

UNIVERSIDADE DO MINHO

Exercício nº 3

**Conhecimento não simbólico: Redes
Neuronais Artificiais**

Mestrado Integrado em Engenharia Informática

Sistemas de Representação de Conhecimento e
Raciocínio

Ano Letivo 2016/2017

2º Semestre

A73674 – Alexandre Lopes Mandim da Silva

A74219 – Hugo Alves Carvalho

A74260 – Luís Miguel da Cunha Lima

Braga,

21 de Maio de 2017

Resumo

A componente prática da unidade curricular de Sistemas de Representação de Conhecimento e Raciocínio, inserida no 3º ano do plano curricular do Mestrado Integrado em Engenharia Informática pressupõem a realização de trabalhos de grupo na forma de um conjunto de exercícios. Este será, portanto, o terceiro dos exercícios práticos, tendo por principal tema “Conhecimento não Simbólico: Redes Neurais Artificiais” com o qual se pretende abordar o problema de deteção de níveis de exaustão através de dados recolhidos da interação humano-computador.

Para o efeito, foi desenvolvido um estudo elucidativo deste panorama que, ao longo deste relatório, iremos analisar e explicar todas as decisões e soluções indicadas aos desafios propostos no enunciado deste terceiro exercício prático.

Índice

1.	Introdução	4
2.	Preliminares	5
3.	Descrição do Trabalho	6
3.1.	Normalização de Dados	6
3.2.	Atributos mais significativos	6
3.3.	Fórmulas	7
3.4.	Sets	8
3.5.	Análise dos níveis de Fadiga com 7 níveis	8
3.6.	Análise da tarefa em execução	9
3.7.	Identificar existência e ausência de exaustão	9
3.8.	Melhor escala de identificação de Exaustão	10
4.	Conclusões e Sugestões	12
5.	Bibliografia	13

1. Introdução

O principal objetivo deste exercício é aprofundar e refinar os conceitos do grupo sobre conhecimento não simbólico na representação de conhecimento e no desenvolvimento de mecanismos de raciocínio, utilizando Redes Neurais Artificiais (RNAs) para a resolução de problemas.

O caso de estudo deste projeto, está relacionado com a biometria como a ciência que estuda as características físicas e/ou comportamentais dos seres vivos. Esta tecnologia de informação refere-se a tecnologias que avaliam e examinam as características do corpo humano, tais como, tamanho da mão, impressão digital, forma de escrever no teclado e forma do olhar. Neste trabalho focaremos as características comportamentais pois estas permitem uma análise de padrões comportamentais de forma não-intrusiva e, até mesmo, invisível e desconhecida para o utilizador. Para isso foi fornecido um conjunto de dados recolhidos da interação humano-computador através dos dispositivos físicos rato e teclado para a realização de testes.

2. Preliminares

De forma a conseguirmos concretizar o enunciado proposto foi essencial, através das aulas desta Unidade Curricular, compreender conceitos teóricos e métodos de aplicação dos mesmos, onde fazemos aqui uma pequena introdução teórica.

As Redes Neurais Artificiais, são modelos simplificados do sistema nervoso central do ser humano. Trata-se de uma estrutura extremamente interconectada de unidades computacionais, designadas por neurónios, com capacidade de aprendizagem. Assemelha-se ao comportamento do cérebro em vários aspetos:

- O conhecimento é adquirido a partir de um ambiente, através de um processo de aprendizagem
- O conhecimento é armazenado nas conexões, também designadas por ligações ou sinapses, entre neurónios.

Baseia-se na conexão de unidades de processamento em que a sua rede neuronal é constituída por:

- Neurónio: unidades de processamento, responsável por coletar e combinar informação vindas de outros neurónios;
- Dendrite: tem como função, receber os estímulos transmitidos pelos outros neurónios;
- Axónio: responsável pela transferência de informação para outras unidades de processamento.

A rede neuronal admite vários parâmetros de um acontecimento como input e, faz esta informação percorrer a sua rede até que é retornado um ou mais valor/valores de output.

A aprendizagem da rede é definida pela regra de transferência que a rede neuronal implementa, isto será outro parâmetro que decidirá o funcionamento da rede. O cálculo do valor de ativação dos neurónios é afetado pela informação que chega aos mesmos, pelos dados de input e pelo valor de ativação anterior (armazenado em memória).

3. Descrição do Trabalho

3.1. Normalização de Dados

Para a realização deste trabalho foi-nos fornecido um conjunto de dados biométricos recolhidos da interação humano-computador através dos dispositivos físicos rato e teclado, que são usados para a deteção de exaustão. Estes dados estão organizados numa tabela como a que apresentamos em seguida:

Performance.KDTMean	Performance.MAMean	Performance.MVMean	Performance.TBCMean	Performance.DDCMean	Performance.DMSMean	Performance.AEDMean	Performance.ADMSLMean	ExhaustionL	Performance.Task
KDT	MA	MV	TBC	DDC	DMS	AED	ADMSL	Exaustão	Tarefa
0,685482821	0,036346519	0,041677644	-0,085541958	0,099002914	5,8186E-05	-0,007109841	0,016009068	4	Work
0,021535506	0,038227324	0,042273998	-0,054892	-0,016446983	0,002066559	0,01371244	0,058274226	4	Work
0,001741047	0,046606792	0,047566659	-0,09511432	0,523803686	0,0082762	-0,023689567	0,11327566	5	Work
-0,000330925	0,052205249	0,058021096	-0,081694699	0,00776611	-0,010530433	0,022886232	-0,011112783	5	Work
0,007564058	0,077287098	0,080671542	-0,078175055	0,047051655	0,018521599	0,060834655	0,08417121	5	Work
0,002186559	0,031077892	0,0359264	-0,100597523	-0,008544334	-0,04993608	-0,020878154	0	5	Work
0,002486693	-0,029336934	-0,019338111	-0,098791685	-0,060463461	-0,007040177	0,046152375	-0,036877204	5	Work
0,074282284	0,017823444	0,031911325	-0,063129499	-0,045115742	-0,000161297	-0,030760794	-0,00109404	3	Work
0,000795447	0,008310959	0,024699565	-0,105763821	0,02788493	-0,008940139	0,027590013	-0,031462857	3	Work
0,016142811	-0,034599702	-0,013872403	-0,077475062	0,085032005	0,002159164	-0,025269483	-0,011187123	3	Work
-0,001215787	-0,029261809	-0,023433191	-0,104800004	0,055173689	-0,006416741	-0,026214323	-0,020215798	4	Work
0,000968846	0,037135449	0,048592672	-0,064033004	-0,054109249	-0,005066853	-0,002647724	-0,016323122	3	Work
-0,000205899	0,030628131	0,031298191	-0,104903055	-0,005017731	0,020073438	-0,01929643	0,049802605	4	Work

Figura 1 - Excerto de Dados fornecidos

Os valores apresentados nas oito primeiras colunas estão normalizados no intervalo $[-1,1]$, enquanto que o nível de exaustão está no intervalo entre $[1,7]$ em cada um tem um significado próprio. A última coluna, correspondente ao *Performance Task*, pode assumir três valores: “Work”, “Office” e “Programming”. Para conseguirmos resolver o problema com o uso de redes neuronais da melhor maneira possível, foi necessário normalizar esta coluna, atribuindo valores inteiros a cada uma das *tags*. Deste modo, “work” passou a ser identificado pelo número 1, “office” pelo número 2 e “programming” pelo número 3.

Por outro lado, e após esta modificação do ficheiro de dados, chegou-se à conclusão que as *tags* de performance não estavam distribuídas de igual forma ao longo do ficheiro de dados, o que mais tarde poderia trazer problemas na identificação dos casos de treino e testes para as redes neuronais artificiais. Deste modo, para evitar esta situação, decidiu-se “baralhar” as linhas após estas serem lidas:

```
dados <- read.csv("C:\\Users\\hugoc\\Dropbox\\Trabalho SRCR\\3\\exaustao.csv", header=TRUE, sep=";", dec=".")
#"baralhar" as linhas lidas
dados <- dados[sample(nrow(dados)), ]
```

Figura 2 - Leitura de Dados

3.2. Atributos mais significativos

De modo a identificar os atributos mais significativos para a representação do conhecimento do problema em análise, decidiu-se utilizar o programa “Weka” onde recolhemos esta informação para os níveis de exaustão e para a tarefa em execução.

Assim, como podemos verificar nas imagens seguintes, obtivemos os seguintes resultados:

- Exaustão: KDT Mean, MA Mean, DDC Mean e Performance Task

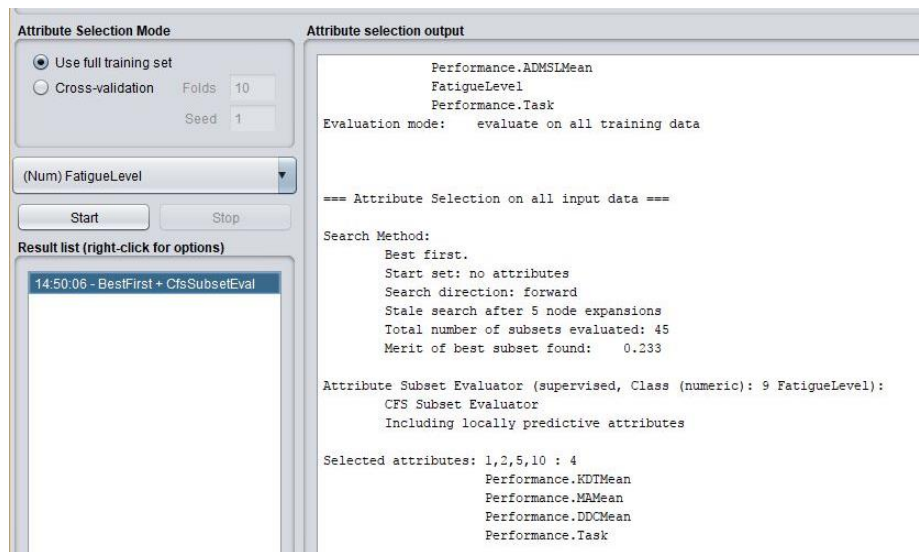


Figura 3 - Printscreen do programa "Weka" para análise dos atributos mais significativos da Exaustão

- Performance Task: KDT Mean, DDC Mean, DMS Mean, AED Mean, ADMSL Mean e Fatigue Level (Exaustão)

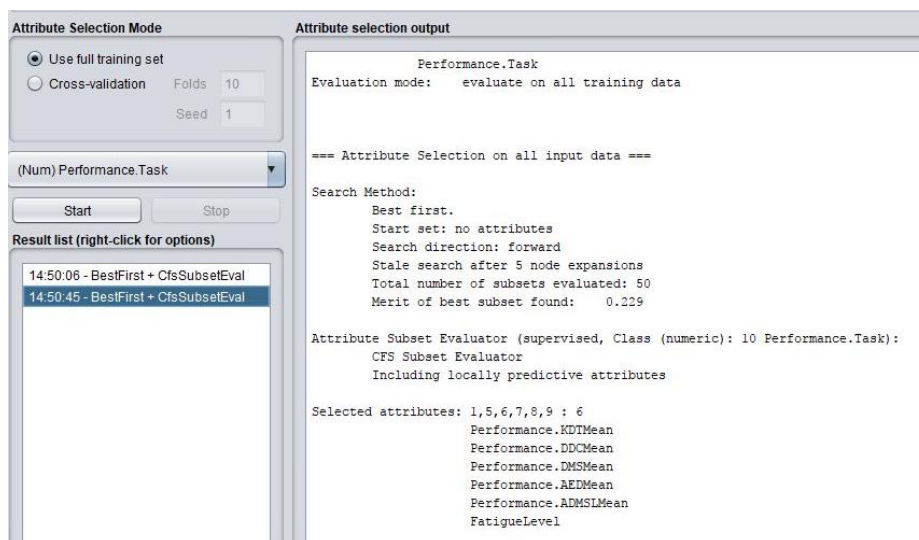


Figura 4 - Printscreen do programa "Weka" para análise dos atributos mais significativos da Performance Task

3.3. Fórmulas

Após identificados os atributos mais significativos para cada um dos elementos que pretendemos analisar, prosseguiu-se com o desenvolvimento das fórmulas que serão passadas à rede neuronal. Para isso, com base nos resultados obtidos no subtópico descrito

anteriormente, o grupo optou por utilizar duas fórmulas, uma para níveis de Fadiga e outra para a tarefa em execução.

```
# formula da fadiga
formulaFadiga <- FatigueLevel ~ Performance.KDTMean + Performance.MAMean + Performance.DDCMean + Performance.Task
```

Figura 5 - Fórmula utilizada para os níveis de Fadiga

```
# formula da tarefa em execução
formulaPerformanceTask <- Performance.Task ~ Performance.KDTMean + Performance.DDCMean + Performance.DMSMean +
Performance.AEDMean + Performance.ADMSLMean + FatigueLevel
```

Figura 6 - Fórmula utilizada para a tarefa em execução

3.4. Sets

Para a realização de testes após treinar a rede, foi necessário definir *sets* em que só estejam presentes as colunas dos atributos que são utilizadas na respetiva fórmula. Deste modo, foram criados dois *subsets*:

```
#subset para treino fadiga
teste.01 <- subset(teste, select = c("Performance.MAMean", "Performance.KDTMean", "Performance.DDCMean", "Performance.Task"))
```

Figura 7 - Subset usado para treino da fórmula da fadiga

```
#subset para tarefa em execução
teste.01 <- subset(teste, select = c("Performance.KDTMean", "Performance.DDCMean", "Performance.DMSMean",
"Performance.AEDMean", "Performance.ADMSLMean", "FatigueLevel"))
```

Figura 8 - Subset usado para o treino da fórmula da tarefa em execução

3.5. Análise dos níveis de Fadiga com 7 níveis

Neste subtópico serão demonstrados os resultados que obtivemos para a fórmula da fadiga, bem como respetivas alterações que fizemos para tentar obter um melhor resultado.

Rede Neuronal	Intervalo de Treino	Intervalo de Testes	Threshold	RMSE
(6,4)	[1:600,]	[601:844,]	0.2	1.1975
(6,4)	[1:600,]	[601:844,]	0.1	1.1645
(6,4)	[1:700,]	[701:844,]	0.1	1.0630
(6,4)	[1:700,]	[701:844,]	0.05	1.0793
(6,4)	[1:700,]	[701:844,]	0.01	-----
(10,8)	[1:600,]	[601,844,]	0.2	1.4031
(10,8)	[1:600,]	[601,844,]	0.1	1.4859
(10,8)	[1:700,]	[701,844,]	0.1	1.3935

(10,8)	[1:700,]	[701,844,]	0.05	1.5426
(6,4,2)	[1:700,]	[701,844,]	0.1	1.1542

Tabela 1 - Resultados com a fórmula da Fadiga em 7 níveis

3.6. Análise da tarefa em execução

À semelhança do subtópico anterior, serão demonstrados os resultados que obtivemos para a fórmula da tarefa em execução, bem como respetivas alterações que fizemos para tentar obter um melhor resultado.

Rede Neuronal	Intervalo de Treino	Intervalo de Testes	Threshold	RMSE
(6,4)	[1:600,]	[601:844,]	0.2	1.5081
(6,4)	[1:600,]	[601:844,]	0.1	1.5190
(6,4)	[1:700,]	[701:844,]	0.1	1.4719
(6,4)	[1:700,]	[701:844,]	0.05	1.5478
(6,4)	[1:700,]	[701:844,]	0.01	1.3743
(10,8)	[1:600,]	[601:844,]	0.2	1.5901
(10,8)	[1:600,]	[601:844,]	0.1	1.5176
(10,8)	[1:700,]	[701:844,]	0.1	1.6180
(10,8)	[1:700,]	[701:844,]	0.05	1.5833
(6,4,2)	[1:700,]	[701:844,]	0.1	1.5679

Tabela 2 - Resultados com a fórmula da tarefa em execução

3.7. Identificar existência e ausência de exaustão

Nesta parte do trabalho prático é pretendido que se altere a estala dos níveis de fadiga, tendo como objetivo identificar, simplesmente, a existência ou ausência de fadiga. Assim, a escala anterior que envolvia um conjunto de 7 valores, passa agora a ter apenas 2 valores:

- 1 – Existência de Fadiga
- 2 – Ausência de Fadiga

Para conseguirmos alterar a escala, foi necessário analisar os níveis anteriores e compreender quais correspondiam a cada um dos novos casos. Deste modo, foi considerado que os níveis 1, 2 e 3 correspondem a ausência de fadiga, enquanto que os níveis 4, 5, 6 e 7 correspondem à existência de fadiga.

```
dados$FatigueLevel[dados$FatigueLevel <= 3] <- 1
dados$FatigueLevel[dados$FatigueLevel > 3] <- 2
```

Figura 9 - Código para alteração da escala

Após esta alteração da escala, realizamos então uma análise semelhante às descritas nos subtópicos anteriores. Em seguida apresentamos os resultados verificados com esta alteração:

Rede Neuronal	Intervalo de Treino	Intervalo de Testes	Threshold	RMSE
(6,4)	[1:600,]	[601:844,]	0.2	0.3883
(6,4)	[1:600,]	[601:844,]	0.1	0.3944
(6,4)	[1:600,]	[601:844,]	0.05	0.4184
(6,4)	[1:700,]	[701:844,]	0.1	0.3576
(6,4)	[1:700,]	[701:844,]	0.05	0.3998
(6,4)	[1:700,]	[701:844,]	0.01	0.4321
(10,8)	[1:600,]	[601,844,]	0.2	0.4341
(10,8)	[1:600,]	[601,844,]	0.1	0.5867
(10,8)	[1:700,]	[701,844,]	0.1	0.6009
(10,8)	[1:700,]	[701,844,]	0.05	0.5714
(6,4,2)	[1:700,]	[701,844,]	0.1	0.4859

Tabela 3 - Resultados com a fórmula da Fadiga em 2 níveis

3.8. Melhor escala de identificação de Exaustão

Nesta parte do trabalho, o objetivo passou por tentar encontrar uma melhor escala para representar os níveis de exaustão. Deste modo, o primeiro passo foi analisar o ficheiro de dados fornecido e compreender como estavam distribuídos os níveis de exaustão. Para isso, recorreu-se novamente ao programa “Weka” onde obtivemos um gráfico com esta mesma distribuição.

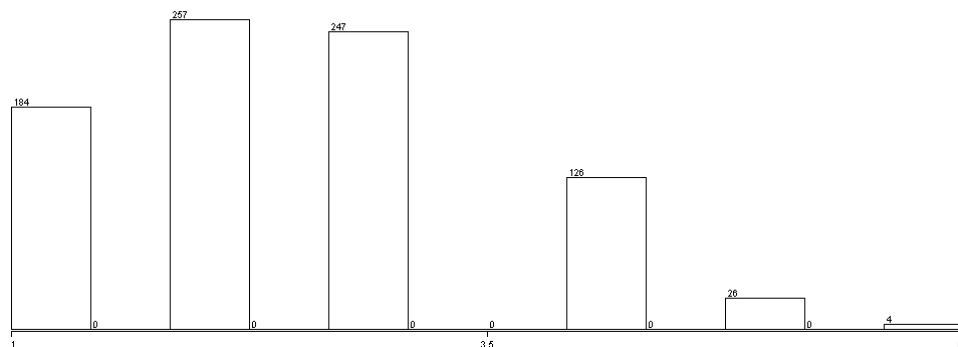


Figura 10 - Gráfico com a distribuição dos níveis de exaustão

Como podemos verificar no gráfico, os níveis 4, 5, 6 surgem com menor frequência, enquanto que o nível 7 nunca surge em nenhum registo. Assim, decidiu-se agrupar estes 4 níveis num só, elaborando uma nova escala com 4 níveis, obtendo assim uma melhor distribuição de valores.

```
dados$FatigueLevel[dados$FatigueLevel = 1 ] <- 1
dados$FatigueLevel[dados$FatigueLevel = 2 ] <- 2
dados$FatigueLevel[dados$FatigueLevel = 3 ] <- 3
dados$FatigueLevel[dados$FatigueLevel >= 4] <- 4
```

Figura 11 - Código para alteração para a melhor escala

Com esta alteração, procedeu-se a uma nova análise de resultados, que é demonstrada em seguida:

Rede Neuronal	Intervalo de Treino	Intervalo de Testes	Threshold	RMSE
(6,4)	[1:600,]	[601:844,]	0.2	1.1908
(6,4)	[1:600,]	[601:844,]	0.1	1.1629
(6,4)	[1:600,]	[601:844,]	0.05	1.1959
(6,4)	[1:700,]	[701:844,]	0.1	1.2190
(6,4)	[1:700,]	[701:844,]	0.05	1.2133
(6,4)	[1:700,]	[701:844,]	0.01	-----
(10,8)	[1:600,]	[601,844,]	0.2	1.6730
(10,8)	[1:600,]	[601,844,]	0.1	1.4329
(10,8)	[1:700,]	[701,844,]	0.1	1.4233
(6,4,2)	[1:700,]	[701,844,]	0.1	1.5321

Tabela 4 - Resultados com uma melhor escala para os níveis de fadiga

4. Conclusões e Sugestões

O presente trabalho de grupo, em todas as suas fases teve como principal objetivo a familiarização da utilização do conhecimento não simbólico, como forma de conhecimento e construção de mecanismos de análise e raciocínio para a resolução de problemas.

Foram aplicadas noções da linguagem em R que foram lecionados ao longo das várias semanas nas aulas desta Unidade Curricular, sendo estes desenvolvidos e aprimorados ao longo que o trabalho ia sendo realizado, uma vez que foi necessário analisar e criar mecanismos de raciocínio para resolver um conjunto de novas situações que nos eram propostas.

De uma forma geral, os resultados produzidos foram bastante satisfatórios e enriquecedores para todos os elementos do grupo, pois conseguimos cumprir todas as tarefas propostas e entender o porquê da sua existência e necessidade bem como a importância da sua utilização para situações práticas.

5. Bibliografia

- [Cortez, 2000] CORTEZ, Paulo, NEVES, José,
“Redes Neurais Artificiais”,
Departamento de Informática, Universidade
do Minho, Portugal, 2000