



SEMM

Sistema de Emergência Médica

Apresentação

Primeira parte

Descrição do problema de análise dos dados

O Sistema de Emergência Médica (*SEMM*) é um trabalho que consiste na representação de múltiplos hospitais e na distribuição de vários pacientes por estes através de diversas ambulâncias. Cada ambulância contém uma equipa médica que deverá ser atribuída tendo em conta a distância a que se encontram do paciente e as necessidades do mesmo. O paciente, após ter sido colocado na ambulância correta, deverá ser encaminhado para o hospital para receber tratamento. Este é selecionado com base na gravidade do ferimento, na distância do paciente até ao hospital e nas especialidades que este contém.

O programa foi subdividido em duas partes para uma melhor abordagem dos problemas:

1. Tendo em conta as necessidades do paciente, a especialidade da ambulância, a distância paciente-ambulância e a disponibilidade das ambulâncias. Será que foi escolhida a ambulância adequada? (Classificação)
2. Tendo em conta as necessidades do paciente, a especialidade do hospital e a distância. Será que foi escolhido o hospital adequado? (Classificação)

Experiências realizadas

Experiência 1

Objetivo: encontrar a ambulância adequada;

Variáveis independentes:

- **Illness:** (polinomial), corresponde à doença do paciente;
- **Ambulance** (polinomial), **Specialty:** (polinomial), corresponde à especialidade da ambulância;
- **Distance A->P:** (Integer), corresponde à distância entre a ambulância e o paciente;
- **Availability:** (Boolean), corresponde à disponibilidade da ambulância;

Variável Dependente:

- **Accepted** (Boolean), indica se a ambulância foi aceite ou não.

Experiência 2

Objetivo: encontrar o hospital adequado;

Variáveis independentes:

- **Illness:** (polinomial), corresponde à doença do paciente;
- **Hospital Specialty** (polinomial), corresponde à especialidade do hospital
- **Distance** (Integer), corresponde à distância entre o hospital e o paciente.

Variável Dependente:

- **Accepted** (Boolean), indica se o hospital foi aceite ou não.

Estatísticas sobre os dados realizados

De forma a encontrar estatísticas adequadas aos problemas a analisar, os dados que recolhemos foram guardados num ficheiro csv. Cada linha corresponde a:

- **Paciente:** id, doença e posição;
- **Ambulância:** id, posição, distância ao paciente, especialidade, disponibilidade e aceitação;
- **Hospital:** id, posição, distância ao paciente, especialidade e aceitação.

column index	attribute meta data information			
0	ID Patient	column ...	polyno...	attribute ▼
1	Illness	column ...	polyno...	attribute ▼
2	Patient X	column ...	integer	attribute ▼
3	Patient Y	column ...	integer	attribute ▼
4	ID Ambulanc	column ...	polyno...	attribute ▼
5	Ambulance S	column ...	polyno...	attribute ▼
6	Ambulance X	column ...	integer	attribute ▼
7	Ambulance Y	column ...	integer	attribute ▼
8	Distance A→f	column ...	integer	attribute ▼
9	Availability	column ...	polyno...	attribute ▼
10	Accepted	column ...	polyno...	attribute ▼
11	Hospital ID	column ...	polyno...	attribute ▼
12	Hospital X	column ...	integer	attribute ▼
13	Hospital Y	column ...	integer	attribute ▼
14	Hospital Spec	column ...	polyno...	attribute ▼
15	Distance	column ...	integer	attribute ▼
16	Chosen Hospi	column ...	polyno...	attribute ▼

Fig. 1 - Estrutura de uma linha do ficheiro csv gerado (Rapidminer)

Estatísticas sobre os dados realizados (Prob. 1)



Fig. 2 - Estatística geradas pelo ficheiro csv para Ambulância (Rapidminer)

Estatísticas sobre os dados realizados (Prob. 2)



Fig. 3 - Estatística geradas pelo ficheiro csv para Hospital (Rapidminer)

Análise dos dados com RapidMiner (Prob. 1)

A maior percentagem de erros é observada nas variações da distância entre o paciente e a ambulância. Prevê-se que os dados relativos a esta variável sejam aproximadamente semelhantes. No entanto, quando acontece um desfasamento entre a posição do paciente e da ambulância num caso pontual, gera-se um erro. Consideremos que a média das distâncias está compreendida entre 5 e 20 (metros). Como o programa prioriza que a especialidade da ambulância coincida com as necessidades do paciente, se a única ambulância disponível com a especialidade pretendida estiver a 40 (metros) de distância existe uma grande diferença face à maioria dos dados.

Conclui-se então que a percentagem de erros estará fortemente relacionada com a disposição dos dados que tratam da distância entre o paciente e a ambulância.

accuracy: 98.87%

	true No	true Yes	class precision
pred. No	2959	23	99.23%
pred. Yes	11	7	38.89%
class recall	99.63%	23.33%	

Fig. 4 - Performance vect de Ambulância

Análise dos dados com RapidMiner (Prob. 2)

Neste caso, os resultados que se observam são causados pelo efeito falado no slide anterior, uma maior disparidade de valores de distância causará uma menor percentagem de acerto. No entanto, como os hospitais não possuem disponibilidade, assume-se que estão sempre disponíveis. Assim, havendo um menor número de variáveis independentes, a percentagem de erro também será menor.

accuracy: 99.20%

	true No	true Yes	class precision
pred. No	2967	21	99.30%
pred. Yes	3	9	75.00%
class recall	99.90%	30.00%	

Fig. 5 - Performance vect de Hospital

Conclusão

Depois de analisar os dados e analisar os resultados inferidos através do RapidMiner, podemos concluir que a precisão (*accuracy*) é bastante elevada, uma vez que temos poucas variáveis independentes nas duas experiências realizadas.

No final deste trabalho, concluimos que teria sido interessante a existência de um maior número de dados relativos a um agente e uma maior complexidade na forma como os agentes se comportam. A abrangência do tema causou no grupo alguma incerteza sobre os dados a inferir, fazendo com que o produto final fosse mais simples do que o inicialmente idealizado.



Informação Adicional

Segunda parte

Processos RapidMiner

Em ambos os processos realizados foi utilizado o operador **Split Validation**, visto os problemas que abordamos com este trabalho eram de classificação (escolher ambulância e/ou hospital adequado/a).

Proporção utilizada:

- 70% dos casos para treino
- 30% dos casos para teste

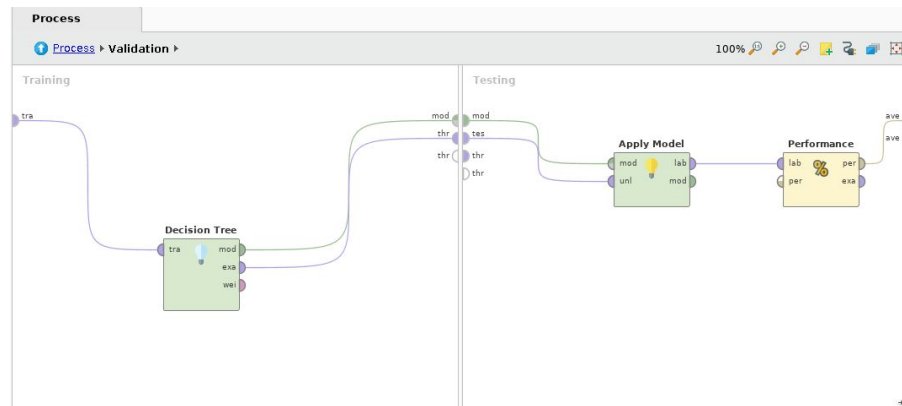


Fig. 6 - Design no *RapidMiner*

Resultados de outras experiências interessantes

O grupo achou interessante criar um novo problema:

3. Como é que a necessidade do paciente e a especialidade da ambulância, afeta o tempo (ou distância) de socorro da ambulância? (Regressão)

Experiência 3

Objetivo: encontrar a ambulância adequada;

Variáveis independentes:

- **Illness:** (polinomial), corresponde à doença do paciente;
- **Ambulance** (polinomial), corresponde à especialidade da ambulância;
- **Availability:** (Boolean), corresponde à disponibilidade da ambulância;
- **Accepted** (Boolean), indica se a ambulância foi aceite ou não.

Variável Dependente:

- **Distance A->P:** (Integer), corresponde à distância entre a ambulância e o paciente;

Como foram utilizadas poucas variáveis independentes, o resultado indica um valor de **root_relative_squared_error** elevado e uma **correlation** mais reduzida do que o pretendido.

root_relative_squared_error

root_relative_squared_error: 0.888

root_mean_squared_error

root_mean_squared_error: 22.164 +/- 0.000

correlation

correlation: 0.460

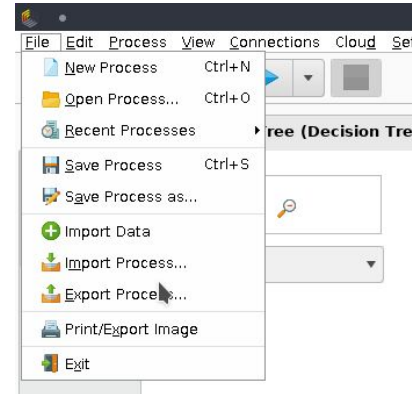
Outras Observações

Em anexo, vão dois ficheiros com os processos utilizados para este trabalho:

- *Ambulance.rmp*;
- *Hospital.rmp*;
- *data.csv* (ficheiro csv com os dados gerados).

Para ser possível testar os processos desenvolvidos, é preciso:

- Abrir o **RapidMiner Studio™**;
- No menu superior escolher **File->Import Process**;



- Escolher o ficheiro pretendido

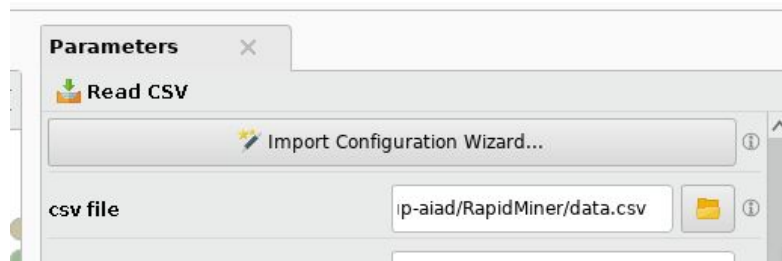
Outras Observações

Após ser importado o ficheiro *rmf*, é necessário dar *setup* à leitura do ficheiro csv:

- Clique em **Read CSV**



- No menu lateral, em **Parameters**, escolher **Import Configuration Wizard**



- Escolher o ficheiro csv pretendido (*data.csv*)

Obrigado/a!



Beatriz de Henriques Martins | up201502858

José Pedro da Silva e Sousa Borges | up201503603

Luís Miguel Cardoso Lopes Correia | up201503342

Agentes e Inteligência Artificial Distribuída
4º Ano - 1º Semestre
16 de dezembro de 2018