#_Toc458254214)7

[**6** **considerações finais**](#_Toc458254215) 18

[REFERÊNCIAS 2](#_Toc458254217)18

1. **INTRODUÇÃO**

Um sistema de unidades móveis está instalado em vários veículos no tráfego. Estas unidades móveis incluem os dispositivos de comunicações sem fios e aparelhos que determina a localização de cada veículo. Monitorar a posição de um veículo, tal como uma função do tempo revela também a velocidade do veículo. Posição e informações de velocidade são transmitidas periodicamente pelos veículos a uma estação de monitoramento central e aos veículos vizinhos. Na estação de monitoramento central, a entrada coletiva de um conjunto de veículos é processada para fornecer um gráfico instantâneo das condições de tráfego na área. Avisos de atrasos ou atualizações sobre as condições de tráfego na estrada à frente são usados ​​como parte de um sistema de estrada para automóveis inteligentes. Veículos dentro de uma região se comunicam com outros dentro de uma rede na qual as informações de difusão são processadas localmente nos respectivos veículos para estimar possíveis problemas com antecedência e considerar o cálculo de uma estrada alternativa e/ou verificar com a estação central de monitoramento para mais informações. Se fora do alcance da estação central de monitoramento, os veículos formam uma rede de área local para a troca e atualização de informações, e quando qualquer veículo da rede está dentro do alcance da estação central de monitoramento, os dados de rede de área local são enviados para ajudar a atualizar as informações de tráfego em geral.

# OBJETIVOS

O objetivo deste trabalho é investigar o uso de algoritmos de aprendizado de máquina online para problemas de regressão, predizer o horário de partida de maneira a minimizar o tempo de viagem, por exemplo, deseja-se obter o melhor horário de partida saindo de Rio das Ostras para o Rio de janeiro, passando o menor tempo possível viajando. Os algoritmos a serem investigados são regressão linear online e rede neural artificial (perceptron) online.

# ORGANIZAÇÃO DA PROPOSTA

No Capítulo 2 são descritos algoritmos de problemas de regressão online a serem utilizados neste trabalho.

No Capítulo 3 é apresentada a metodologia de investigação a ser utilizada para o desenvolvimento da dissertação.

No Capítulo 4 são apresentadas as atividades a serem desenvolvidas e um cronograma de trabalho.

No Capítulo 5 são apresentadas as considerações finais.

1. **ALGORITMOS DE APRENDIZADO DE MÁQUINA ONLINE PARA PROBLEMAS DE REGRESSÃO**

A Aprendizagem de Máquina é a ciência que faz com que os computadores exerçam seu papel de forma natural sem que pareçam explicitamente programados para tal (NG, A., 2016). Na última década, a aprendizagem de máquina foi responsável pelo surgimento dos carros automáticos, recursos de reconhecimento de fala, otimizou as buscas na web e possibilitou um avanço enorme na compreensão do genoma humano estando difundida com frequência na vida das pessoas.

A forma de organizar os algoritmos de aprendizado de máquina é útil porque força a pensar sobre os papéis dos dados de entrada e o processo de preparação do modelo e seleciona o que é o mais adequado para o problema, a fim de obter o melhor resultado.

Uma característica comum em sistemas de aprendizado de máquina é através de um grande volume de informações, identificar padrões para extrair conclusões e realizar uma tarefa e quanto mais tempo e informações absorvidas, melhores se tornam as respostas.

Nas últimas décadas, com a crescente complexidade dos problemas a serem tratados computacionalmente e do volume de dados gerados por diferentes setores, tornou-se clara a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, reduzindo a necessidade de intervenção humana e dependência de especialistas. Para isso, essas técnicas deveriam ser capazes de criar por si próprias, a partir da experiência passada, uma hipótese, ou função, capaz de resolver o problema que deseja tratar. Um exemplo simples é a descoberta de uma hipótese na forma de uma regra ou conjunto de regras para definir que clientes de um supermercado devem receber material de propaganda de um novo produto, utilizando para isso dados de compras passados dos clientes cadastrados na base de dados do supermercado. A esse processo de indução de uma hipótese (ou aproximação da função) a partir da experiência passada dá-se o nome Aprendizado de Máquina (FACELI *et al*., 2011).

O modelo de Regressão pode conter uma ou dezenas de variáveis de entrada e uma variável de saída e têm o mesmo padrão geral. Há algumas variáveis independentes que, quando tomadas em conjunto produzem um resultado, uma variável dependente. O modelo de regressão é usado para prever o resultado de uma variável dependente desconhecida, dados os valores das variáveis independentes.

Algoritmos de aprendizado online são aqueles que recebem suas entradas como uma sequência de partes e processam cada uma recebida sem conhecimento daquelas que estão por vir (FIAT, WOEGINGER, 1998). Ao processar uma parte, o algoritmo deve tomar decisões irrevogáveis que afetem sua solução. Por isso, algoritmos online são interessantes para tratar problemas em que não se tem todas as informações. Nesta seção são descritos dois algoritmos de aprendizado online: regressão online e redes neurais artificiais online.

# REGRESSÃO LINEAR online

**Regressão Linear Clássica:** Considere o problema de aproximar o conjunto de dados por uma função linear , já que a função de regressão ótima, dada pelo mínimo do funcional, + ), onde C é um valor  pré-determinado, e , são variáveis de folga representando restrições superiores e inferiores nas saídas do sistema, é uma função é desconhecida. Usando uma função de perda insensível de outra forma , a solução é dada por, , definida na Eq 1, ou alternativamente, dado pela Eq. 2, para a qual devem ser respeitadas as restrições definidas pelas Eq. 3 e 4.

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (2) |

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

**Regressão Online:** No contexto de aprendizado online a função candidata (geralmente chamada de hipótese) é construída através da minimização do risco empírico examinando um exemplo de treinamento por vez. Dessa maneira, inicia-se com uma hipótese inicial e, a cada iteração, o algoritmo examina um exemplo e atualiza a hipótese atual de acordo com uma regra de correção específica. Com o objetivo de derivar essa regra de correção segue-se as ideias do algoritmo Perceptron (ROSENBLATT, 1958) usando a abordagem da descida do gradiente estocástica. Considerando o risco empírico definido na seção 2, define-se o seguinte custo:

que deve ser minimizado em relação a. Assim, para cada par de pontos , a seguinte regra de correção é aplicada à hipótese atual

(2.3.1.1)

em que η > 0 é geralmente chamado de taxa de aprendizado e denota o gradiente da função de perda em relação a .

Um aspecto importante dessa abordagem é que se , o que é verdadeiro para a maioria das funções de perda, a atualização acima precisa ser efetuada somente nos casos em que. Caso contrário, a hipótese atual já atingiu o mínimo para o exemplo e não é necessário proceder qualquer correção, isto é, . Nesse sentido, funções de perda que são baseadas no conceito de tubo são bem adequadas para esse esquema, uma vez que o exemplo somente afetará a hipótese atual caso encontre-se fora do tubo.

# rede neural artificial ONLINE

**Redes Neurais Tradicionais (Aprendizado em batch):** As redes neurais artificiais têm sua inspiração biológica originada de um neurônio típico, formado por dendritos, o corpo celular e os axônios, que junto com outros neurônios podem ter várias combinações com as mais variadas topologias. A ideia básica é sumarizada na abordagem Reducionista: reproduzindo-se com suficiente detalhe a suposta "máquina" biológica responsável pelo comportamento inteligente, (ou seja, o cérebro), um comportamento inteligente emergirá do sistema. Historicamente, os primeiros trabalhos nessa área iniciaram-se na década de 40 com a Cibernética. Todavia, somente a partir da década de 80 esta tecnologia tomou impulso para a representação do conhecimento e para o aprendizado de máquina. Na figura 1 apresenta a descrição dos componentes do neurônio artificial.

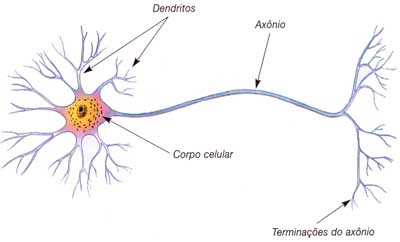


Figura 1: Esquema simplificado de um neurônio.

Fonte: PAIVA, Anabelle Barroso de/ CEFET/SC.

**Sinapse:** No interior da membrana há uma alta concentração de moléculas específicas de neurotransmissores. Os neurotransmissores ficam armazenados em vesículas especiais denominadas vesículas sinápticas. Quando a membrana é despolarizada, isto é, quando o potencial do citoplasma se torna maior que o potencial de repouso, os canais da membrana permitem íons como o cálcio fluir para dentro da célula, estes ativarão a liberação dos neurotransmissores (tal como acetilcolina) constituindo-se assim a sinapse, que é a ligação entre a terminação axônica e os dendritos e que permite a propagação dos impulsos nervosos de uma célula a outra. As sinapses podem ser excitatórias ou inibitórias. As sinapses excitatórias cujos neuro-excitadores são os íons sódio permitem a passagem da informação entre neurônios e as sinapses inibitórias, cujos neuro-bloqueadores são os íons potássio, bloqueiam a atividade da célula, impedindo ou dificultando a passagem da informação.

**O Neurônio Artificial:** O corpo faz a soma ponderada do produto dos pesos, são a intensidade da força sináptica e podem ser fixos ou treináveis implementando as ligações entre as unidades e a intensidade com que o sinal é transmitido de um neurônio ao outro. O modelo do neurônio artificial de acordo com (Dayhoff, 1992) é apresentado na Figura 2. São elementos do neurônio artificial a função soma, a função de transferência e a função de ativação:

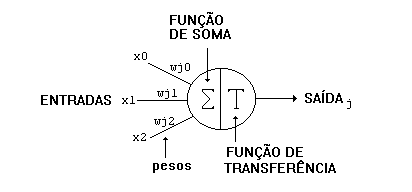
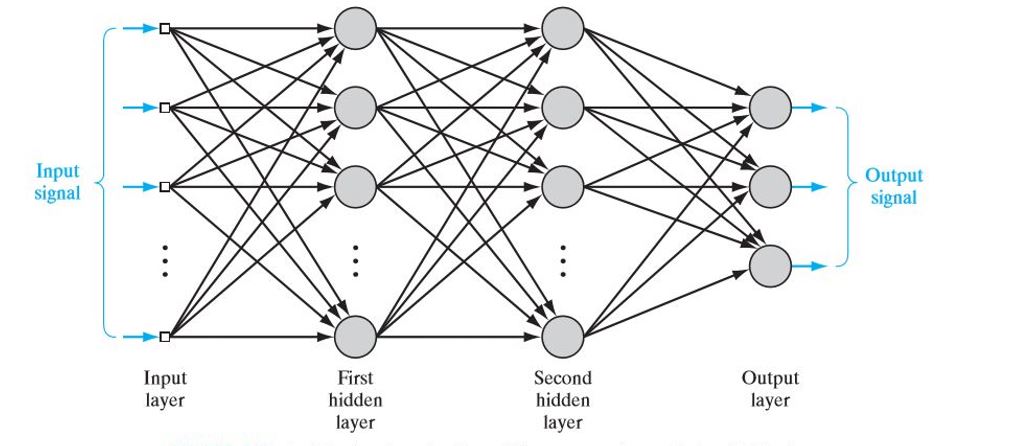


Figura 2: Modelo do Neurônio Artificial.  
Fonte: (TAFNER, 1998).

* **Função Soma** - as entradas depois de multiplicadas pelos pesos, recebem o nome de entradas ponderadas, ou unidades sigma.
* **Função de Transferência** - depois de acumulado o valor somado, o neurônio deve comparar esse valor com um valor de limiar (um valor estipulado), e atingindo-o, o valor é então passado adiante através da saída. A esse processo chamamos de função de transferência. Se o valor não atinge o limiar, o sinal não é passado adiante. A lógica neural expõe, que a intensidade dos sinais de entrada, dispara ou não, o sinal do neurônio, fazendo com que este estimule o neurônio seguinte. A função de transferência define e envia para fora do neurônio o valor passado pela função de ativação.
* **Função de ativação** - A função de ativação antecede a função de transferência, e tem por atribuição repassar o sinal para a saída do neurônio. A função de ativação é uma função de ordem interna, cuja atribuição é fazer acontecer um nível de ativação dentro do próprio neurônio, ou seja, uma decisão é tomada pelo próprio neurônio sobre o que fazer com o valor resultante do somatório das entradas ponderadas. Essa decisão terá efeito restrito ao próprio neurônio. Em modelos simples de redes neurais, a função de ativação pode ser a própria função soma das entradas ponderadas do neurônio. Em modelos mais complexos, a função de ativação possui um processamento atribuído. Esse processamento pode usar o valor prévio de saída como uma entrada para o próprio neurônio. Após ter sido processado pela função de ativação, é então passado pela função de transferência que produzirá o valor de saída do neurônio.

**O Algoritmo back-propagation**



**Figura 3**: Diagrama arquitetural do perceptron multicamadas com duas camadas escondidas

**Fonte:** (Simon, H., 2009)

A figura 3 apresenta o layout arquitetônico de um perceptron. O gráfico de sinal de fluxo correspondente para a aprendizagem back-propagation, incorporando as fases para frente e para trás dos cálculos envolvidos no processo de aprendizagem, é apresentada na Figura 4 para o caso de L=2 e 3. A parte superior do gráfico de fluxo de sinal para a passagem para frente. A parte inferior do gráfico de fluxo de sinal é responsável pelo passe para trás, o que é referido como um gráfico de sensibilidade para calcular os gradientes locais no algoritmo back-propagation (Narenda and Parthasaraty, 1990). A atualização sequencial dos pesos é o método preferido para a execução online do back-propagation. Por este modo da operação, os ciclos do algoritmo através da amostra de treino como a seguir:

1. **Inicialização** – Assumindo do princípio que nenhuma informação prévia está disponível, escolher os pesos sinápticos e os limiares de uma distribuição uniforme cuja variância é escolhida para fazer o desvio padrão dos campos locais induzidos dos neurônios que estão na transição entre as partes padrões e lineares da função de ativação sigmoide.
2. **Apresentação de exemplos de treinamento** – Apresentar exemplos de período de treinamento da rede. Para cada exemplo na amostra, ordenada de alguma forma, realizar a sequência para frente e para trás cálculos descritos sob os pontos 3 e 4 respectivamente.
3. **Computação Para Frente** - Permitir que um exemplo de formação do período seja denotado por (**x**(n), **d**(n)), com o vetor de entrada **x**(n) aplicado na camada de entrada de gânglios sensoriais e o vetor da resposta desejada **d**(n), apresentada à camada de saída de nós da computação. Calcular os campos locais induzidos e sinais de funções, procedendo através da rede, camada por camada. O campo local induzido para neurônio na camada

|  |  |
| --- | --- |
|  |  |

Onde é a saída (função) sinal do neurônio na camada anterior na iteração n, e (n) é o peso sináptico do neurônio na camada que é alimentado a partir do neurônio na camada Para , então, , e é o viés aplicado para o neurônio na camada . Assumindo o uso da função sigmoide, o sinal de saída do neurônio na camada é

=

Se o neurônio é na primeira camada escondida (), então

Onde é o th elemento do vetor de entrada **x**. Se neurônio é na camada de saída ( , onde é referido como a profundidade da rede), então

Cálculo do sinal de erro

|  |  |
| --- | --- |
|  |  |
|  |  |

Onde é o th elemento do vetor de resposta desejado **d**.

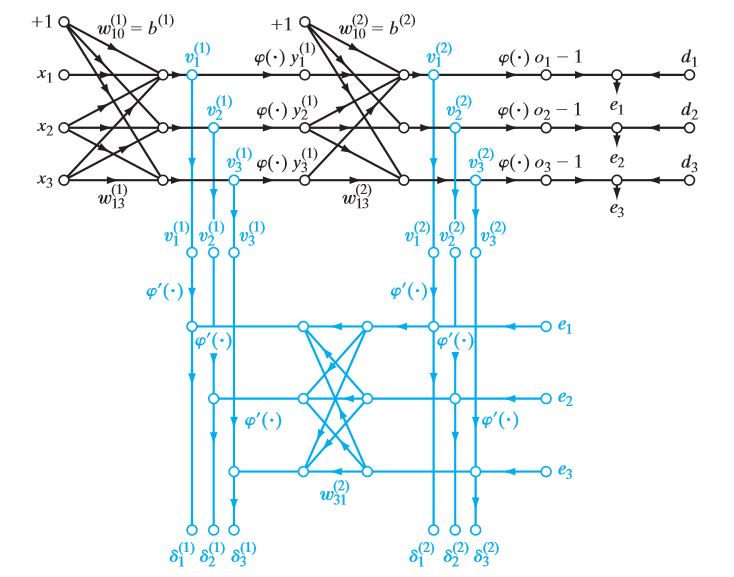
1. **Computação para trás**. Calcula o s (gradientes locais) da rede, definido por

Onde o primordial no (.) denota diferenciação em relação ao argumento. Ajuste dos pesos sinápticos da camada na rede de acordo com a regra delta generalizada

Onde η é o parâmetro da taxa de aprendizado e é a constante impulso

1. **Iteração** -Iterar os cálculos frente e para trás nos pontos 3 e 4, apresentando novas épocas de exemplos de treinamento para a rede até que o critério de parada escolhido for atendido.

*Notas*: A ordem de apresentação de exemplos de treinamento deve ser aleatória de período a período. O impulso e parâmetro da taxa de aprendizagem são tipicamente ajustado (e normalmente diminuído) como um número de iterações de treinamento aumenta.



**Figura 4**: Resumo do gráfico de fluxo de sinal do aprendizado Back-Propagation. Parte de cima do gráfico passe para frente. Parte baixa do Gráfico, passe para trás.

**Fonte:** (Simon, H., 2009)

1. **TECNOLOGIAS E TRABALHOS RELACIONADOS**

# TECNOLOGIAS EXISTENTES

O Waze (Suporte Google, 2016) é um aplicativo que compartilha informações em tempo real sobre o trânsito, acidentes, policiais, vias interditadas e condições climáticas. O aplicativo coleta as informações adicionadas pelo usuário e as analisa imediatamente para fornecer aos outros usuários (Wazers) a melhor rota para seus destinos, 24 horas por dia. Dirigir com o aplicativo aberto é uma maneira de contribuir com sua melhoria, quanto mais pessoas dirigirem com o Waze aberto, melhor será a navegação. As informações do percurso são utilizadas para calcular a velocidade média, encontrar erros e melhorar o traçado das vias, mapas e navegação além de aprender os sentidos de condução e conversões. O Waze funciona melhor nas suas rotas diárias e não há navegação disponível para o transporte público, bicicletas e caminhões.

Para que o aplicativo possa fornecer com precisão a rota mais rápida ou mais curta, de acordo com suas configurações, necessita de dados precisos de todos os segmentos e rotas vizinhas. O Waze coleta dados de cada trecho dirigido com o aplicativo aberto.  Na próxima vez em que uma via específica for percorrida, o aplicativo saberá como comparar os dados entre cada rota e saberá sugerir a rota ideal, conforme as preferências do usuário.

Algumas vezes, existem rotas melhores do que as sugeridas, problemas de roteamento como estes acontecem porque o sistema funciona em tempo real e usa a média dos dados estatísticos, por isso, o aplicativo sugere que algumas vezes, o usuário siga o caminho sugerido para que o Waze possa coletar novas informações e usá-las nos algoritmos de roteamento.

Quando o Waze não oferece a melhor rota, pode ser que houve um erro no mapa em algum lugar ao longo da rota preferida do usuário ou ausência de informações corretas de velocidade e trânsito na rota que está tentando enviar instante.

Para que todas as funções do Waze possam operar, é necessária uma conexão de dados ativa no dispositivo móvel para aprimorar com informações em tempo real sobre o trânsito e garantir um mapa atualizado. Se houver uma conexão intermitente, os dados serão obtidos dos servidores para alertas de trânsito e perigos, mas pode não ser capaz de fornecer informações confiáveis. Além disso, se o Waze não conseguir restabelecer a conexão com os servidores, o usuário não conseguirá enviar alertas, pois estes ou problemas não são armazenados no mapa para um posterior envio.

O Google Maps é um serviço de pesquisa e visualização de mapas e imagens de satélite da Terra gratuito na web fornecido e desenvolvido pela empresa Google.

Atualmente, o serviço disponibiliza mapas e rotas para qualquer ponto em vários países como nos Estados Unidos, Canadá, União Europeia, Austrália e Brasil. Disponibiliza imagens de satélite do mundo todo, com possibilidade de um zoom nas grandes cidades, como Nova Iorque, Paris, São Paulo, Rio de Janeiro e Brasília. Nos Estados Unidos o Google fez uma parceria com órgãos públicos, que incluirão as linhas de trem americanas e seus cruzamentos com rodovias ao Google Maps.

Juntamente com o lançamento da versão brasileira do Google Maps, a empresa introduziu o Local Business Center, ferramenta que permite com que qualquer empresa faça seu cadastro e seja encontrada no Google Maps. No cadastro as empresas podem preencher seus dados cadastrais, horário de atendimento, formas de pagamento, logotipo e fotos.

Com uma conta Google, é possível destacar as rotas, pontos e áreas, gerar comentários e compartilhar os respectivos links de acesso ao mapa criado. Também é possível gerar um arquivo KML para integração com o Google Earth.

O Google Maps sincroniza as pesquisas realizadas no computador para smartphones Android possibilitando o download do mapa, para uso “off-line”. O aplicativo recalcula o trajeto dinamicamente com base nos próximos padrões de trânsito, ajudando o usuário a evitar congestionamentos.

# TRABALHOS RELACIONADOS

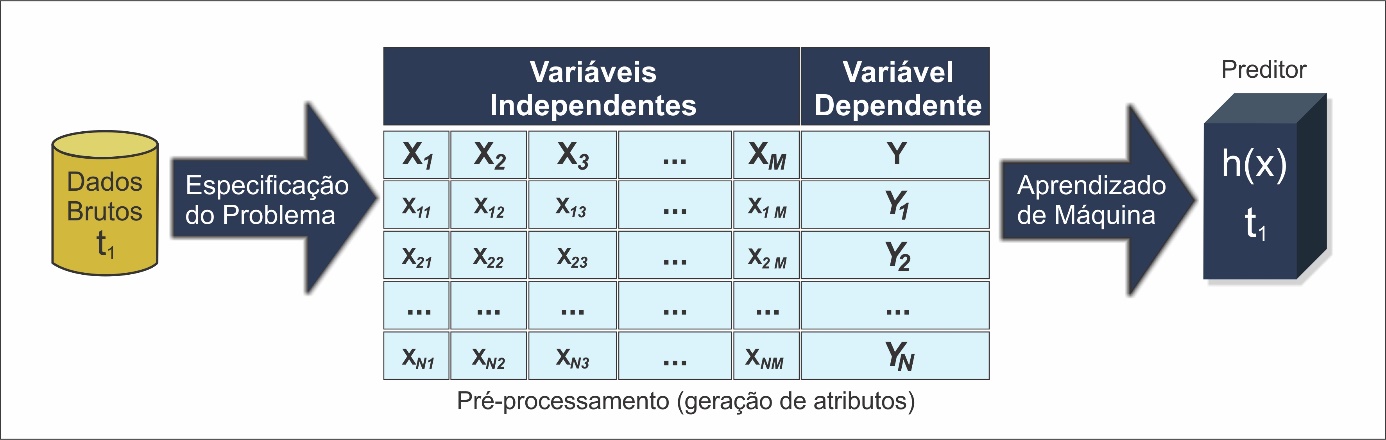
Em Science Direct, Sci-hub, Google Scholar, Scielo, ACM, IEEExplore, Periodicos Capes, Scielo e Scopus foram usados os termos de busca “traffic data prediction”, “traffic data analysis”, “traffic data analysis machine learning” e “traffic data analysis regression” no entanto, os dois trabalhos mais relacionados ao apresentado nesta proposta são descritos a seguir.

Em (WIBISONO, A. et al., 2015) a extração de informações através de sensores distribuídos tem sido amplamente utilizada em várias regiões. Veículo de extração de dados de tráfego é uma das maneiras de coletar informações sobre condição de tráfego. A pretensão de prever e visualizar as condições de tráfego em uma determinada região da estrada. Os dados de tráfego foram obtidos a partir do Departamento de Transportes do Reino Unido. Estes dados são recolhidos usando centenas de sensores durante 24h. Assim, o tamanho dos dados é muito grande. A fim de obter o comportamento da condição de tráfego, é necessário analisar o enorme conjunto de dados, que se obteve a partir dos sensores. As utilizações dos métodos de extração convencionais de dados não são suficientes para a utilização, devido ao processo de conhecimento construção de ponta que deve armazenar dados temporários na memória. O fato de que os dados vão se tornando maior ao longo do tempo, é preciso encontrar um método que poderia se adaptar automaticamente para processar dados sob a forma de fluxos. O método chamado FIMT-DD (Fast Incremental Modelo Trees-Deriva Detection) foi usado para analisar e prever o grande conjunto de dados de tráfego. Baseado no sistema de previsão de que foi desenvolvido, foi também visualizada a previsão da condição de fluxo de tráfego dentro do ponto de sensor gerado no mapa real de simulação. A presente proposta também trabalha com fluxo de dados dinâmico, além disso, retorna o melhor horário de saída do veículo.

Em (LEE, J. et al., 2015) são descritos novos conhecimentos entre congestionamento do tráfego e tempo usando tecnologia de processamento de um grande conjunto de dados armazenados. Os fatores climáticos como temperatura e precipitação em zonas residenciais e destinos turísticos afetam o fluxo de tráfego nas estradas circundantes. As mudanças no congestionamento do tráfego devido ao clima são avaliadas usando linear múltipla. A análise de regressão cria um modelo de predição e previsão de congestionamento do tráfego em uma base diária. Para a análise de regressão, foram usados 48 fatores de previsão meteorológica e seis variáveis ​​dummy(variáveis utilizadas para representar numericamente uma variável qualitativa) para expressar os dias da semana. A regressão linear múltipla do modelo final é então proposta com base nas três etapas de análise de (i) a criação de um modelo de regressão completa, (ii) a remoção das variáveis, e análise dos resíduos. Para verificar a sua previsibilidade, o modelo avalia o congestionamento do tráfego, prever o congestionamento do tráfego com o congestionamento de tráfego real. Usando o método de avaliação erro médio percentual absoluto, mostra que o modelo de regressão linear múltipla final tem uma precisão da previsão de 84,8%. Informações de condições de tempo não será considerado como variável de condição para o resultado final do trabalho proposto.

1. **METODOLOGIA DA PESQUISA**

A metodologia para o desenvolvimento deste trabalho consiste na continuidade de estudos de referencial bibliográfico que aborde a aplicação dos algoritmos de aprendizado online para problemas de regressão. Neste estudo, serão utilizados os algoritmos de regressão linear online e de construção de redes neurais artificiais perceptron online, que foram descritos no Capítulo 2.  Na Figura 5 é apresentado um diagrama esquemático das atividades a serem executadas para atingir o objetivo dessa pesquisa.



1 – Coleta dos Dados

Figura 5: Diagrama esquemático das atividades.  
Fonte: Elaboração própria.

**1 - Coleta dos dados:** Nessa etapa, os dados são coletados da Tomtom através de informações de medições anônimas de GPS combinadas com fontes tradicionais, como vans de mapeamento móvel e sensores rodoviários. A coleta de dados é realizada em determinado tempo *t*. Na Tabela 1 é apresentada uma amostra de dados que foram coletados na cidade do Rio de Janeiro (da Barra da Tijuca até o Centro). Foram coletados dados de 3 rotas distintas:

* Uma rota considerada curta, do bairro São Cristóvão até o Centro;
* Outra considerada de tamanho médio, do bairro Catete até o Centro; e
* Por último uma rota considerada longa, do bairro Barra da Tijuca até o Centro.

Consideramos rotas de tamanhos distintos para podermos verificar o comportamento dos regressores em diferentes rotas.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tempo** | **Distância(m)** | **Tempo Total(s)** | **Tempo de Atraso(s)** | **Data e hora de partida** | **Data e hora de chegada** |
| t | 5946 | 1098 | 0 | 2015-12-23T20:37:49-02:00 | 2015-12-23T20:56:06-02:00 |
| t+1 | 5946 | 1078 | 0 | 2015-12-23T20:52:56-02:00 | 2015-12-23T21:10:53-02:00 |
| t+2 | 5946 | 1052 | 0 | 2015-12-23T21:08:03-02:00 | 2015-12-23T21:25:35-02:00 |
|  | 5946 | 1026 | 0 | 2015-12-23T21:23:09-02:00 | 2015-12-23T21:40:15-02:00 |
|  | 5946 | 983 | 0 | 2015-12-24T21:47:23-02:00 | 2015-12-24T22:03:45-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T22:02:27-02:00 | 2015-12-24T22:18:44-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T22:17:31-02:00 | 2015-12-24T22:33:48-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T22:32:35-02:00 | 2015-12-24T22:48:52-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T22:47:39-02:00 | 2015-12-24T23:03:56-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T23:02:43-02:00 | 2015-12-24T23:19:00-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T23:17:46-02:00 | 2015-12-24T23:34:03-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T23:32:51-02:00 | 2015-12-24T23:49:08-02:00 |
|  | 5946 | 978 | 0 | 2015-12-24T23:47:54-02:00 | 2015-12-25T00:04:11-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T00:02:57-02:00 | 2015-12-25T00:19:14-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T00:18:01-02:00 | 2015-12-25T00:34:18-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T00:33:04-02:00 | 2015-12-25T00:49:21-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T00:48:08-02:00 | 2015-12-25T01:04:25-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T01:03:12-02:00 | 2015-12-25T01:19:29-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T01:18:15-02:00 | 2015-12-25T01:34:32-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T01:33:19-02:00 | 2015-12-25T01:49:36-02:00 |
|  | 5946 | 978 | 0 | 2015-12-25T01:48:22-02:00 | 2015-12-25T02:04:39-02:00 |

Tabela 1: Uma amostra dos dados disponíveis para o trabalho.   
Fonte: Dados coletados da Tomtom.

**2 - Pré-processamento dos dados:** Nessa etapa, os dados coletados devem ser transformados de tal maneira que possam ser utilizados em formato atributo-valor por algoritmos de aprendizado. É interessante que os atributos a serem utilizados para construção do regressor leve em consideração características do tráfego diário. Sendo assim, Na Tabela 1 são apresentados 5 atributos coletados. A variável a ser predita é a variável dependente, que no nosso caso é o tempo de deslocamento de uma rota (atributo Tempo Total da Tabela 1), dada a rota a ser executada (conhecida pela coleta dos dados) e a data e o horário de partida (atributo Data e Hora da Partida na Tabela 1). Nos dados coletados, as rotas possuem sempre a mesma distância, e portanto a princípio essa informação será descartada. A partir do atributo Data e Hora da Partida, serão extraídos outros atributos:

**Hora da partida**: Número real

**É feriado?:** (0=Não/1=Sim)

**Véspera de feriado?:** (0=Não/1=Sim)

**Final de semana?:** (0=Não/1=Sim)

**Terça/Quarta/Quinta?:** (0=Não/1=Sim)

**Segunda?:** (0=Não/1=Sim)

**Sexta?:** (0=Não/1=Sim)

**Sábado?:** (0=Não/1=Sim)

**Domingo?:** (0=Não/1=Sim)

**Pós-feriado?:** (0=Não/1=Sim)

**Pré-feriado?:** (0=Não/1=Sim)

Todos esses atributos serão utilizados como atributos independentes, cujos valores compõem o vetor x. A cada exemplo x de treinamento é associado um valor y, que é o tempo de execução da rota.

**3 - Indução dos Modelos:** Nessa etapa o modelo é revisto a partir do exemplo de treinamento. Serão utilizados os algoritmos de aprendizado Regressão Linear Online e Redes neurais Online.

**4 - Avaliação dos Modelos:** Os algoritmos Regressão Linear online e Redes Neurais Artificiais online são avaliados. Métricas para avaliação de algoritmos online serão utilizados nessa etapa, ainda a serem selecionadas..

1. **ATIVIDADES E CRONOGRAMA**

As seguintes atividades serão realizadas:

1 – Coleta e pré-processamento de dados

2 – Indução e Avaliação dos Modelos

3 - Escrita e Submissão do Artigo

4 - Escrita e Dissertação

5 - Defesa da Dissertação

6 - Conclusão

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2015 | | | | | | | 2016 | | | | | |
| Atividades\Meses | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. **considerações finais**

A presente proposta tem por objetivo apresentar o referencial teórico relacionado ao objetivo do projeto, bem como analisar as tecnologias e soluções acadêmicas relacionadas a este projeto. Assim foram apresentados os algoritmos que serão utilizados na execução do projeto para fazer uma análise quanto aos atributos utilizados e identificar qual apresenta o melhor resultado, segundo algum teste estatístico a ser realizado.

# REFERÊNCIAS

ALMEIDA, Maria Aparecida Fernandes. **Introdução ao Estudo das Redes Neurais Artificiais.**2000. Disponível em: <https://intranet.dcc.ufba.br/pastas/mecateam/material\_de\_estudo/ia/iaconexionista/rna.pdf>. Acesso em: 23 maio 2016.

BROWNLEE, Jason. **A Tour of Machine Learning Algorithms.**2013. Disponível em: <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>. Acesso em: 02 maio 2016.

FIAT, A.; WOEGINGER, G. J. **Online Algorithms:**The State of the Art (Lecture Notes in Computer Science). Springer, 1998.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social.**São Paulo: Atlas, 1999.

HAYKIN, S. **Neural Networks and Learning Machines.** Terceira edição. Pearson, 2009

KIVINEN, J.; SMOLA, A.; WILLIAMSON, R. **Online learning with kernels.**IEEE Transactions On Signal Processing, 2004. 52 v.

LEE, J. et al. **A Prediction Model of Traffic Congestion Using Weather Data**. IEEE International Conference on Data Science and Data Intensive Systems, 2015.

NOVIKOFF, A. B. **On convergence proofs for perceptrons.**In: Proceedings of the Symposium on the Mathematical Theory of Automata, 1963. 12 v.

SMOLA, A. J.; SCHÖLKOPF, B. **On a kernel–based method for pattern recognition, regression, approximation and operator inversion.**Algorithmica, 1998.

Suporte Google. 2016. Disponível em: <<https://support.google.com/waze/answer/6078702?hl=pt-BR>>. Acesso em: 23 março 2016.

TAFNER, Malcon Anderson. **O Que São as Redes Neurais Artificiais.**1998. Disponível em: <http://www.cerebromente.org.br/n05/tecnologia/rna.htm>. Acesso em: 26 maio 2016.

VAPNIK, V. **The Nature of Statistical Learning Theory.**New York: Springer, 1995.

VAPNIK, V. **Statistical Learning Theory.**New York: Wiley, 1998.

WENJIAN, W. et al. **Online prediction model based on support vector machine**. 2007. Disponível em <[www.sciencedirect.com](http://www.sciencedirect.com)>. Acesso em: 03 julho 2016.

WIBISONO, Ari et al. **Traﬃc big data prediction and visualization using Fast IncrementalModel Trees-Drift Detection (FIMT-DD).**2015. Disponível em: <https://www.researchgate.net/publication/284121757\_Traffic\_Big\_Data\_Prediction\_and\_Visualization\_using\_Fast\_Incremental\_Model\_Trees-Drift\_Detection\_FIMT-DD>. Acesso em: 26 maio 2016.

WIKIPÉDIA. **Google Maps.**2016. Disponível em: <https://pt.wikipedia.org/wiki/Google\_Maps>. Acesso em: 03 jun. 2016.