**Human Activity Recognition Using Smartphones**

**Reconhecimento de Padrões**

**2016/2017**

Gabriel Angel Amarista Rodrigues 2016211454

Maria José Mateus Branco 2013134952

**Introdução**

Atualmente, os smartphones são utilizados em diversas atividades do quotidiano, nomeadamente para monitorização da atividade humana. Assim, foi desenvolvido um trabalho com o objetivo de desenvolver classificadores para reconhecimento da atividade humana.

O projeto foi dividido em duas partes, na primeira é considerado apenas o caso binário (walking e not walking) e na segunda é estudado um caso com seis classes (walking, walking upstairs, walking downstairs, sitting, standing e laying).

**Procedimento**

Inicialmente, os dados foram carregados, estando divididos em dois conjuntos diferentes, um conjunto de treino e outro de teste. A cada um destes conjuntos corresponde uma matriz que contém todas as *features* (*X\_train* e *X\_test*) e um vetor com a classe correspondente a cada uma das instâncias (*y\_train* e *y\_test*). Após este passo, foi feito o *scaling* dos dados, de forma a que estes estejam todos contidos no intervalo de valores entre -1 e 1.

Uma vez que existem 561 *features* disponíveis para analisar, e que existe a possibilidade de algumas delas apresentarem informação insignificante para a classificação ou terem informação redundante, procedeu-se à redução da dimensionalidade dos dados, usando as técnicas de PCA e LDA, para além destas técnicas foi também testado o método de Kruskal Wallis para selecionar as *features* com maior importância na classificação. Ao longo desta primeira fase do projeto foram usadas 3 *features* para a realização de todos os testes. Todas as funções que foram usadas para redução de dimensionalidade, como PCA, Kruskal Wallis, encontram-se desenvolvidas no ficheiro *FeatureProcess.m.*

Após a redução da dimensionalidade, foram testados apenas classificadores simples, o Fisher LDA e outros baseados nas distâncias mínimas. Assim, foram feitos diferentes testes para verificar qual a combinação que apresenta uma melhor performance no caso binário. No cenário *multiclass*, apenas foram desenvolvidos e testados classificadores baseados nas distâncias mínimas, uma vez que o não foi desenvolvido nenhum classificador do tipo Fisher LDA para distinguir as seis classes diferentes. As funções que permitem desenvolver estes classificadores, encontram-se no programa *Classifier.m*.

No desenvolvimento dos classificadores, foram feitos vários testes, dividindo o conjunto de treino em várias partes e fazendo validação cruzada, com o objetivo de avaliar a capacidade de generalização dos modelos.

Após a escolha dos classificadores com melhor *performance*, estes foram aplicados ao conjunto de teste e analisados os resultados.

Para proceder à redução de dimensionalidade, criação e teste dos classificadores, as funções necessárias são chamadas no programa *testing\_script.m*.

**Resultados**

**Cenário Binário (Conjunto de Treino)**

**Kruskal Wallis + Fisher LDA**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3490 | 0 |
| Not Walking | 30 | 3840 |

Accuracy=0.996

Precisão=0.9915

Especificidade=0.992

Sensibilidade/Recall=1.000

**Kruskal Wallis + Minimun Euclidean Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3260 | 0 |
| Not Walking | 10 | 4090 |

Accuracy=0.999

Precisão=0.997

Especificidade=0.998

Sensibilidade/Recall=1.000

**Kruskal Wallis + Minimun Mahalanobis Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3390 | 0 |
| Not Walking | 20 | 3950 |

Accuracy=0.997

Precisão=0.994

Especificidade=0.995

Sensibilidade/Recall=1.000

**PCA + Fisher LDA**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3540 | 0 |
| Not Walking | 0 | 3820 |

Accuracy=1.000

Precisão=1.000

Especificidade=1.000

Sensibilidade/Recall=1.000

**PCA + Minimun Euclidean Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3650 | 0 |
| Not Walking | 10 | 3700 |

Accuracy=0.999

Precisão=0.996

Especificidade=0.997

Sensibilidade/Recall=1.000

**PCA + Minimun Mahalanobis Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3310 | 0 |
| Not Walking | 10 | 4040 |

Accuracy=0.999

Precisão=0.997

Especificidade=0.998

Sensibilidade/Recall=1.000

**LDA + Fisher LDA**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 3300 | 0 |
| Not Walking | 20 | 4040 |

Accuracy=0.997

Precisão=0.994

Especificidade=0.995

Sensibilidade/Recall=1.000

**LDA + Minimun Euclidean Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 10 | 3480 |
| Not Walking | 3860 | 10 |

Accuracy=0.003

Precisão=0.003

Especificidade=0.003

Sensibilidade/Recall=0.003

**LDA + Minimun Mahalanobis Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 10 | 3440 |
| Not Walking | 3910 | 0 |

Accuracy=0.001

Precisão=0.003

Especificidade=0.000

Sensibilidade/Recall=0.003

**Cenário Binário (Conjunto de Teste)**

Após a verificação dos resultados obtidos aplicando os classificadores ao conjunto de treino foi possível escolher os classificadores desenvolvidos através das combinações PCA+Fisher LDA, PCA + Minimun Mahalanobis Distance e Kruskal Wallis + Minimun Euclidean Distance para proceder à sua aplicação ao conjunto de teste e assim verificar a sua *performance,* uma vez que foram os que apresentaram melhores resultados na fase de treino*.*

**PCA+Fisher**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 1387 | 0 |
| Not Walking | 5 | 1555 |

Accuracy=0.998

Precisão=0.996

Especificidade=0.997

Sensibilidade/Recall=1.000

**PCA + Minimun Mahalanobis Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 1387 | 0 |
| Not Walking | 3 | 1557 |

Accuracy=0.999

Precisão=0.998

Especificidade=0.998

Sensibilidade/Recall=1.000

**Kruskal Wallis + Minimun Euclidean Distance**

|  |  |  |
| --- | --- | --- |
|  | Walking | Not Walking |
| Walking | 1387 | 0 |
| Not Walking | 6 | 1554 |

Accuracy=0.998

Precisão=0.996

Especificidade=0.996

Sensibilidade/Recall=1.000

**Cenário com 6 classes (Conjunto de Treino)**

**Kruskal Wallis + Minimun Euclidean Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 930 | 250 | 100 | 0 | 0 | 0 |
| Walking Upstairs | 200 | 780 | 150 | 0 | 0 | 0 |
| Walking Downstairs | 100 | 130 | 630 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 1060 | 190 | 10 |
| Standing | 0 | 0 | 0 | 1030 | 310 | 10 |
| Laying | 0 | 10 | 0 | 1190 | 280 | 0 |

Accuracy=0.505

**Kruskal Wallis + Minimun Mahalanobis Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 510 | 350 | 270 | 0 | 0 | 0 |
| Walking Upstairs | 380 | 590 | 180 | 0 | 0 | 0 |
| Walking Downstairs | 120 | 80 | 1050 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 1070 | 210 | 10 |
| Standing | 0 | 20 | 0 | 920 | 350 | 30 |
| Laying | 0 | 0 | 0 | 980 | 220 | 20 |

Accuracy=0.488

**PCA + Minimun Euclidean Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 580 | 270 | 280 | 0 | 0 | 0 |
| Walking Upstairs | 160 | 770 | 200 | 0 | 0 | 0 |
| Walking Downstairs | 240 | 100 | 750 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 800 | 430 | 50 |
| Standing | 0 | 0 | 0 | 520 | 750 | 0 |
| Laying | 0 | 10 | 0 | 10 | 0 | 1440 |

Accuracy=0.692

**PCA + Minimun Mahalanobis Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 430 | 630 | 190 | 0 | 0 | 0 |
| Walking Upstairs | 190 | 620 | 170 | 0 | 0 | 0 |
| Walking Downstairs | 260 | 110 | 650 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 820 | 440 | 100 |
| Standing | 0 | 0 | 0 | 840 | 640 | 0 |
| Laying | 0 | 30 | 0 | 10 | 130 | 1100 |

Accuracy=0.579

**LDA + Minimun Euclidean Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 490 | 350 | 260 | 0 | 0 | 0 |
| Walking Upstairs | 150 | 840 | 150 | 0 | 0 | 0 |
| Walking Downstairs | 70 | 140 | 630 | 0 | 0 | 0 |
| Sitting | 0 | 10 | 0 | 160 | 910 | 220 |
| Standing | 0 | 0 | 0 | 320 | 950 | 190 |
| Laying | 0 | 30 | 0 | 350 | 900 | 240 |

Accuracy=0.450

**PCA + Minimun Mahalanobis Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 550 | 540 | 290 | 0 | 0 | 0 |
| Walking Upstairs | 150 | 740 | 150 | 0 | 0 | 0 |
| Walking Downstairs | 170 | 60 | 650 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 910 | 90 | 320 |
| Standing | 0 | 0 | 0 | 990 | 100 | 390 |
| Laying | 0 | 10 | 0 | 920 | 100 | 220 |

Accuracy=0.4312

**Cenário com 6 classes (Conjunto de Teste)**

Após a verificação dos resultados obtidos aplicando os classificadores ao conjunto de treino foi possível escolher a combinação o classificador baseado na combinação PCA + Minimun Euclidean Distance e proceder à sua aplicação ao conjunto de teste e assim verificar a sua *performance,* uma vez que foi o que apresentoumelhores resultados na fase de treino*.*

**PCA + Minimun Euclidean Distance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Walking Upstairs | Walking Downstairs | Sitting | Standing | Laying |
| Walking | 238 | 136 | 122 | 0 | 0 | 0 |
| Walking Upstairs | 54 | 386 | 31 | 0 | 0 | 0 |
| Walking Downstairs | 130 | 58 | 232 | 0 | 0 | 0 |
| Sitting | 0 | 4 | 0 | 273 | 214 | 0 |
| Standing | 0 | 1 | 0 | 115 | 416 | 0 |
| Laying | 0 | 2 | 0 | 6 | 0 | 529 |

Accuracy=0.704

**Discussão**

Depois de aplicadas as diferentes combinações entre as técnicas de redução de dimensionalidade e os diferentes tipos de classificadores, seguidas da validação cruzada, foi possível verificar que para o caso binário os que apresentaram uma maior capacidade de generalização foram as combinações PCA+Fisher LDA, PCA + Minimun Mahalanobis Distance e Kruskal Wallis + Minimun Euclidean Distance. No caso *multiclass,* o classificador que se destacou foi o que foi desenvolvido com base no PCA para redução da dimensionalidade e na distância euclidiana mínima para classificação dos dados.

Relativamente aos dois últimos testes no cenário binário, os resultados não foram satisfatórios, uma vez que os métodos usados para o cálculo das distâncias não suportam valores complexos e o *output* do método LDA contém números complexos, desta forma o classificador não apresenta a *performance* desejada.

De acordo com os resultados na fase de treino foram aplicados os classificadores desenvolvidos aos dados para teste. No cenário binário, os resultados foram muito satisfatórios uma vez que a *accuracy* observada nos três testes foi aproximadamente igual a 1. Relativamente ao cenário onde se pretende fazer a distinção entre as seis classes, a *accuracy* obtida foi de 70%, o que permite verificar que existe uma relação entre as *features* existentes e o estado da pessoa (a caminhar, sentada, em pé,…), desta forma os resultados poderão ser melhorados aplicando classificadores mais complexos na fase posterior do trabalho.