13th Global Congress on Manufacturing and Management, GCMM 2016

Research on search engine of knowledge adaptation system based on large scale data set

Zhou Zhiwei\(^a\), Guo Xiaomin\(^a\), Wei Lin\(^a\), Zhang Mengmeng\(^a\), Lian Zhikang\(^a\), Zhao Yamin\(^a\), Zhang Jianhua\(^*\)

*Zhengzhou University, No.100 Science Avenue, Zhengzhou 450001, China

Abstract

The world has entered the era of knowledge economy, knowledge is becoming the core of social production. Any organization or individual can only continue to strengthen the application of knowledge, in today's changeable, fast changing, change in the competition to stand out. Provides a rich source of data explosion of information and knowledge for knowledge application, knowledge management research direction of past and existing knowledge, needs the shortcomings, which introduced the large-scale data adapter based on knowledge. Large data sets in the memory not only need to consume a large amount of storage space, but also put forward higher requirements for data retrieval, and the existing search engine is not very good to meet the needs of knowledge adaptation system. Based on large data sets, this paper mainly studies the retrieval system of knowledge adaptation system. The selection of the algorithm for the search engine, the analysis of the deficiencies, and put forward a targeted improvement program-kNN algorithm to increase the pre process, expand the scope of application and effect of the engine. Finally, the efficiency and effectiveness of the model are verified by the example of forest cover.

© 2016 The Authors. Published by Elsevier Ltd. Peer-review under responsibility of the organizing committee of the 13th Global Congress on Manufacturing and Management.

Keywords: Search Engine; kNN Algorithm; Rough Set; Variation Coefficient Method

1. Introduction

In recent years the global acceleration into the era of Knowledge Economy\(^{[1,2]}\), the most valuable production
resources in social production activities has been turned into Knowledge. The rapid development of information technology has brought the explosion of data volumes, but due to the amount of data on a huge scale, large data sets not only need to consume large amounts of storage space in the storage, at the same time, the data matching and retrieval are there will be a bottleneck.

In view of this, we proposed the knowledge search engine adaptation system based on large-scale data sets. Among them, the author of searching engine to do some modification, such as using the rough sets and the variation coefficient method of combining the distance formula of kNN algorithm used by the condition attribute weights allocation, make it more relevant applications. Use k - means to do division algorithm to accelerate the efficiency of search engine, etc. Therefore, this article research results have certain practical value, improvements to the enterprises in knowledge management, knowledge applications link has a certain reference value.

2. Search engine

Search engine is the foundation of the whole knowledge adaptation system stage, efficient and accurate retrieval out with unanswered questions matching degrees higher case set has important influence on the quality of the fit for follow-up, now commonly used retrieval algorithm mainly are: nearest neighbor, inductive reasoning and knowledge guidance strategy three categories.

Table 1. Comparation of common case retrieval algorithm.

<table>
<thead>
<tr>
<th></th>
<th>kNN</th>
<th>Inductive reasoning strategy</th>
<th>Knowledge guidance strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior training</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Training time-consuming</td>
<td>No</td>
<td>Longer</td>
<td>Medium</td>
</tr>
<tr>
<td>Retrieval efficiency</td>
<td>Medium</td>
<td>Faster</td>
<td>General</td>
</tr>
<tr>
<td>Algorithm complexity</td>
<td>Easy</td>
<td>More complex</td>
<td>Complex</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Poor</td>
<td>Strong</td>
<td>General</td>
</tr>
<tr>
<td>Implementation difficulty</td>
<td>Easy</td>
<td>General</td>
<td>Difficult</td>
</tr>
</tbody>
</table>

Although some system design of case retrieval and storage system when they do different optimization, but the overall strategy still belongs to the category of the above three methods. According to the comparison table 1 case retrieval algorithm and the design features of massive knowledge adaptation framework, this paper, the author selected kNN algorithm for case retrieval.

3. Adapter case sets determined based on large data sets

3.1. Case similar level characterization and method

In order to characterize the target case \( C_q \) and case similarity between case, usually the condition attribute values ofthe two after the similarity calculation. Commonly used the similarity calculation of Euclidean distance, min had distance and vector inner product\([3-5]\), but generally adopts Euclidean distance to do calculation. For under test case \( C_q \) (no solution to the target case temporarily), its conditions set \( a_{q-j} (j = 1, 2, \ldots, m^q) \) Which the number \( m^q \) for cases under the condition of limited conditions. By comparing the retrieval algorithm source in case similarity \( sim(C_i, C_q) \) between the case \( C_i \) and case under test \( C_q \). If \( \exists sim(g) \in sim(C_i, C_q) \) and \( sim(g) = 1 \) set up in the case-based reasoning, to retrieve cases in the case bank indicates that the source \( C_i \) and target case \( C_q \) are identical in all conditions, at this time may consider direct reuse in solving unanswered question \( C_q \), namely "empty adaptation".
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات