Cognitive Location-Aware Information Retrieval by Agent-Based Semantic Matching

Eddie C. L. Chan, The Hong Kong Polytechnic University, China
George Baciu, The Hong Kong Polytechnic University, China
S. C. Mak, The Hong Kong Polytechnic University, China

ABSTRACT
This paper proposes semantic TFIDF, an agent-based system for retrieving location-aware information that makes use of semantic information in the data to develop smaller training sets, thereby improving the speed of retrieval while maintaining or even improving accuracy. This proposed method first assigns intelligent agents to gathering location-aware data, which they then classify, match, and organize to find a best match for a user query. This is done using semantic graphs in the WordNet English dictionary. Experiments will compare the proposed system with three other commonly used systems and show that it is significantly faster and more accurate.

Keywords: Agent, Information Retrieval, Location-Aware, Semantic, TFIDF

INTRODUCTION
One of the most challenging problems in retrieving the location-aware information is to understand the behavior of users and how it suits the current location. Wireless tracking applications are popular ways to help user navigate that may make use of current location-aware information. However, very often, the information retrieved in common search engines is both excessive and unstructured from the user’s point of view. The basic requirement of an effective location-aware retrieval system is that it should match user queries and provides accurate information where users access location-aware (e.g., pervasive computing-enabled) applications and services. The retrieval system should do this in an organized and efficient way.

In recent years, researchers have focused on how to provide higher accuracy and faster retrieval by making use of keywords and textual semantics. However, the results so far have been unsatisfactory. First, the truth-conditional semantics that are often applied provide only a very limited account of meaning. Second, information is derived from few sources and the information from those sources is not structured. In these circumstances, the value of the
information that can be extracted is very limited. Third, information collection is an expensive, time-consuming process that is often carried out manually. This makes it very difficult to build, maintain and grow comprehensive databases. Forth, some approaches, such as Naïve-Bayes (Danesh et al., 2007) and K-Nearest Neighbors (K-NN) classifiers (Weiss et al., 1996) machine learning approaches ignore the semantic meaning in text classification. This leads to inadequate search results. Finally, location-aware information is distributed in different locations and some information, traffic information for example, they will quickly become out-of-date so location-aware information must be updated frequently.

This paper proposes an agent-based semantics retrieval system for location-aware information. The proposed system uses the WordNet (WordNet, 2004) dictionary to construct the semantics graph structure of a location. This graph structure sets the weight values between edges and nodes (words) of the semantics and the classic information retrieval, term-frequency-time-inverse-document-frequency (TFIDF) technique is modified with semantics weight values. As the location-aware information is distributed in different locations, the approach also implement agents using the IBM Aglets Agent Software Development Kit (IBM, 2004). These mobile network agents are programs that can be dispatched from one computer and transported to a remote computer for execution. Arriving at the remote computer, they present their credentials and obtain access to local services and data. The remote computer may also serve as a broker by bringing together agents with similar interests and compatible goals, thus providing a meeting place at which agents can interact. The proposed system uses four types of agents: one each to gather, classify, match, and organize information.

The proposed system offers a number of benefits. First, the proposed semantics graph structure provides a hierarchical structure for the location-aware information. Second, the proposed agent system obviates the need for extensive manual information grasping. Third, the modified TFIDF with semantics weights improves the accuracy of matching keywords and provides a more meaningful result from the point of view of semantics. Forth, the agent can update information directly by communicating with its neighbor agents. Finally, it is fast and cost-effective.

The rest of paper is organized as follows: Section 1 presents the related work of information retrieval techniques and agent-based information systems. Section 2 describes cognitive semantics TFIDF technique in our system. Section 3 presents the system implementation and architecture of agent-based cognitive semantics retrieval system. Section 4 describes the experiment setup and result. Section 5 presents the case study. Finally, Sections 6 offers our conclusion and future work.

1. RELATED WORK

In this section, we summarize current research works of the information retrieval techniques, agent-based information retrieval systems and positioning system in the area of cognitive informatics and cognitive computing (Baciu et al., 2009; Wang, 2003, 2007, 2009).

1.1 Information Retrieval Techniques

There are a number of information retrieval techniques. These include, Naïve Bayes classifier (Lewis & Ringutte, 1994), linear or quadratic discriminated analysis (LDA or QDA) (Hull et al., 1996), artificial neural network (Tan et al., 2006), back-propagation-based technique (Nawi et al., 2008) and C4.5 (Ruggieri, 2002) rule learning technique. Naïve Bayes classifier (Lewis & Ringutte, 1994) suffers from its assumption that words are independent of each other and are, therefore, less accurate than a more complex model. While standard parametric methods such as LDA, QDA (Hull et al., 1996) and regression are mostly criticized as being dependent on strong assumptions...
Related Content

Protoforms of Linguistic Database Summaries as a Human Consistent Tool for Using Natural Language in Data Mining
[www.irma-international.org/article/protoforms-linguistic-database-summaries-human/2788/](www.irma-international.org/article/protoforms-linguistic-database-summaries-human/2788/)

Online Machine Learning
[www.irma-international.org/chapter/online-machine-learning/76468/](www.irma-international.org/chapter/online-machine-learning/76468/)

Computer-Aided Detection of Polyps in CT Colonography by Means of Feature Selection and Massive-Training Support Vector Regression
[www.irma-international.org/chapter/computer-aided-detection-polyps-colonography/62230/](www.irma-international.org/chapter/computer-aided-detection-polyps-colonography/62230/)

An Action Guided Constraint Satisfaction Technique for Planning Problem

Toward an Agent Framework that Assists the Individuals with Mild Memory Impairment