Chapter 9

Newshound Revisited: The Intelligent Agent That Retrieves News Postings

Jeffrey L. Goldberg
Analytic Services Inc. (ANSER), USA

Shijun S. Shen
Tygart Technology, Inc., USA

INTRODUCTION

There has been a lot of research done in the area of Intelligent Internet Agents. In this chapter, we would like to report our experience in implementing such an agent. It is called Newshound, and it can be trained to recognize a desired topic, and then scan Usenet newsgroups looking for new examples of that topic. Recently, Newshound has been in use by law enforcement personnel, and in response to their feedback, we have extended its capabilities. Also, we introduce two additional intelligent agents: Chathound and Webhound. Finally, we describe the inter-agent communication layer, the facilitator for cooperation between ANSER’s intelligent agents.

Organization

This chapter is organized into eight sections: (1) Introduction; (2) Newshound, describes the requirements and implementation of an Intelligent Internet Agent; (3) Chathound and Webhound, describing briefly two additional agents; (4) Inter-Agent Communications Layer, outlines the common database supporting inter-agent communications; (5) Future Work; (6) Conclusions; (7) Acknowledgments; and (8) References.
NEWSHOUND

Newshound Requirements

The original purpose of Newshound is to look for specified, trainable content in Usenet newsgroups. By specified and trainable, we mean that given a set of example postings (positive and negative), it must be able to discern a classifier function and find new postings that are "like" the positive examples. Newshound is to operate as an intelligent agent. It must allow a human agent to specify the parameters of operation, including the news server, the newsgroups in which to look, and the categories of what to be on the look-out for. After having the parameters of operation selected, the intelligent agent must then operate autonomously, only requiring interaction whenever the user desires to check the results of what has been matched so far, or to change the parameters of operation. Once a Newshound agent has found postings that match its category(s), the human agent instructs the Newshound agent as to which of the results are correct and which are not. This last requirement is called user-feedback and retraining and allows the originally learned text categorizers to be refined and personalized.

Newshound Implementation

Newshound is an intelligent Internet agent that recognizes postings of interest to a human user from Usenet newsgroups. It uses text categorization technology (Goldberg, 1996a) to train a classifier function for each desired category based on a set of examples. The classifiers, or text categorizers, are then used to recognize documents (Usenet postings) that are like the positive examples. It has been employed in a pilot program with an organization of the federal government and is being operationally tested by Special Agents.

Newshound Architecture

Newshound is a complex and dynamic software system. As shown in Figure 1, it makes a connection from the computer on which the Newshound client is running to a news server, examines the articles one by one, and compares them to its text categorizers. If they match, it takes a snapshot of the posting and stores it in a database. Newshound also makes a connection to a database server (not shown) on the local area network (LAN).

In the upper right-hand corner of Figure 1, the offline learning component is shown. This produces the classifier functions, or text categorizers. For each desired category, the input to the learning algorithm is a set of pre-labeled training documents, and the output is a text categorizer. Currently, the algorithms being applied for text categorization are batch algorithms, i.e., the categorizers are learned offline, prior to the performance part of the system when the categorizers are used
Related Content

A Survey on the use of Emotions, Mood, and Personality in Ambient Intelligence and Smart Environments
Carlos Ramos, Goreti Marreiros and Ricardo Santos (2011). Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives (pp. 88-107).
www.igi-global.com/chapter/survey-use-emotions-mood-personality/54654?camid=4v1a

ADAM: A Multi-Agent System for Autonomous Database Administration and Maintenance
www.igi-global.com/article/adam-multi-agent-system-autonomous/2387?camid=4v1a
A Dynamically Optimized Fluctuation Smoothing Rule for Scheduling Jobs in a Wafer Fabrication Factory
www.igi-global.com/article/dynamically-optimized-fluctuation-smoothing-rule/60657?camid=4v1a

Soft Computing Approaches for Human-Autonomous Agent Communication
www.igi-global.com/article/soft-computing-approaches-human-autonomous/74826?camid=4v1a