Automating the Selection of Stories for \textit{AI in the News}

Liang Dong$^{1}$, Reid G. Smith$^{2,*}$, and Bruce G. Buchanan$^{3}$

$^{1}$ School of Computing, Clemson University, SC, USA
ldong@clemson.edu
$^{2}$ Marathon Oil Corporation
rgsmith@marathonoil.com
$^{3}$ Computer Science Department, Univ. of Pittsburgh, PA, USA
buchanan@cs.pitt.edu

\textbf{Abstract.} It is relatively easy, albeit time-consuming, for a person to find and select news stories that meet subjective judgments of relevance and interest to a community. NewsFinder is an AI program that automates the steps involved in this task, from crawling the web to publishing the results. NewsFinder incorporates a learning program whose judgment of interestingness of stories can be trained by feedback from readers. Preliminary testing confirms the feasibility of automating the service to write \textit{AI in the News} for the AAAI.

\textbf{Keywords:} News crawler, machine learning, supervised classification, SVM, artificial intelligence, AAAI, AITopics.

1 Introduction

Selecting interesting news stories about AI, or any other topic, requires more than searching for individual terms. The AAAI started collecting current news stories about AI and making them available to interested readers several years ago, with manual selection and publishing by an intelligent webmaster.

Current news stories from credible sources that are considered relevant to AI and interesting to readers are presented every week in five different formats: (i) posting summarized news stories on the \textit{AI in the News} page of the AITopics web site [2], (ii) sending periodic email messages to subscribers through the “AI Alerts” service, (iii) posting RSS feeds for stories associated with major AITopics, (iv) archiving each month’s collection of stories for later reference, and (v) posting each news story into a separate page on the AITopics web site.$^{1}$

Manually finding and posting stories that are likely to be interesting is time-consuming. Therefore, we have developed an AI program, NewsFinder, that collects news stories from selected sources, rates them with respect to a learned measure of goodness, and publishes them in the five formats mentioned. Off-the-shelf implementations of several existing techniques were integrated into a working system for the AAAI.

---

$^{1}$ Anyone may view current and archived stories and subscribe to any of the RSS feeds; alerts are available only to AAAI members.
Traditional recommender systems [9] require recording a user’s preference and using techniques such as non-negative matrix factorization [12] to find users with similar tastes. Then, recommendations are based on the preferences of similar users. In our approach, we learn the characteristics of the items preferred by users and classify new items with respect to those.

The NewsFinder Program

The work of NewsFinder is implemented in four loosely-coupled program modules as in Fig. 1: (A) Crawling; (B) Training; (C) Ranking; (D) Publishing. The first three are independent from each other and the last two usually run together.

Fig. 1. NewsFinder Procedure Diagram

1.1 Crawling

In the crawling phrase, the program collects a large number of recent news stories about AI. Since crawling the entire web for stories mentioning a specific term like ‘artificial intelligence’ brings in far too many stories, we restrict the crawling to about two dozen major news publications. This makes a story more credible and more likely to interest an international audience. The system administrators (AI subject matter experts) maintain a list of news sources, chosen for their international scope, credibility, and stability. These include The BBC, The New York Times, Forbes, The Wall Street Journal, MIT Technology Review, CNET, Discovery, Popular Science, Wired, The Washington Post, and The Guardian. Others can be added to the list.

NewsFinder collects the latest news stories via either in-site search or RSS feeds from those sources and filters out blogs, press releases, and advertisements. If a source