# Estado da Arte – Machine Learning

Problemática em causa: Aprendizagem Automática (AA) no problema dos transportes públicos terrestres (autocarros, metro, comboio, etc.).

Para que possamos usufruir de todas as qualidades de machine learning temos que antes garantir a qualidade dos dados utilizados. A tecnologia de smart cards permite uma recolha de dados fidedigna e eficaz tendo sido o método de recolha utilizado na pesquisa conduzida por Florian Toqué (Florian Toqué, 2017).

Os dados recolhidos são utilizados para prever a necessidade da deslocação dos utilizadores, é notável referir que existem diferenças entre o estudo da Origem-Destino e da fluidez de viagem. O estudo em causa foca-se na fluidez contrastando a importância da previsão a curto prazo, mas não esquecendo a possibilidade do planeamento a longo prazo, que seria possível com o aproveitamento de machine learning.

Existem vários exemplos que comprovam a utilidade de *machine* *learning* no desenvolvimento e melhoramento de horários nos autocarros nas grandes metrópoles, inclusive, um estudo feito por autores portugueses baseado na cidade do Porto (João Mendes-Moreira, 2015). Foram analisados dados recolhidos pelo *Automatic Vehicle Location* (AVL), equipamentos de comunicação baseados em localização GPS que enviam o *status* com a atual localização dos vários autocarros da frota.

O método de estudo utilizado por João Mendes-Moreira consiste em formar vários *clusters* (viagens, etc.) com informação seletiva, que são estudados de forma a obter vários valores de interesse, sendo o mais importante para a nossa pesquisa o TTP, ou *travel* *time prediction,* que consiste na duração total estimada das viagens. Tal como o que foi desenvolvido ao longo do projeto, devemos começar por criar um perfil do dia que contem, para uma determinada rota, informação sobre as viagens. A aplicação deste tipo de estudo permite uma análise mais detalhada e aprofundada.

Devem ser criados intervalos para os quais diferentes dados são agrupados. Um exemplo prático seria a divisão dos horários em 4 *schedules,* o primeiro seria de 1 de Janeiro até 15 de Julho, altura onde se iniciam as férias de verão para a maioria dos alunos, o segundo de 15 de Julho até 15 de Setembro, onde se dá a reabertura das escolas, o terceiro seria o intervalo de dias de trabalho compreendido entre 15 de Setembro e 31 de Dezembro, o ultimo seriam exclusivamente feriados e fins-de-semana. A criação destes 4 *schedules* deve-se à disparidade de valores prevista, é sabido que os estudantes compõe uma grande parte dos usuários de transportes públicos incluindo autocarros, por isso deve-se ter em mente que a carga horário dos meses de verão será muito reduzida em comparação com a de fim e inicio de ano. Para que haja mais detalhe no estudo devemos estudar estes intervalos independentes uns dos outros.

Um dos problemas enfrentados pela abordagem atual é que os algoritmos atuam apenas sobre turnos de motorista já definidos que são baseados em definições prévias que não podem ser avaliados por algoritmos automáticos. Isto cria um aumento no custo das operações por criar a necessidade de disponibilizar mais autocarros do que o que seria ideal, o mau aproveitamento de turnos faz com que também sejam pagas horas desnecessárias. A utilização de machine learning é a unica que garante que todos os fatores são considerados e avaliados.

Quanto à metodologia que envolve o *machine learning* propriamente dito, implica a aplicação de técnicas como o ‘*consensus clustering’* e *‘rule induction’* para descobrir informação relevante em grandes quantidades de dados.

*Machine Learning* pode ser utilizado também para a estimativa e previsão das exigências de viagens como se lê no artigo de Florian Toqué. Este explora de forma mais aprofundada a utilização de métodos de machine learning, tendo atenção às vantagens a curto e longo prazo. Utilizando as *digital footprints* existentes como dados a ser estudados é possível criar ferramentas que permitem prever a fluidez de passageiros nas cidades grandes e aumentar os níveis dos serviços e agendamento de transportes.

As vantagens da utilização de machine learning no desenvolvimento das soluções pretendidas é aparente, permite que sejam processadas múltiplas rotas em simultâneo e garante um maior aprofundamento de dados. Uma vez que é capaz de processar dados a um ritmo maior, seria possível diminuir os custos tanto para a empresa responsável por planear as rotas, como também, eventualmente, aos clientes, com o intuito de manter um preço competitivo.

Tem-se ainda em conta outro método de recolha de dados, que nos dias de hoje é uma evolução dos smart cards e permite obter uma base de dados mais completa, a leitura e armazenamento da informação proveniente dos sensores de telemóvel. Como foi mencionado inicialmente, a recolha de dados e os dados escolhidos para os estudos de machine learning são extremamente importantes e uma utilização errada vai limitar o seu potencial e criar resultados menos fidedignos. Tendo em conta a aplicação que está a ser desenvolvida, atualmente, seria necessária uma recolha de dados tanto de smart cards como de sensores de telemóvel uma vez que as aplicações móveis começam a ter a capacidade de libertar o utilizador dos cartões (Jahangiri, 2015).

O estudo dos sensores de telemóvel permite uma avaliação de diferentes transportes, cria a possibilidade de criar *smart scheduleing* para todos. Isto traria incríveis benefícios para os utilizadores que teriam acessos a horários com uma complementaridade quase perfeita.Um dos grandes desafios para o desenvolvimento deste tipo de horários seria a capacidade de diferenciar os vários meios de transporte com acesso apenas aos sensores.

Em suma, os benefícios da integração de machine learning são claros, a troca para uma abordagem deste género é aconselhada para reduzir custos e possibilitar uma evolução do nível de transportes em Portugal, a criação de horários interligados entre os diferentes transportes terrestres já existe, não é novidade, mas com machine learning estes seriam aperfeiçoados ao ponto de reduzir custos de manutenção, diminuir tempos de espera para os utilizadores e criar horários mais apetecíveis para os condutores, com intervalos regulados da melhor forma e uma diminuição do numero de regras quebradas pelos horários atuais. Para criar uma cidade metrópole inteligente é fundamental analisar todos os dados de forma inteligente.

# Referencias

1. Florian Toqué, M. K. (2017). *Short & Long Term Forecasting of Multimodal Transport Passenger.* Montréal, Canada.: IEEE.
2. Jahangiri, A. (2015). Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data. *IEEE Transactions on Intelligent Transportation Systems*, 2406 - 2417.
3. João Mendes-Moreira, L. M.-M. (2015). Validating the coverage of bus schedules: A Machine Learning approach. *Information Sciences*, 299-313.