Introducción a las Tecnologías del Habla

Trabajo Práctico 3: Aprendizaje Automático

Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

Leandro Lovisolo
LU 645/11

Segundo Cuatrimestre de 2012

Introducción

El objetivo de este trabajo práctico es construir un sistema de reconocimiento automático del género de una persona a partir de una grabación corta de su habla, aplicando técnicas de aprendizaje automático.

Para la realización del sistema se dispone de un corpus de grabaciones recolectadas por todos los alumnos de esta materia durante el TP 1, de las cuales se extrayeron el género del hablante y un conjunto de atributos acústicos que serán utilizados como referencia para entrenar el sistema y evaluar su eficacia.

El sistema deberá implementarse sobre la suite de aprendizaje automático Weka[1], en la que se deberá construir un clasificador que tome como entrada los atributos acústicos de una grabación, y decida en base a estos el género de la persona.

En primer lugar, se deberá implementar como *sistema baseline* un clasificador de reglas RIPPER[2] utilizando como único atributo la media de la frecuencia fundamental del hablante. En el TP1 habíamos visto que la diferencia de este atributo para cada género era significativa y grande. Ahora veremos cuál es su poder predictivo en esta tarea.

Finalmente, se deberá experimentar con diferentes clasificadores y diferentes conjuntos de atributos, en busca de una configuración que arroje buenos resultados. La tasa de aciertos deberá ser mayor o igual a 94%.

Materiales y métodos

Las instancias en la base de datos corresponden a los segmentos del habla sin pausas (inter-pausal units o IPUs) de todas las grabaciones. Cada instancia registra el género del hablante y 1582 atributos acústicos.

Los atributos acústicos en la base de datos fueron extraídos de los archivos de audio con la herramienta openSMILE[3], usando la configuración para el INTERSPEECH 2010 Paralinguistic Challenge[4], y almacenados en formato ARFF para facilitar su lectura desde Weka. Para más información, ver las páginas 30 y 31 del openSMILE book[5].

La base de datos de atributos acústicos, junto con el enunciado completo del TP, pueden descargarse desde la siguiente URL: `http://habla.dc.uba.ar/gravano/ith-2012/tp3/`

Sistema baseline

Se empleó un clasificador `rules.JRip` con opciones de configuración por defecto y cross-validation de 10 folds.

El atributo utilizado fue `F0Final_sma_amean` (media de la frecuencia fundamental.)

[1] `http://www.cs.waikato.ac.nz/ml/weka/`

[2] Repeated Incremental Pruning to Produce Error Reduction (RIPPER.) Ver `http://wiki.pentaho.com/display/DATAMINING/JRip`

[3] `http://opensmile.sourceforge.net/`

[4] `http://emotion-research.net/sigs/speech-sig/paralinguistic-challenge`

[5] `http://sourceforge.net/projects/opensmile/files/openSMILE_book_1.0.0.pdf`

Resultados

Se obtuvo un porcentaje de instancias correctamente clasificadas del **86.9315%**, con la siguiente matriz de confusión:

```
=== Confusion Matrix ===

   a   b   <-- classified as
 660  94 |   a = f
 110 697 |   b = m
```

Mejor sistema desarrollado

La mejor configuración hallada fue un clasificador `trees.J48` con los siguientes atributos:

mfcc_sma[10]_quartile1 Primer cuartil del decimoprimer coeficiente cepstral en las frecuencias de Mel, suavizado por un filtro promedio móvil de ventana de longitud 3.

mfcc_sma[12]_quartile1 Decimotercer coeficiente, idem anterior.

logMelFreqBand_sma[0]_quartile3 Tercer cuartil de la potencia logarítmica de la primer banda de frecuencias de Mel, suavizado por un filtro promedio móvil de ventana de longitud 3.

lspFreq_sma[4]_linregerrA Error lineal, computado como la diferencia entre los valores reales y su aproximación lineal, de pares de 8 líneas espectrales de frecuencias computados a partir de 8 coeficientes LPC (linear predictive coding.)

F0finEnv_sma_amean Media aritmética del contorno de la envolvente de la frecuencia fundamental, suavizada por un filtro promedio móvil de ventana de longitud 3.

shimmerLocal_sma_linregc1 Pendiente de la aproximación lineal de las desviaciones de amplitud locales entre períodos de pitch, suavizadas por un filtro promedio móvil de ventana de longitud 3.

Los atributos fueron seleccionados usando el método de búsqueda `GreedyStepwise`, con evaluador de atributos `ClassifierSubsetEval` y clasificador `rules.JRip`.

Resultados

Se clasificaron correctamente el **95.0673%** de las instancias, y se obtuvo la siguiente matriz de confusión:

```
=== Confusion Matrix ===

   a   b   <-- classified as
 710  44 |   a = f
  33 774 |   b = m
```

Otros experimentos conducidos

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin

fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.