



Primera Práctica Aprendizaje Automático

Tasación y Clasificación Automática de Viviendas

4º Ingeniería en Informática
Curso 2004-05 (Junio)

1. Introducción

La oferta inmobiliaria sobre la venta de pisos, proporciona un conjunto de datos que puede ser interesante tratar con el objetivo de construir un sistema que de forma automática permita clasificar la vivienda según sus características o estimar el precio más adecuado para una vivienda según el mercado, es decir hacer una tasación.

Este tipo de sistema puede resultar interesante tanto desde el punto de vista del comprador de la vivienda como desde el punto de vista del vendedor. Incluso podría resultar de ayuda a los tasadores de los bancos y a las agencias inmobiliarias:

- Estimar el precio de mercado de una vivienda es una tarea que debe realizar tanto la persona que pretende realizar una venta de la misma, como el tasador de un banco para poder ofrecer una hipoteca a un cliente, o como una agencia inmobiliaria a la hora de orientar a los clientes que ponen a la venta un piso sobre el precio más adecuado para el mismo. Para tasar una vivienda es necesario tener en cuenta tanto sus características como el precio de mercado que tienen las viviendas con unas características similares.
- Clasificar una vivienda en distintas clases que reflejen la relación calidad-precio de la vivienda, por ejemplo si la vivienda tiene un precio normal, si es demasiado cara, o por el contrario es un chollo, es más útil desde el punto de vista del comprador.

En esta práctica pretende analizarse la utilidad de alguna de las técnicas de aprendizaje automático estudiadas en teoría, aplicándolas a un dominio como el de la clasificación de viviendas, bien sea con el objetivo de estimar el precio de una vivienda que todavía no se ha tasado, o bien con el objetivo de clasificar la vivienda en clases relativas a lo adecuado de su precio. Trabajaremos con los dos casos.

2. Descripción de la Tarea

Una parte muy importante y en muchos casos laboriosa que hay que llevar a cabo para construir cualquier sistema utilizando aprendizaje automático es conseguir un conjunto de datos apropiado para entrenarlo. Para conseguir estos datos de entrenamiento es necesaria una fuente de información que los proporcione. Además, una vez que se cuenta con los datos habrá que analizarlos y, posiblemente, preprocesarlos y hacer pruebas hasta conseguir un conjunto de datos *adecuado*.

Para llevar a cabo la práctica el alumno deberá elegir un sitio Web en el que se ofrezca información sobre viviendas a la venta en una determinada ciudad que deberá elegir. Esta será la fuente de información, a partir de la cual deberá obtener los datos tanto para entrenar el clasificador, como para evaluarlo.

Entre las características que pueden ayudar a definir el precio de una vivienda pueden encontrarse las siguientes:

- La superficie de la vivienda.
- El número de habitaciones de la vivienda.
- La zona en la que la vivienda está situada.
- La antigüedad de la vivienda.
- La altura en la que está la vivienda en el edificio.
- El estado de la vivienda.
- Etc.

Aunque aquí se han definido algunas características posibles, queda a decisión del alumno la selección del conjunto definitivo, pudiendo eliminar o incluir otros atributos de los definidos anteriormente.

El correcto desarrollo de la práctica necesitará los siguientes pasos:

1. Capturar los ejemplos de entrenamiento en formato WEKA, a partir del contenido de páginas Web.
 - Creación de un directorio llamado *CasosEntrenamiento* que almacenará todos los ejemplos que se utilicen para entrenar al clasificador, es decir, las páginas Web que describan las viviendas.
 - Decidir qué atributos de las viviendas debe contener cada ejemplo de entrenamiento.
 - A partir del contenido del directorio *CasosEntrenamiento* y de los atributos que se decidan en el apartado anterior, generar el conjunto de datos de entrenamiento en formato WEKA. Este paso de puede realizar de forma manual o de forma automática. Si se implementa un método automático, la única restricción que tiene que cumplir el software que se desarrolle es que pueda ser ejecutado en las aulas del laboratorio. En cualquier caso, el conjunto de datos de entrenamiento deberá contener al menos información de 40 viviendas.
2. Utilizar WEKA para construir los clasificadores adecuados para las tareas que se han comentado anteriormente. En el caso de la tasación de viviendas, se trata de estimar el precio de la vivienda, por lo tanto, la clase será numérica. En el caso de la clasificación de viviendas según su relación calidad-precio, la clase será discreta.
 - Generar los ejemplos de test en formato WEKA, para ello se deberán conseguir las páginas Web que forman el conjunto de test de la misma forma que los casos de entrenamiento, situándolas en el directorio *CasosTest*. El fichero con formato WEKA se obtendrá de forma similar a los datos de entrenamiento (bien sea de forma manual o de forma automática). El conjunto de datos de test deberá contener al menos información de 40 viviendas.
 - Seleccionar cuáles son los atributos más relevantes.
 - Utilizar para cada tarea (tasación y clasificación) un clasificador de los estudiados en teoría que sea adecuado.
3. Repetir los pasos anteriores tantas veces como sea necesario, hasta encontrar los clasificadores adecuados, con el conjunto de atributos adecuado.

3. Entrega de la Práctica

La práctica se debe entregar antes del día **14 de abril de 2004**.

Para realizar la entrega debereis:

- Enviar un e-mail a rfuentet@inf.uc3m.es con Subject: Práctica 1 de AA, en el que aparezcan los nombres de los integrantes del grupo y en el que se manden adjuntos los siguientes ficheros:
 - Directorio de páginas Web de entrenamiento comprimido entrenamiento.tgz
 - Directorio de páginas Web de test comprimido test.tgz
 - Los ficheros con los datos de entrenamiento y los datos de test en formato WEKA. En caso de que se implemente un método automático de creación de ficheros en formato WEKA, habrá que incluir tanto los ficheros fuentes como el ejecutable y los ficheros necesarios para generar el ejecutable (makefile) si procede.
- Entregar una memoria breve (de unas 10-15 páginas) en papel con la siguiente estructura:
 1. **Introducción** (1 hoja de descripción de la práctica desde vuestro punto de vista)
 2. **Análisis de atributos y Generación de Datos:** Análisis inicial de los atributos seleccionados para caracterizar las viviendas. En caso de que se implemente, descripción del método automático para la generación de los datos de entrenamiento y test en formato WEKA.
 3. **Construcción del clasificador:** Descripción de las pruebas que se han realizado hasta conseguir el clasificador final y descripción del clasificador conseguido. El hecho de que se realicen pruebas puede suponer cambios en los atributos seleccionados, cualquier cambio que se realice se debe explicar en este apartado.
 4. **Análisis de resultados:** Análisis de los resultados sobre los conjuntos de entrenamiento y de test. Análisis acerca de la calidad real del clasificador.
 5. **Conclusiones:** En las que además de cualquier tipo de conclusión a la que se llegue tras la realización de la práctica, se debe incluir:
 - Opinión acerca del proceso de construcción de un clasificador. ¿Cuáles son las principales dificultades que has encontrado?.
 - Opinión acerca de la aplicación de este tipo de aprendizaje automático a problemas reales. ¿Se te ocurren otras aplicaciones?.
 - Opinión sobre la conveniencia de la práctica, de lo que se ha aprendido, ... etc.

Las conclusiones son una parte importante de la práctica, no las descuidéis.