



Universidade do Minho

Redes Neurais para classificação de Ressonâncias Magnéticas de Difusão

Mestrado Integrado em Engenharia Informática

(2 º Semestre 2016/2017)

Grupo 4:
67751 Pedro Silva
69302 João Farinha
Orientado por:
Professor Victor Alves

Braga
Junho de 2017

Resumo

Neste relatório é apresentada o processo de investigação realizado, na tentativa de criar uma rede neuronal capaz de classificar ressonâncias magnéticas de difusão.

Primeiro é feita uma breve introdução às ferramentas e tecnologias utilizadas.

De seguida são descritos os conceitos básicos sobre as ressonâncias magnéticas de difusão, de modo a criar um melhor entendimento sobre o tema e a utilidade desta rede.

Depois são explicados os dados fornecidos e o processo de tratamento executado sobre eles.

De seguida é explicado em detalhe o processo de investigação, mencionando os problemas que surgiram durante a investigação e as soluções desenvolvidas para tais problemas, acompanhado da evolução dos resultados ao longo do processo.

Por fim é feita uma análise crítica sobre o trabalho realizado, expressando as opiniões sobre o sucesso da investigação e o método de trabalho utilizado.

Índice

Introdução	4
Ressonâncias Magnéticas de Difusão	4
Dados fornecidos	5
Processo de investigação	6
Primeira rede	6
Segunda rede	8
Tentativas de resolução do problema de mínimo local	9
Conclusão e trabalhos futuros	11
Bibliografia	11

Introdução

Este documento apresenta as decisões tomadas durante o processo de investigação, na tentativa de desenvolvimento de uma rede neuronal de classificação de DTI's.

Para as redes serem rapidamente desenvolvidas foi tirado partido da framework Keras a correr em Python versão 3.4, com o backend Theano.

Foi também tirado partido de diversas bibliotecas do Python como o matplotlib para análise de histogramas, o Nibabel para leitura do tipo de dados padrão e o Numpy para o processamento de dados após a leitura.

Esta investigação tem como objetivo averiguar a possibilidade de desenvolver uma rede neuronal capaz de classificar DTI, ou seja capaz de determinar se a DTI que lhe é fornecida pertence a um indivíduo saudável ou doente.

Isto é interessante pois a obtenção e posterior análise de uma DTI é um processo demorado, no entanto se uma rede determinar com confiança e rapidez, se um paciente se encontra doente ou não, pode ser poupado tempo e recursos na análise.

Para além disso se esta investigação for um sucesso, será possível realizar investigações sobre DTI's com menor pré-processamento, tornando a obtenção de um diagnóstico mais expedita.

Ressonâncias Magnéticas de Difusão

Ressonâncias magnéticas de difusão (DWI), é um método que utiliza a difusão das moléculas da água de modo a gerar contraste nas imagens de ressonâncias magnéticas.

Com a utilização deste método, é possível mapear o processo de difusão das moléculas, em tecidos biológicos, com a utilização de espécimes vivos e de forma não invasiva.

Esta difusão das moléculas interage com vários obstáculos, tais como macromoléculas, fibras e membranas, revelando assim detalhes microscópicos sobre a arquitetura do tecido cerebral, quer este se encontre num estado normal, ou danificado.

Um tipo especial de ressonâncias magnéticas de difusão, são as **diffusion tensor imaging (DTI)**, que tem sido utilizado extensivamente para mapear a tractografia de matéria branca no cérebro, entre outros tipos de tecido.

Numa DTI, cada voxel, contém um ou mais pares de parâmetros, o ritmo de difusão e a direção preferida de difusão, descritos em termos de um espaço tridimensional, para o qual esse parâmetro é válido.

As propriedades de um único voxel de uma DTI, são normalmente obtidos com a utilização de várias DWIs.

A maior quantidade de informação contida nos voxels das DTIs, torna-a extremamente sensível a patologias subtis no cérebro.

É através da utilização de DTIs, que é pretendido criar uma rede que consiga determinar se a DTI é relativa a tecido de um cérebro danificado, ou saudável.

Dados fornecidos

Desde da ressonância magnética inicial, até esta se tornar uma DTI, são realizados vários passos de processamento.

Este processo é composto por vários passos complexos e demorados.

Os dados fornecidos correspondem aos vários estágios de processamento.

Na totalidade foram fornecidos dados sobre 106 análises e a sua respetiva classificação.

Cada análise é composta por 62 imagens de 128x128 dimensão.

Sendo de especial importância os dados no estágio dtifit_V1/V2/V3, que fornece informação sobre o movimento em cada voxel. Esse movimento é modelado através de uma elipse que é descrita matematicamente através de três vetores próprios.

Neste tipo de dados existe uma quarta dimensão que mapeia três imagens com os valores dos vetores ao longo do eixo xx,yy e zz, respetivamente.

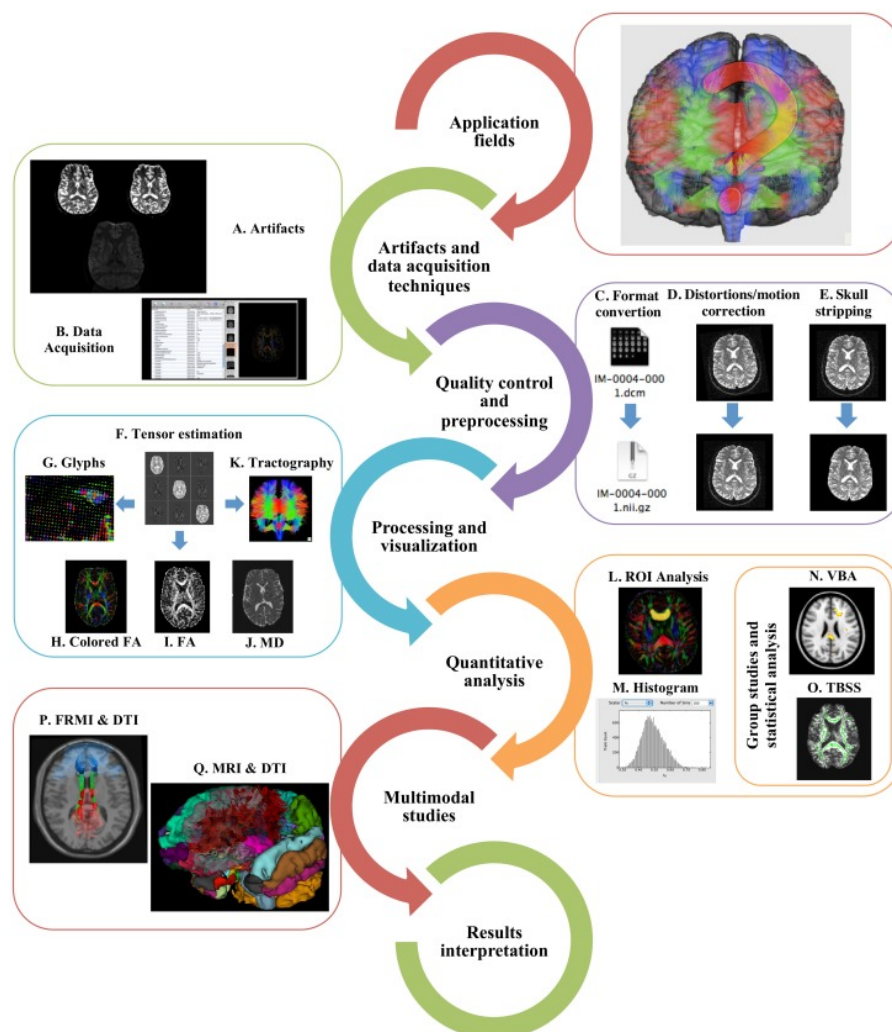
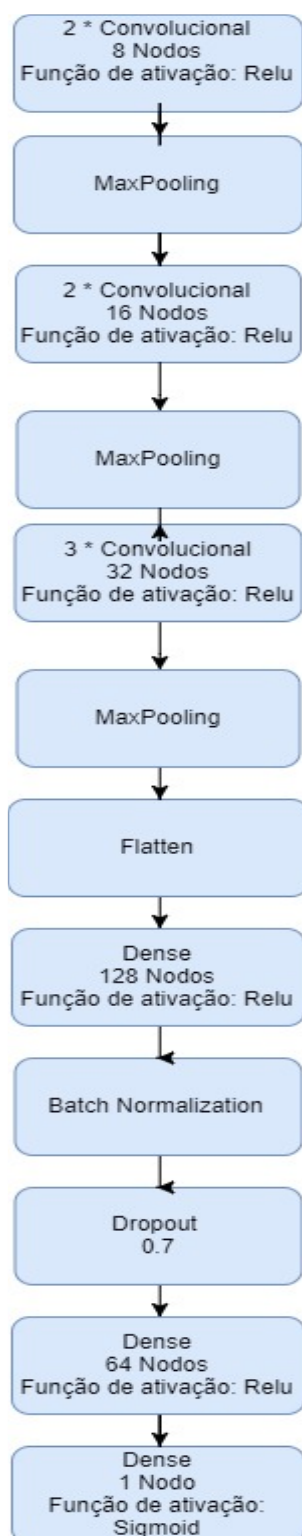


Fig. 1 - Fluxo de trabalho habitual

Processo de investigação

Para obter maior coerência nos resultados todos os testes foram executados com uma distribuição de 86 casos para treino, 10 para validação e 10 para teste. Os dados encontravam-se distribuídos originalmente com valores entre -1 e 1 no entanto como parte do pré-processamento estes foram normalizados para o intervalo de 0 a 1.

Primeira rede



Após uma breve análise dos dados fornecidos, foi desenvolvida uma rede com o objetivo de obter, rapidamente um resultado base.

Para esta rede foram utilizados os dados na fase do tratamento de difusão média.

A razão de ter sido escolhido este tipo de dados, foi a vontade de obter um resultado o mais rapidamente possível, e sendo os dados nesta fase de tratamento os de menor dimensão, foram escolhidos na esperança de obter menores tempos de treino.

Como arquitetura da rede foi desenhada uma baseada num artigo sobre a utilização de redes neuronais para classificação de ressonâncias magnéticas.

Como tal foi desenvolvida uma rede com a arquitetura, representada à esquerda.

Sobre esta arquitetura foram realizados vários testes, dos quais alguns se encontram de seguida discriminados:

- Teste 1
 - 5 epochs
 - aproximadamente 600 segundos de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.185, val_acc: 0.8000
- Teste 2
 - 200 epochs
 - aproximadamente 2 horas de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000
- Teste 3
 - 5 epochs

- aproximadamente 600 segundos de tempo total de treino
- função de ativação das camadas intermédias alterada para tanh
- Resultado: loss: 10.1207, acc: 0.3721, val_loss: 12.8945, val_acc: 0.2000
- Embora o resultado final tenha sido diferente a evolução da aprendizagem teve o mesmo comportamento
- Teste 4
 - 5 epochs
 - aproximadamente 600 segundos de tempo total de treino
 - uma camada densa de 512 nodos, adicionada entre a parte de camadas convolucionais e a parte das camadas densas
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000

Durante a realização dos variados testes começou a tornar-se evidente um problema de possível mínimo local, afinal todas as configurações testadas convergiam para o mesmo valor de precisão (0.6279), para além do resultado final ser sempre o mesmo, ao analisar os resultados intermédios era observado que após o resultado de 0.6279 de precisão ser obtido a rede nunca mais obtinha outro resultado, independentemente do total de epochs de treino.

Outros problemas detetados nesta fase foi a fraca compreensão dos dados, e do artigo em qual foi baseada a arquitetura da rede.

Nesta fase foram utilizados dados relativos a etapa de processamento da difusão média, no entanto esses dados não são os mais corretos a utilizar, para o tipo de problema a ser resolvido.

E no caso do artigo em qual foi baseada a arquitetura da rede, as ressonâncias a serem classificadas eram ressonâncias magnéticas tradicionais e não ressonâncias magnéticas de difusão, o que significa que a rede não foi tão adequada a este problema específico como era inicialmente esperado.

Segunda rede



Nesta segunda rede, foram feitas várias alterações para tentar resolver os problemas encontrados na primeira rede. Foram reduzidas o número de camadas convolucionais e alteradas a posição relativa colocando as camadas de maior tamanho antes das menores, de forma a reduzir o número de parâmetros ao longo da rede.

A representação visual da arquitetura encontra-se à esquerda.

Nesta rede foram utilizados dados noutra fase de tratamento, pois estes novos dados são mais adequados para o objetivo. Especificamente passaram a ser usados o conjunto de dados no formato L1 em junção com os de formato V1, sendo que estes representam a direção e o comprimento dos vetores de difusão das DTIs.

Foram removidas as camadas de dropout e batch normalization, pois foi determinado que a rede ainda não se encontra numa fase de desenvolvimento, que justifique estas medidas que gastam recursos desnecessariamente.

O principal foco destas alterações era ultrapassar o problema de mínimo local que foi detetado, na rede anterior com a utilização do outro tipo de dados, que está a condicionar os resultados finais.

- Teste 1
 - 5 epochs
 - aproximadamente 450 segundos de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000
- Teste 2
 - 200 epochs
 - aproximadamente 1:30h de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000

Mesmo com a utilização de um tipo de dados diferente, e de uma rede com uma arquitetura radicalmente diferente, o problema dos resultados finais não se alterarem manteve-se, como foi observado nos testes realizados.

Tentativas de resolução do problema de mínimo local

Após a nova configuração de rede não ter resolvido o problema de mínimo local, foi estabelecida uma diferente abordagem na tentativa de o solucionar.

A primeira abordagem tomada foi a alteração dos hiperparâmetros utilizados pelo otimizador.

Até aqui foi utilizado o otimizador sgd(stochastic gradient descent) com learning rate de 0.01, momentum 0 e decay 0.

Em que learning rate controla o tamanho dos saltos executados pelo algoritmo em cada iteração, momentum faz com que a otimização acelere na direção de otimizações anteriores e o decay faz com que as iterações finais tenham menos efeito na aprendizagem que as iniciais.

Ao alterar estes valores no otimizador é esperado obter resultados diferentes.

O problema persistiu mesmo com a alteração do otimizador sgd, para o otimizador adam.

- Teste 1
 - learning rate = 0.1, momentum= 0.9, decay= 1e-6
 - aproximadamente 420 segundos de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000
- Teste 2
 - otimizador adam
 - aproximadamente 430 segundos de tempo total de treino
 - Resultado: loss: 0.6663 - acc: 0.6279 - val_loss: 0.5528 - val_acc: 0.8000
- Teste 3
 - learning rate = 0.1, momentum= 0.9, decay= 1e-3
 - aproximadamente 430 segundos de tempo total de treino
 - Resultado: loss: 5.9321, acc: 0.6279, val_loss: 3.1885, val_acc: 0.8000
- Teste 4
 - learning rate = 0.1, momentum= 0.5, decay= 1e-6
 - aproximadamente 430 segundos de tempo total de treino
 - loss: 5.9321 - acc: 0.6279 - val_loss: 3.1885 - val_acc: 0.8000

No entanto nenhuma das alterações apresentou mudanças nos resultados finais.

Por fim foi realizada uma abordagem de tratamento de dados de modo a tentar resolver o problema, pois todas as tentativas de resolução deste problema, através da alteração da rede foram infrutíferas.

Como tal foi realizada uma análise mais cuidada dos dados que são utilizados.

Foi detetado que os valores dos dados eram todos muito próximos de zero, o que pode estar a afetar a capacidade de aprendizagem da rede.

Ora de modo a tentar obter uma distribuição de valores mais acentuada, submeteu-se as imagens a uma equalização do histograma, antes de serem alimentadas à rede.

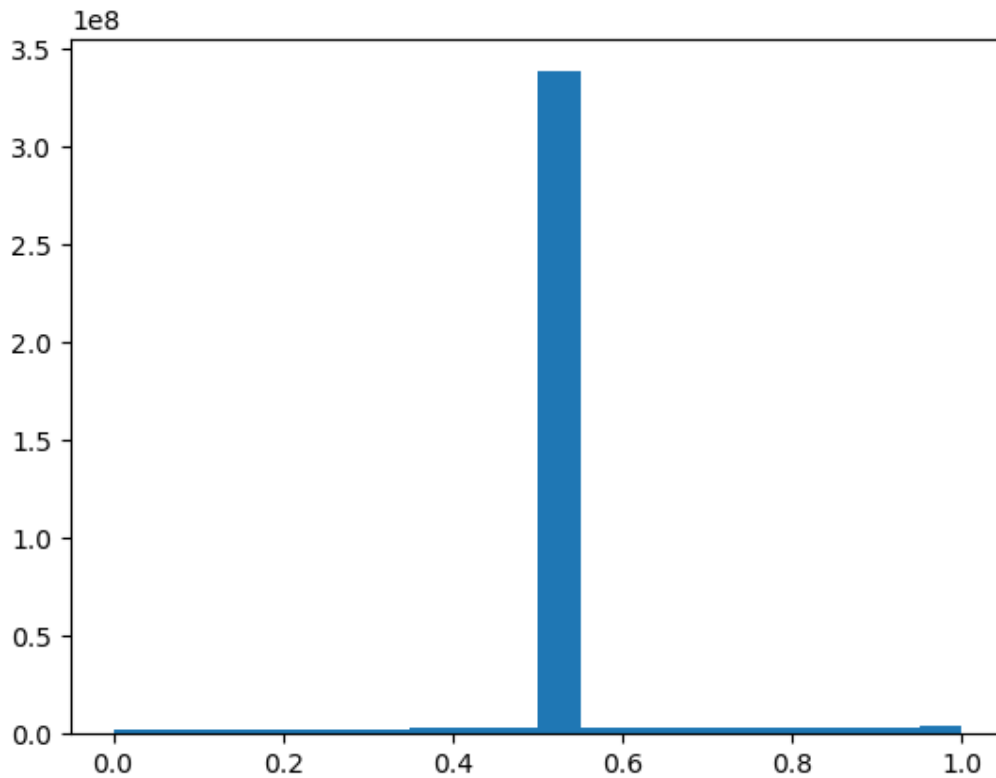


Fig. 2 - Histograma antes da equalização

- Teste:
 - 5 epochs, otimizador adam, dados com equalização de histograma
 - aproximadamente 430 segundos de tempo total de treino
 - loss: 0.6664 - acc: 0.6279 - val_loss: 0.5527 - val_acc: 0.8000

No entanto este novo passo de pré-processamento não teve qualquer efeito nos resultados finais.

Conclusão e trabalhos futuros

Quando se inicia um processo de investigação, não há garantias que esse processo termine com resultados positivos.

O próprio conceito de investigação implica que quando se começa não se sabe o resultado final.

Este processo tinha como objetivo determinar se era possível desenvolver uma rede neuronal capaz de classificar DTI's.

Idealmente neste momento seria aqui estabelecido se isso é impossível ou possível.

No entanto não podemos fazer tais afirmações com os resultados obtidos.

Por um lado não podemos determinar que é impossível, pois durante o processo as redes desenvolvidas classificavam os exames com 62.79% de precisão.

No entanto com esse nível de precisão também não podemos determinar que é possível, pois para além de ser uma precisão insuficiente para o ramo médico, foi também obtido por todas as redes testadas, o que torna o valor suspeito.

Como tal temos de concluir a investigação como inconclusiva, não conseguimos afirmar ou negar a possibilidade de utilização de redes neuronais para a classificação de DTI's.

Quanto ao método de trabalho, a vontade de obter resultados rápido, refletiu-se negativamente no processo, pois fez com que tenha sido iniciado o desenvolvimento, sem ainda ter o total entendimento sobre os dados fornecidos, o que se refletiu na utilização de dados que não eram os mais corretos para a resolução deste problema.

No futuro de modo a tentar obter uma conclusão sobre esta investigação, seriam realizados testes, e aplicadas técnicas de processamento de imagem sobre os dados de modo a tentar resolver o problema de mínimo local.

Bibliografia

Soares, J. M., Marques, P., Alves, V., & Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in Neuroscience*, 7, 31. <http://doi.org/10.3389/fnins.2013.00031>

Nielsen, M. A. (1970, January 01). *Neural Networks and Deep Learning*. Retrieved April 04, 2017, from <http://neuralnetworksanddeeplearning.com/>

Keras: Deep Learning library for Theano and TensorFlow. (n.d.). Retrieved April 04, 2017, from <https://keras.io/>

Sergey Korolev,& Amir Safiullin,& Mikhail Belyaev,& Yulia Dodonova (23, January 2017).Residual and plain convolutional neural networks for 3D brain MRI classification. Retrieved April 04,2017 from <https://arxiv.org/pdf/1701.06643.pdf>