

# **Redes Neuronais para predição de químicos biodegradáveis**

# *Relatório Final*

# Grupo E5\_1

# Inteligência Artificial

# 3º ano do Mestrado Integrado em Engenharia Informática e Computação

# Elementos do Grupo:

# José Salgado de Magalhães Taveira Gomes - ei10157

# Ricardo Jorge Matos Figueiredo – ei11081

# 

# 31 de Maio de 2015

# Objetivo

A determinação da biodegradabilidade de um químico é efetuada através de uma rede neuronal de múltiplas camadas com recurso ao algoritmo de Backpropagation, algoritmo esse que é de aprendizagem supervisionada. Este algoritmo serve para treinar a rede que sabe quais são os valores esperados e tendo em conta esses valores são feitos ajustes aos pesos das arestas de modo a que o output se aproxime cada vez mais do resultado, ou seja, aprenda. A rede depois de treinada é sujeita a um teste cujo objetivo é a obtenção de um erro suficientemente baixo que possa ser desprezado, para que a rede consiga determinar com eficiência a biodegradabilidade ou não de um químico.

# Especificação

Para a realização do projeto foi proposto o uso do algoritmo de Backpropagation com aprendizagem supervisionada. Usando este algoritmo consegue-se determinar com bastante precisão se um químico é ou não biodegradável. Para gerar a rede neuronal, o utilizador terá de inserir um conjunto de dados que representam o número de nós por camada. A geração da rede e as operações efetuadas sobre a mesma fazem uso da API do Neuroph.

Desenvolvimento

O desenvolvimento do projecto foi realizado em ambiente Windows, com a ajuda do IDE Netbeans com recurso a linguagem Java, como já foi referido, usamos a API do Neuroph, facilitando assim a criação e uso de redes neuronais.

A aplicação encontra-se dividida em 3 módulos principais: interface de linha de comandos (CLI), lógica (logic), leitura de dados de ficheiros (cvsFileIO).

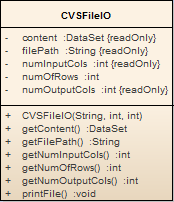


Fig 1 - Diagrama de Classes do pacote cvsFileIO

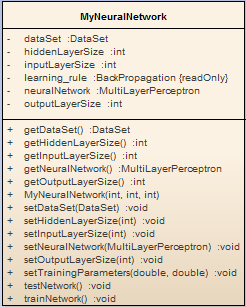


Fig 2 - Diagrama de Classes do pacote logic

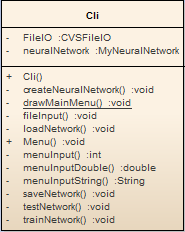


Fig 3 - Diagrama de Classes do pacote cli

É de referir que a aplicação está construída de modo a permitir guardar a rede neuronal treinada, o utilizador poderá sair da aplicação e, quando voltar, carregar a sua rede e assim continuar o trabalho sem voltar a alterar as opções e a treinar a rede com o *dataset*. Durante a esta fase, foi também necessário normalizar os valores usados no *dataset* entre 0 e 1, pois usamos o SIGMOID como transfer function e os valores que estavam a ser obtidos não eram próximos dos esperados. 0 e 1 são também os valores esperados do nosso *dataset* (1 se for biodegradável, 0 se não o for). A nossa aplicação usa 70% do *dataset* para treino e 30% para testes.

Experiências

Nesta experiência iremos tentar determinar disparidades de sucesso entre o uso de um número de nós diferentes na camada intermédia, de modo a tentar determinar o número ideal de nós para minimizar o erro na rede.

Rede Neuronal 1

* Número de nós de Entrada: 41
* Número de nós da camada intermédia: 150
* Número de nós de saída: 1
* Max Error: 5%
* Learning Rate: 0,5
* Erro da Rede: 4,858%

Rede Neuronal 2

* Número de nós de Entrada: 41
* Número de nós da camada intermédia: 100
* Número de nós de saída: 1
* Max Error: 5%
* Learning Rate: 0,5
* Erro da Rede: 4,773%

Rede Neuronal 3

* Número de nós de Entrada: 41
* Número de nós da camada intermédia: 80
* Número de nós de saída: 1
* Max Error: 5%
* Learning Rate: 0,5
* Erro da Rede: 4,767%

Rede Neuronal 4

* Número de nós de Entrada: 41
* Número de nós da camada intermédia: 50
* Número de nós de saída: 1
* Max Error: 5%
* Learning Rate: 0,5
* Erro da Rede: 4,977%

Rede Neuronal 5

* Número de nós de Entrada: 41
* Número de nós da camada intermédia: 10
* Número de nós de saída: 1
* Max Error: 5%
* Learning Rate: 0,5
* Erro da Rede: 4,995%

Com os valores obtidos criou-se um gráfico de regressão linear que aproxima os valores obtidos de uma função, sendo visível que o número de neurónios da rede intermédia será perto de 90 para um erro mínimo de aproximadamente 4,75%.

Fig 4 – Gráfico de percentagem de erro da rede e número de nós da camada intermédia

Conclusões

Dados os resultados das experiências efectuadas, podemos concluir que existe um valor para o qual o número de nós da camada intermédia tende que consegue minimizar o erro total da rede. Podemos dizer com segurança que esse valor se encontra algures entre os 50 e os 100 nós, podendo ver também que para o gráfico gerado, aproximadamente 90 nós foi o número de nós que para as mesmas condições nos obteve um erro total da rede menor. Concluindo, a rede neuronal ideal varia de *dataset* para *dataset* e neste caso em específico, seria uma rede que rondasse 90 nós na camada intermédia.

Recursos

# Blbilografia:

<http://archive.ics.uci.edu/ml/datasets/QSAR+biodegradation>

<http://en.wikipedia.org/wiki/Backpropagation>

<http://en.wikipedia.org/wiki/Artificial_neural_network>

<http://neuroph.sourceforge.net/documentation.html>

Software:

Netbeans

NeurophStudio

Excel