

Curso de Engenharia de Sistemas e Informática - 5º Ano Análise Inteligente de Dados Ficha de Trabalho N.º 12

Objectivo: Estudo das Redes Kohonen para segmentação e das Regras de Associação para efectuar análise de cesto de compras utilizando o Nó APRIORI.

I – Redes Kohonen.

Tentaremos segmentar um data set que contém informação acerca dos participantes em cursos SPSS. O fich. de dados UKTraining.txt contém informações relativas ao tipo de cursos que 2000 indivíduos frequentaram. Este ficheiro deve ser descarregado do anexo a esta ficha para a área de trabalho do aluno.

- 1. Comece com um painel de Stream vazio.
- Coloque um Nó Var.File no painel Stream e faça-o ler o fich. texto UKTraining.txt com campos separados por tabs. Este ficheiro contém os nomes dos campos na 1.ª linha.
- 3. Ligue um nó Type ao nó source já existente e edite-o.
- 4. Altere para simbólicos todos os campos fazendo click botão direito no cabeçalho da coluna Type e seleccionando Auto Symbol.
- 5. Altere o tipo dos seguintes campos de acordo com a lista:

Total Spend	Auto Real
Number of Courses	Auto Real
Id	Typeless

- Coloque em NONE a direcção para os campos: Id, Privilege Card, Sector, Mini Subscription, Total Spend, Number of Courses, Venue, Subscription.
- 7. Ligue uma Tabela e execute o stream. Verifique se o tipo dos campos ficaram correctamente definidos.
- Ligue um nó Train Kohonen ao nó Type e dite o nó Kohonen. Verifique a caixa das Opções p/ Perito e altere a largura e comprimento para 3. Execute esta parte do stream.
- 9. Uma vez terminado o treino da rede, o modelo surge na paleta dos modelos gerados. Coloque-o no painel Sream e ligue-o ao nó Type.
- 10. Ligue um nó Table ao nó Knet gerado e e execute essa parte do stream. Interprete o significado dos dois novos campos.
- 11. Ligue um nó Derive que permita criar um novo campo de nome Cluster, contendo uma concatenação dos dois campos a que se aludiu na questão anterior, que deverá verificar utilizando um novo nó table.
- 12. Inspeccione a distribuição de registos pelos diversos clusters gerados e interprete os resultados.
- 13. Crie um novo nó Knet, mas desta vez com 16 clusters possíveis. Compare resultados com o nó Knet anterior e tire conclusões.
- 14. Para visualizar espacialmente a distribuição dos registos relativos ao diversos clusters gerados, ligue um nó Plot ao nó Knet gerado em 9 e edite-o. Faça o gráfico dos campos \$KX-Kohonen versus \$KY-Kohonen, utilizando agitação de 0.3 em ambos

os eixos.

- 15. O que visualizou? Repita o gráfico, mas desta vez sem agitação. Efectuando uma análise comparativa entre os dois gráficos Plot, interprete a funcionalidade da agitação de 0.3 introduzida?
- 16. Crie também um gráfico Plot relativo ao modelo de 16 clusters e compare com o gráfico da questão anterior. Conformam-se as conclusões a que chegou na resposta à questão 13? Há outros elementos que possam trazer mais algum conteúdo explicativo? Enumere-os em caso afirmativo.
- 17. No gráfico (da questão 14), marque rectângulos à volta dos clusters mais relevantes. Crie um **Nó Derive Set**.
- 18. Ligue o nó derive gerado (chamado Region) ao nó do modelo gerado.
- 19. Coloque um nó table e interprete o conteúdo do campo derivado agora gerado.
- 20. Ligue um nó Web ao nó derive e produza um gráfico do campo derivado (região) e os quatro cursos introdutórios: Introdução ao SPSS Introdução ao SPSS e Estatística Introdução à Estatística FastTrack
- 21. Verifique que a caixa **Use true values only for flags** está ligada. Coloque **Thresholds em overall percentages**. Coloque **connection limits** em 1%, 2% e 3%, respectivamente.
- 22. Execute o gráfico Web. Há alguns padrões relativos aos clusters?
- 23. Produza um gráfico Web semelhante, com os mesmos parâmetros, utilizando os campos seguintes:

Region Mini Subscription or 5 day deal Subscription Previlege card

- 24. Há mais alguns padrões a salientar?
- 25. Guarde uma cópia do stream criado com o nome ficha9.str.

II - Regras de Associação.

Tentaremos fazer uma análise de cesto de compras utilizando os dados do exercício anterior, para verificar que cursos induzem outros cursos ou são adquiridos em conjunto.

Uma regra de associação apresenta-se sob a forma de Condição1 & Condição2 & ... & Condição n => Conclusão ou de forma inversa Conclusão <= Condição1 & Condição2 & ... & Condição n. Independentemente da ordem, temos uma ou mais condições (antecedente) e uma conclusão (consequente). Um exemplo de regra será Combustível & Chocolate => Jornais, que interpretada significará algo como "Indivíduos que compram combustível e chocolate, também poderão adquirir chocolate". Claro que esta regra poderá ter um campo de aplicabilidade mais ou menos alargado e ser mais ou menos precisa. Daí que seja importante incluir um conjunto de medidas, que permitam aferir do interesse de uma dada regra.

O Clementine, quando produz um conjunto de regras de associação, também dá medidas que indicam a força da associação para cada regra. Essas medidas são denominadas de cobertura e precisão, apresentando-se assim as regras como

Conclusão <= Condição 1 & ... Condição n (Número de registos: Cobertura, Precisão).

1. Carregue o stream guardado no exercício anterior, se não estiver ainda no painel stream.

- 2. Edite o nó Type.
- A maioria dos valores dos parâmetros podem ser mantidos tal como foram utilizados no exercício anterior, menos as direcções dos campos do curso, que devem ser alterados para BOTH, de forma a que possam aparecer em regras quer como condições quer em conclusões, menos os campos *Introdução ao SPSS*

Introdução ao SPSS e Estatística FastTrack,

Que devem ser postos a IN (estamos interessados em saber que cursos induzem outros, então, assumindo que a 1.ª participação num curso seja em um curso introdutório, estes campos actuarão apenas como condições).

- 4. Ligue um nó APRIORI ao nó Type e edite o nó APRIORI.
- Ponha o parâmetro Minimum Coverage (cobertura mínima) em 1% e o Minimum Accuracy (precisão mínima) em 50%. Coloque em true o Only True values for flags. Execute esta parte do stream.
- 6. Deve ter-se apercebido que o aspecto gráfico do modelo gerado é algo diferente dos símbolos dos outros modelos gerados anteriormente. È mostrado apenas parte do diamante / pepita, estando o resto ainda rodeado de pedra. Porquê esta alegoria? Parece que o modelo não está completamente refinado...
- 7. **Inspeccione** o modelo não refinado resultante e ordene as regras por confiança e cobertura.
- 8. Que pode concluir das regas geradas? Essas regras fazem sentido?
- 9. Suponha agora que pretendia utilizar as regras relativas a uma dada conclusão ex. frequentar curso "Criar Modelos Preditivos". Actue no sentido de o poder fazer, criando um modelo refinado dessas regras.
- 10. Inspeccione o modelo gerado, interpretando as regras geradas e respectiva cobertura e confiança.
- 11. Coloque esse no painel stream e ligue-lhe um nó table. Execute essa parte do stream e interprete os novos campos gerados, verificando se correspondem às conclusões chegadas na questão anterior.