

Trabalho 3 (Opção 2) - 2017/1
Reconhecimento de Atividades com Aprendizado de Máquina

A área de Reconhecimento de Atividades [1,2,3] tem como meta identificar o que uma pessoa está fazendo, dentre um conjunto de possibilidades em um contexto específico. Quando esse problema é resolvido com uma técnica de Aprendizado de Máquina, usualmente um classificador é aprendido, a partir de dados coletados de sensores, tais como celulares, ou dispositivos usáveis. Neste trabalho, você usará técnicas de Aprendizado de Máquina para criar um classificador que consiga detectar qual atividade está sendo executada por uma pessoa, de acordo com a base de dados. Neste trabalho, experimentaremos diferentes tipos de classificadores, técnicas de seleção de atributos, e técnicas de combinação de classificadores.

Para a execução do trabalho, você deverá seguir os seguintes passos:

1 - Escolher a base de dados . Algumas opções são como seguem:

<https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones>
<http://www.cis.fordham.edu/wisdm/dataset.php>
http://ps.cs.utwente.nl/Blog/Activity_Recognition_DataSet_Shaoib.rar
<http://www.opportunity-project.eu/challengeDataset>

2 - Experimentos com a ferramenta de aprendizado de máquina escolhida. Você pode escolher qualquer ferramenta de sua preferência, atentando para a existência das implementações requeridas pelo trabalho. Algumas opções são:

<http://scikit-learn.org/stable/>
<http://www.cs.waikato.ac.nz/ml/weka/>
<https://orange.biolab.si/>

Os experimentos são como seguem:

- (a) Comparar abordagens de redução de dimensionalidade (PCA [4]) e seleção de atributos (RFE [5] e Randomized Lasso [6]) . Escolha duas dentre essas opções. O objetivo dessa fase é reduzir o conjunto de atributos, ou os agrupando, ou os removendo.

- (b) Seleção de parâmetros com GridSearch [7]. Algoritmos de Aprendizado de Máquina requerem diversos parâmetros a serem usados durante o processo de otimização. O método de GridSearch é útil para essa escolha de parâmetros.
- (c) Seleção de classificadores. No mínimo três classificadores devem ser escolhidos, dentre as opções: Naive Bayes, SVM, Decision tree, Random Forest, AdaBoost [8,9]. Observe que as duas últimas são métodos de Ensemble [10] e é interessante que ao menos um deles seja escolhido.
- (d) Combinação de Classificadores com Stacking. Para essa fase, você deverá combinar os resultados obtidos a partir das combinações de fases anteriores, com os classificadores, usando a abordagem de Stacking [11], e escolher um dos classificadores anteriores para ser o classificador após a combinação. O método de stacking considera que os atributos são as saídas de classificadores, e tenta aprender um outro classificador a partir de tais atributos e das classes originais.

Referências:

- [1] Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005, July). Activity recognition from accelerometer data. In *Aaaí* (Vol. 5, No. 2005, pp. 1541-1546).
- [2] Bao, L., & Intille, S. S. (2004, April). Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing* (pp. 1-17). Springer Berlin Heidelberg.
- [3] Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2), 74-82.
- [4] Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.
- [5] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- [6] Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics.
- [7] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951-2959).
- [8] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.
- [9] Mitchell, T. M. (1997). *Machine learning*. WCB.

[10] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Springer Berlin Heidelberg.

[11] Rooney, N., Patterson, D., & Nugent, C. (2004, November). Reduced ensemble size stacking [ensemble learning]. In *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on* (pp. 266-271). IEEE.

Instruções Adicionais

- O grupo deve ser composto de 2 a 5 componentes.
- O código e um relatório descrevendo o trabalho devem ser enviados por um dos integrantes do grupo na tarefa do Classroom, até o dia 06.07. Todas as instruções para a execução do trabalho, bem como comentários ao longo do código, devem ser incluídos. Por favor, especificar o nome de todos os componentes.
- Cada integrante do grupo deve enviar um relatório como sua tarefa no Classroom, detalhando sua porcentagem de participação no trabalho e a porcentagem de participação dos demais integrantes do grupo. Exemplo (supondo grupo de 3): se você fez metade do trabalho e cada um dos seus colegas contribuiu com 25%, reporte qual foi a sua contribuição correspondente aos 50% e no que seus colegas contribuíram, com 25% cada