

This paper appears in the publication, Data Mining and Knowledge Discovery Technologies edited by D. Taniar © 2008, IGI Global

Chapter XIV

A Machine Learning Approach for One-Stop Learning

Marco A. Alvarez, Utah State University, USA SeungJin Lim, Utah State University, USA

Abstract

Current search engines impose an overhead to motivated students and Internet users who employ the Web as a valuable resource for education. The user, searching for good educational materials for a technical subject, often spends extra time to filter irrelevant pages or ends up with commercial advertisements. It would be ideal if, given a technical subject by user who is educationally motivated, suitable materials with respect to the given subject are automatically identified by an affordable machine processing of the recommendation set returned by a search engine for the subject. In this scenario, the user can save a significant amount of time in filtering out less useful Web pages, and subsequently the user's learning goal on the subject can be achieved more efficiently without clicking through numerous pages. This type of convenient learning is called one-stop learning (OSL). In this chapter, the contributions made by Lim and Ko (2006) for OSL are redefined and modeled using machine learning algorithms. Four selected supervised learning algorithms:

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

334 Alvarez & Lim

support vector machine (SVM), AdaBoost, Naive Bayes, and Neural Networks are evaluated using the same data used in Lim et al. (2006). The results presented in this chapter are promising, where the highest precision (98.9%) and overall accuracy (96.7%) obtained by using SVM is superior to the results presented by Lim et al. (2006). Furthermore, the machine learning approach presented here demonstrates that the small set of features used to represent each Web page yields a good solution for the OSL problem.

A Machine Learning Approach for One-Stop Learning

Using the Web, a global repository of information, for educational purposes requires more accurate and automated tools than general-purpose search engines. Innovative tools should support the learning experience and focus the attention of the learner on his or her desired target subject. A typical learner would be interested in going directly to the point and learn without spending time with useless or non-informative pages. In this context, harvesting the Web using current search engines and technologies, however, looking for concepts, subjects, or general information usually imposes significant overhead that is denoted when the user spends time in filtering irrelevant pages or when he or she is simply distracted with advertisements, latest news, or attractive but not suitable Web sites for learning.

Before the advent of the Web, students and occasional learners studied new subjects by reading books or well-known articles in which they could find all the required information. Certainly, these primary sources of information can be considered adequate and sufficient for learning the subject when the learner satisfies his or her aspirations with them. In most cases, there is no need to look for additional resources for the same subject. This conventional learning strategy is called one-stop learning (OSL) in Lim and Ko (2005). On the other hand, when considering the Web as a repository for learning, the learners very often rely on available general-purpose search engines like Google, Yahoo, or Microsoft Live in finding suitable materials for OSL. Here, it must be emphasized that these search engines were not designed with the specific goal of assisting educational activities. The use of such engines for one-stop learning needs to be revisited in order to optimize the time that learners spend searching for self-contained sources of knowledge/information. One clear advantage of existing search engines is the fact that they maintain billions of updated pages already indexed for fast search and retrieval. Previously proposed strategies for OSL using the Web take advantage of the results returned by search engines (Lim et al., 2005, 2006). The major motivation in this chapter is to present a machine learning approach for the OSL problem making use of existing search

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/machinelearning-approach-one-stop/7523

Related Content

Applying Dynamic Causal Mining in E-Government Modeling

Yi Wang (2009). Social and Political Implications of Data Mining: Knowledge Management in E-Government (pp. 200-220). www.irma-international.org/chapter/applying-dynamic-causal-mining-government/29072

Classification and Visualization of Alarm Data Based on Heterogeneous Distance

Boxu Zhaoand Guiming Luo (2018). *International Journal of Data Warehousing and Mining (pp. 60-80).*

www.irma-international.org/article/classification-and-visualization-of-alarm-data-based-onheterogeneous-distance/202998

A Novel Hybrid Algorithm Based on K-Means and Evolutionary Computations for Real Time Clustering

Taha Mansouri, Ahad Zare Ravasanand Mohammad Reza Gholamian (2014). International Journal of Data Warehousing and Mining (pp. 1-14). www.irma-international.org/article/a-novel-hybrid-algorithm-based-on-k-means-and-evolutionarycomputations-for-real-time-clustering/116890

Analytical Processing Over XML and XLink

Paulo Caetano da Silva, Valéria Cesário Times, Ricardo Rodrigues Ciferriand Cristina Dutra de Aguiar Ciferri (2012). *International Journal of Data Warehousing and Mining (pp. 52-92).*

www.irma-international.org/article/analytical-processing-over-xml-xlink/61424

From Data to Knowledge: Data Mining

Tri Wijaya (2013). Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1150-1162).

www.irma-international.org/chapter/data-knowledge-data-mining/73489