Inferencia de lenguajes racionales estocásticos

Rafael C. Carrasco Jiménez

Tesis de Doctorado

Facultad: Escuela Politécnica Superior

Director: Dr. José Oncina Carratalá

1997

Universidad de Alicante
Departamento de Lenguajes
y Sistemas Informáticos

Inferencia de lenguajes racionales estocásticos

Rafael C. Carrasco Jiménez
Tesis doctoral

Mayo 1997

La presente memoria constituye la tesis doctoral presentada por Rafael C. Carrasco Jiménez para la obtención del título de Doctor en Informática y ha sido desarrollada bajo la dirección del Dr. Jose Oncina Carratalá, profesor del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante

Para Carmen.

Agradecimientos: A todos aquellos que contribuyeron al desarrollo de esta tesis, bien con su dirección (Jose Oncina), sus sugerencias (Enrique Vidal, Mikel Forcada), o su insustituible aportación técnica (Emilio Corbí, Pepe Verdú). También a todos mis compañeros del Departamento de Lenguajes y Sistemas Informáticos, donde tan buena acogida he recibido.

El hombre no se libera de la acción por no emprenderla.
Baghavad Gita

Índice General

Presentación

En esta memoria se presentan una serie de algoritmos que permiten la identificación de forma eficiente de lenguajes racionales (esto es, lenguajes reconocidos por autómatas finitos) a partir de muestras aleatorias. El significado de la inferencia, en particular para los lenguajes estocásticos, los criterios de éxito en el aprendizaje y otros conceptos básicos son presentados en el capítulo 1. El capítulo 2 contiene una introducción a los autómatas finitos estocásticos. En el capítulo 3 se discuten distintos métodos para evaluar la calidad de los modelos obtenidos. Dichos métodos miden la distancia entre las distribuciones de probabilidad propuestas y las correctas basándose en la teoría de la información. Algunos técnicas y métodos tradicionales usados en el aprendizaje de lenguajes estocásticos son revisados en el capítulo 4. En el capítulo 5 se presenta un algoritmo nuevo para la identificación de lenguajes racionales de cadenas y en el capítulo 6, otro algoritmo para la identificación de lenguajes racionales de árboles. En el capítulo 7 se explora la posibilidad de que las redes neurales recurrentes de segundo orden identifiquen de forma robusta lenguajes racionales a partir de ejemplos estocásticos. Por último se realiza una discusión sobre los problemas abiertos y las posibles continuaciones de este trabajo.