



Universidade do Minho

Escola de Ciências da Universidade do Minho

Departamento de Informática

Mestrado em Matemática e Computação

Mestrado Integrado em Engenharia Informática

Redes Neurais Recorrentes para previsão do fluxo de tráfego rodoviário

Alunos:

Andreia Costa (PG37013)

Henrique Faria (A82200)

Paulo Barbosa (PG40160)

Rui Teixeira (PG37021)

Docentes:

Bruno Fernandes

Victor Alves

Unidade Curricular: Classificadores e Sistemas Conexionistas

Maio
2020

Conteúdo

1	Introdução	1
2	<i>Dataset</i>	2
2.1	<i>Traffic Flow Braga</i>	2
2.2	<i>Traffic Incidents Braga</i>	3
2.3	<i>Weather Braga Descriptions</i>	3
2.4	<i>Weather Braga</i>	4
2.5	Preparação dos dados	4

1 Introdução

2 *Dataset*

Aquando da apresentação do presente trabalho foram disponibilizados dados referentes a duas cidades: Braga e Porto, sendo que o grupo escolheu os dados relativos à cidade de Braga para trabalhar.

Os dados encontram-se distribuídos em 4 *datasets*:

- *Traffic Flow Braga Until 20191231*;
- *Traffic Incidents Braga Until 20191231*;
- *Weather Braga Descriptions Until 20191231*;
- *Weather Braga Until 20191231*.

Todos os *datasets* contêm dados relativos ao período entre 15 Janeiro 2019 e 31 Dezembro 2019.

2.1 *Traffic Flow Braga*

O *dataset* "Traffic Flow Braga" é constituído pelos seguintes atributos:

- *city_name*;
- *road_num*;
- *road_name*;
- *functional_road_class_desc*;
- *current_speed*;
- *free_flow_speed*;
- *speed_diff*;
- *current_travel_time*;
- *free_flow_travel_time*;
- *time_diff*;
- *creation_date*.

2.2 *Traffic Incidents Braga*

- *city_name*;
- *description*;
- *cause_of_incident*;
- *from_road*;
- *to_road*;
- *affected_roads*;
- *incident_category_desc*;
- *magnitude_of_delay_desc*;
- *length_in_meters*;
- *delay_in_seconds*;
- *incident_date*;
- *latitude*;
- *longitude*.

2.3 *Weather Braga Descriptions*

- *city_name*;
- *cloudiness*;
- *atmosphere*;
- *snow*;
- *thunderstorm*;
- *rain*;
- *sunrise*;
- *sunset*;
- *creation_date*.

2.4 *Weather Braga*

- *city_name*;
- *temperature*;
- *atmospheric_pressure*;
- *humidity*;
- *wind_speed*;
- *clouds*;
- *precipitation*;
- *current_luminosity*;
- *sunrise*;
- *sunset*;
- *creation_date*.

2.5 Preparação dos dados

Após análise dos quatro *datasets* concluiu-se que, antes de se desenvolver o modelo para a previsão DE QQ COISA, era necessário fazer uma prévia preparação dos dados.

Começou-se por fazer o tratamento do *dataset Weather_Descriptions_Braga*, tendo-se removido as colunas: *city_name*, *snow* e *cloudiness*. A coluna *snow* apresentava apenas *missing values*, daí se ter optado pela sua remoção. Relativamente à coluna *cloudiness*, optou-se por fazer a remoção da mesma, uma vez que existe uma coluna que está diretamente relacionada com esta, a coluna *cloud*, e que não apresenta *missing values*.

De seguida, procedeu-se à remoção das colunas *city_name* e *precipitation* do *dataset Weather_Braga*. A remoção da coluna *precipitation* deveu-se ao facto desta apenas apresentar um único valor, o 0.

De modo a unir o resultado da preparação dos dados feita para os *datasets* anteriores, recorreu-se ao nodo *Joiner*, e uniram-se os *datasets* por *creation_date*, tendo-se efetuado, de seguida, a extração da data e do tempo.

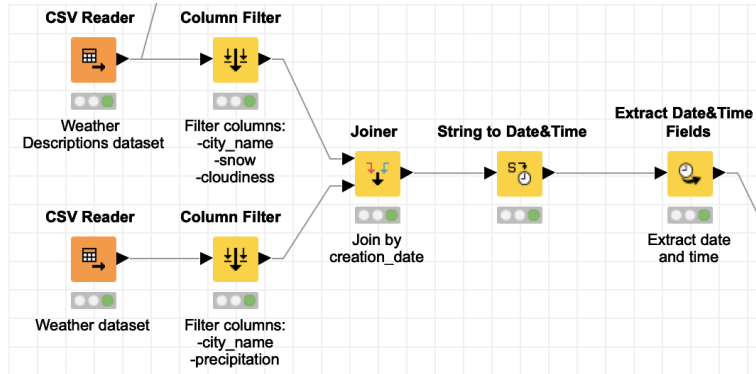


Figura 1: Preparação dos *datasets* *Weather_Descriptions_Braga* e *Weather_Braga*.

De seguida, procedeu-se à preparação do *dataset* *Traffic_Flow_Braga*, procedendo-se à remoção das colunas *city_name* e *road_name*, seguida da extração da data e hora e agrupamento dos dados por *road_num*, hora, dia do mês e mês. De modo a juntar este *dataset* ao obtido anteriormente, recorreu-se ao nodo *Joiner*, unindo-se os *datasets* por hora, dia do mês e mês.

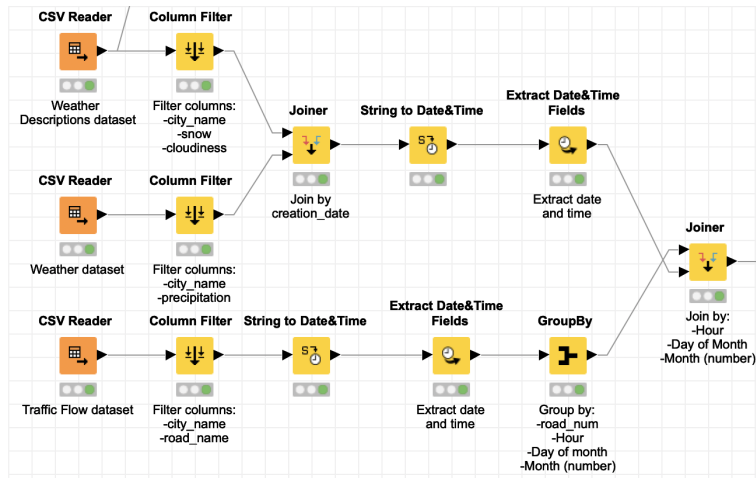


Figura 2: Preparação do *dataset* *Traffic_Flow_Braga*.

Após a junção dos *datasets*, eliminou-se a coluna *creation_date* e transformaram-se os valores "N/A", das colunas *rain*, *thunderstorm* e *atmosphere*, em *missing values*, recorrendo ao nodo *String Manipulation*. De seguida, fez-se um *merge* das colunas *rain* e *thunderstorm*, tendo-se alterado alguns dos valores ("trovoada com chuva fraca-; "chuva fraca", "trovoada com chuva forte-; "chuva forte" e "trovoada-; "chuva"), tendo-se removido, no final, a coluna

thunderstorm. Por fim, com os valores da coluna *clouds* construíram-se 6 intervalos:

1. céu claro: $] - \text{inf}, 17[$;
2. céu pouco nublado: $[17, 34[$;
3. nuvens dispersas: $[34, 51[$;
4. nuvens quebradas: $[51, 68[$;
5. nublado: $[68, 85[$;
6. muito nublado: $[85, + \text{inf}[$;

Por fim, tratou-se o *dataset Traffic_Incidents_Braga*, procedendo-se à extração do dia e da hora e removendo-se as colunas *cause_of_accidents* e *affected_roads*. Após tratado este *dataset*, e recorrendo ao nodo *Joiner*, uniu-se este *dataset* com o obtido anteriormente por hora, dia do mês e mês. Deste modo, uniram-se os 4 *datasets* iniciais num único.

Após se ter apenas um *dataset*, e tendo em conta que as colunas *atmosphere* e *rain* apresentam muitos *missing values*, procedeu-se ao tratamento dos mesmos.

Começou-se, então, por tratar os *missing values* da coluna *atmosphere*, tendo-se separado o *dataset* em dois, recorrendo ao nodo *Rule-based Row Splitter*. Um *dataset* apresenta a coluna *atmosphere* apenas com *missing values* e o outro apresenta a coluna *atmosphere* com os vários valores. De seguida, utilizaram-se *Random Forest* para fazer a previsão dos *missing values*, para isso, particionou-se o *dataset* que apresentava os valores do atributo *atmosphere* tendo-se usado 80% dos dados para treino. Após feita a previsão dos *missing values*, procedeu-se à previsão dos *missing values* do atributo *rain*, tendo-se utilizado o mesmo raciocínio.

Python
