Title	Studies on Knowledge Discovery and Its Applications based on Mining Techniques
Author(s)	裴, 明花
Citation	大阪大学, 2008, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/49287
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

論 文 内 容 の 要 旨

In recent years, the scale of data stored in databases has become so large that it is inapplicable to statistical analysis.　Data mining can extract knowledge from huge-scale databases including data that cannot be processed by means of statistical analysis.　Up to now, much work has been done on data mining, so that various algorithms have been proposed considering the characteristics of various data sources, such as enterprise data, Web data and text data.

However, many problems still remain while applying these algorithms to real world data.　For example, when association rules are extracted by having value of a criterion, support, larger than a threshold, rules having high support only represent general knowledge, and there are probably so many useless rules among the rules having low support.　Thus, it is difficult to distinguish important rules for a user having not enough background knowledge.　Another example is knowledge extraction for Semantic Web (SW).　The information processing can be conducted by inferring the relations among the SW resources because the SW resources and the relations among them are defined well.　However, in real world SW, much of the Web resources cannot be matched by inferring the explicitly defined relations since it is impossible to define all the relations.　Thus there are some serious issues, such as the discovery of potential relations for incomplete ontologies and efficient ontology mapping, for using SW resources efficiently.

In this thesis work, some methods to solve the problems in the process of knowledge discovery for real data are proposed, i.e., a method of distinguishing important association rule mining results, a method of categorizing SW resources by discovering potential relations from incomplete ontologies, and a new method of constructing global ontology using Wikipedia Thesaurus.

This thesis consists of five chapters.　In Chapter 1, we introduce the background and motivation of our research, and present the contribution and organization of the thesis.

In Chapter 2, we propose a new criterion for association rule mining named R-repeat, so as to present

differences in customer behavior in association rules when conducting association rule mining to purchase history data. It contributes to discovery of important association rules with high customer recognition by adopting the new criterion R-repeat to conventional methods, which only needs a little additional computation. Moreover, it successfully discovers a valuable customer segment in which the customers repeat purchasing the products in the rules with their preferential products.

In Chapter 3, we propose a categorization method for SW resources for an incomplete ontology. The categorization method consists of four processes, which extract not only the relations defined explicitly in the existing ontology but also the potential relations discovered from it. Though these four methods are not very novel, the design of combination, optimization and work flow for Web resource categorization is the main contribution of this chapter. The effectiveness of the proposed categorization method is confirmed by comparing search results to those of a class tree specification based search.

In Chapter 4, we propose a method of constructing a large-scale ontology, which can be used as an intermediator for ontology mapping. It contributes to a practical approach for ontology mapping by applying intermediator based ontology mapping. We constructed a global ontology using Wikipedia knowledge, and investigated some factors affecting the accuracy of the constructed ontology. A relation detective process is also performed for ontology construction. It represents the possibility and capability of using Wikipedia knowledge for efficient use of SW data.

Finally, we conclude the thesis and present our future work in Chapter 5.

論 文 審 査 の 結 果 の 要 旨

近年、データベースの大規模化とともに、様々なデータソースに対し、マイニング技術を適用した知識発見に対する需要が高まっている。しかし、マイニング技術において数多くの研究が行われているにも関わらず、実応用のための知識発見において、重要な課題が残されている。本論文は、POS データとウェブデータにおけるマイニング技術の適用による知識発見およびその応用に関する研究成果をまとめたものである。その主要な成果を要約すると次の通りである。

⑴相関ルールマイニングにおいて、顧客の購買行動における繰り返しを考慮した評価基準を適用することにより、顧客に重視されるルールを判別する方法を提案している。

⑵実セマンティックウェブにおける意味情報の定義が不足している問題に対して、実データの特徴を考慮し、意味関係の推論や、知識発見の既存手法とその拡張手法を用いて、ウェブリソースの分類方法を提案している。

⑶大規模関連シソーラスの特徴を活かした隠れリレーションの抽出によるグローバルオントロジの構築方法を提案し、様々な影響要素について分析し、グローバルオントロジの構築可能性を示している。

以上のように、本論文はマイニング技術による知識発見および応用に関する成果を挙げた研究として、情報科学に寄与するところが大きい。よって本論文は博士（情報科学）の学位論文として価値あるものと認める。