



Universidade do Minho

Departamento de Informática

Mestrado em Engenharia Informática

Mestrado integrado em Engenharia Informática

Perfil de Machine Learning: Fundamentos e Aplicações

Classificadores e Sistemas Conexionistas

1º/4º Ano, 2º Semestre

Ano lectivo 2018/2019

Trabalho Prático - 2ª Parte

Abril, 2019

Tema	Utilização de modelos de <i>Deep Learning</i> para previsão do fluxo de tráfego rodoviário.
Objetivos de aprendizagem	Com a realização deste trabalho prático pretende-se sensibilizar e motivar os alunos para a conceção e desenvolvimento de modelos de <i>Deep Learning</i> nos mais variados domínios, com particular ênfase no que respeita à construção de modelos capazes de tratar séries temporais.
Enunciado	<p>Com o passar dos anos tem se verificado que a segurança rodoviária se tem tornado num crescente ponto de preocupação da nossa sociedade. Uma primeira, e simples, explicação surge do número substancial de acidentes e mortes que se verificam nas estradas, dia após dia. Uma outra explicação poderá fundamentar-se na poluição atmosférica ou no grau de saturação que se verifica em determinadas zonas e ruas de várias cidades.</p> <p>Uma das abordagens seguidas para combater estes problemas prende-se com a previsão do fluxo de tráfego rodoviário. Essa informação poderá depois ser utilizada para informar e auxiliar autoridades, condutores, pedestres e ciclistas. Numa fase inicial, a modelação do fluxo de tráfego era efetuada recorrendo a modelos essencialmente estatísticos. Contudo, tem-se verificado uma crescente adoção de modelos de <i>Deep Learning</i>, em particular Redes Neurais Recorrentes, para previsão do fluxo de tráfego, a curto e longo prazo, com resultados interessantes.</p> <p>No contexto da segunda componente prática de avaliação, pretende-se que os grupos de trabalho concebam e otimizem modelos de <i>Deep Learning</i>, em particular Redes Neurais Recorrentes, para previsão do fluxo de tráfego, a curto e longo prazo. Este trabalho deverá ser encarado como se de um projeto de indústria se tratasse. Nesse sentido, os grupos deverão seguir uma, de várias, metodologias existentes para planeamento e execução de projetos de <i>Machine Learning</i>. Os dados serão disponibilizados em bruto, sendo necessário proceder ao estudo, análise e processamento dos mesmos. Os dados em questão correspondem a:</p> <ul style="list-style-type: none">• Dados climatológicos, descrevendo a temperatura, pressão atmosférica e humidade, entre outros;• Incidentes, tendo como atributos a descrição do incidente (estradas fechadas ou acidentes rodoviários, entre outros), as estradas afetadas, categoria do incidente, magnitude do atraso, comprimento da fila e o atraso provocado, entre outros;• Fluxo do tráfego em certas ruas de uma cidade. Como atributos encontra-se o nome da rua, velocidade atual e velocidade ótima, tempo que se demora a percorrer a rua e tempo ótimo, entre outros.

Serão disponibilizados dados referentes a três diferentes cidades do país: Guimarães (<http://bit.ly/2UKqA4J>), Braga (<http://bit.ly/2D86OGp>) e Porto (<http://bit.ly/2UvEQPg>). Cabe, a cada grupo de trabalho, escolher qual a cidade que pretende trabalhar. Neste trabalho prático, e como em qualquer projeto de *Machine Learning*, o tratamento e processamento dos dados assume especial relevo.

Os modelos a desenvolver deverão ser definidos em função da configuração do problema a tratar, i.e., modelos *single-step* vs *multistep*, modelos univariável vs multivariável e a forma (*shape*) do input. Os modelos poderão variar, por exemplo, entre simples *Multilayer Perceptrons* até *deep LSTMs*. A otimização do modelo deve focar a sua estrutura e os hiper-parâmetros definidos, utilizando, para isso, técnicas como intuição e experimentação, *grid search* ou computação evolucionária (algoritmo genética). A performance do modelo deve ser validada utilizando *k-fold cross-validation*.

Os resultados obtidos deverão ser objeto de um relatório que contenha, entre outros:

- Quais os domínios a tratar, quais os objetivos e como se propõe a atingi-los;
- Qual a metodologia seguida e como foi aplicada;
- Descrição e análise detalhada do *dataset* escolhido e de todo e qualquer tratamento efetuado;
- Descrição da configuração e enquadramento adotado para o problema em causa;
- Descrição detalhada do modelo desenvolvido, quais as suas características, como e sobre que parâmetros foi realizado o *tuning* do modelo, características do treino, entre outros detalhes que seja oportuno fornecer;
- Sumário dos resultados obtidos e respetiva análise crítica.

Todo o processo deverá ser acompanhado de exemplos e indicações que permitam reproduzir todos os passos realizados assim como os resultados obtidos.

Entrega e avaliação

As submissões deverão ser feitas por correio eletrónico para cesar.analide@di.uminho.pt e para bruno.fmf.8@gmail.com, enviando, **num único ficheiro compactado**, todos os elementos produzidos. Tanto o assunto da mensagem como o ficheiro submetido deverão ser identificados na forma “CSC_TP2GX”, em que X designa o número do grupo de trabalho. As submissões deverão ser realizadas até ao dia 19 de maio de 2019.

A sessão de apresentação do trabalho decorrerá no dia 23 de maio de 2019, na sala DI-0.08, tendo início às 14h00min. Cada grupo disporá de 10 minutos para realizar a apresentação, utilizando os meios que considerar mais adequados.

A avaliação deste trabalho de grupo contará com os seguintes elementos:

- Pelo documento produzido, 75%;
- Pela apresentação realizada do trabalho desenvolvido, 25%.

Referências bibliográficas

Para além do material disponibilizado nas aulas aconselha-se também a consulta das seguintes fontes:

- Long Short-Term Memory, Hochreiter and Schmidhuber, 1997.
- Multi-step Time Series Forecasting with Long Short-Term Memory Networks in Python, Jason Brownlee, 2017.
- Intent prediction of vulnerable road users from motion trajectories using stacked LSTM network, Saleh, Hossny and Nahavandi, 2017.
- LSTM-based traffic flow prediction with missing data, Tian et al., 2018.