

## データマイニングを用いた知識共有化に関する研究

|      |   |
|------|---|
| 著者   | 上田 芳弘   |
| 著者別名 | Ueda, Yoshihiro   |
| 雑誌名  | 博士学位論文要旨 論文内容の要旨および論文審査結果の要旨 / 金沢大学大学院自然科学研究科                                   |
| 巻    | 平成14年9月   |
| ページ  | 133-142   |
| 発行年  | 2002-09-01  |
| URL  | <a href="http://hdl.handle.net/2297/16429">http://hdl.handle.net/2297/16429</a> |

|            |                             |
|------------|-----------------------------|
| 氏名         | 上田 芳弘                       |
| 生年月日       |                             |
| 本籍         | 石川県                         |
| 学位の種類      | 博士(工学)                      |
| 学位記番号      | 博甲第450号                     |
| 学位授与の日付    | 平成13年9月28日                  |
| 学位授与の要件    | 課程博士(学位規則第4条第1項)            |
| 学位授与の題目    | データマイニングを用いた知識共有化に関する研究     |
| 論文審査委員(主査) | 木村 春彦(自然科学研究科・教授)           |
| 論文審査委員(副査) | 船田 哲男(工学部・教授) 村本健一郎(工学部・教授) |
|            | 中山 謙二(工学部・教授) 西川 清(工学部・教授)  |

## 学位論文要旨

### Abstract

In this paper, the data mining is applied to two problems of the knowledge sharing. In the first application, we pay attention to the supporting of the document revising as the subject of the knowledge's externalization that converts the tacit knowledge into the explicit knowledge. The rules for the revising are extracted by applying the association rule to the document revising data that externalize the tacit knowledge of human reviser. In the process of this extraction, the adaptive degree is introduced into the rules for not overlooking low frequent revising. In addition, it becomes possible to revise the documents depending on the domain by learning the adaptive degree in each domain. It is confirmed that the documents are revised based on the notation, the grammar and the meaning, besides the organizational information.

In the second application, we pay attention to the transmission of the operation techniques for the machine tool as the subject of the knowledge's socialization that transmits the tacit knowledge without conversions. The knowledge is extracted and shared with operators by imitating methods that experts set the machine. First, we examine the indispensable data for the imitation besides propose a method combined the AHP with the entropy-model for making much of human subjects. According to the proposed method, the tacit knowledge of the experts is collected without disturbing their operations, and

when operators set the machine, the indispensable knowledge is estimated for the duration of trial-and-error.

Consequently, the novel methods in the data mining for the knowledge's externalization and socialization are proposed and evaluated in the problems of the office work and the manufacturing process.

## 1 はじめに

データマイニングは、最近様々な分野においてその有効性が認識され始めている。本研究では、まずその代表的な手法である相関ルールが、パターン認識の問題である居住者の活動認識に有用であることを示す。ここでは、相関ルールの抽出アルゴリズムに正例と負例の概念を導入し、これに対応する正と負のルールを抽出できるようにする。これらのルールにより認識結果を補正でき、認識率の向上が可能であることを示し、データマイニングの有効性を検証する。

次に、データマイニングを知識共有化の2つの問題に応用することを検討する。第1の応用では、暗黙知から形式知へ変換することによる知識の「表出化」の問題として文書校正支援を取り上げる。ここでは、校正者の暗黙知が表出した結果と考えられる文書校正データに相関ルールを応用し、校正ルールを抽出する。このとき、校正全体から見ると出現頻度の少ない校正も見逃さずに、ルールとして抽出するために、適用度と呼ぶ概念を導入する。また、文書の分野や業務の種類、すなわちドメインごとにこの適用度を学習することによって、ドメイン別に異なる校正ができることを示す。さらに、実際の校正者が行った校正と抽出したルールによる校正を比較し、提案手法により表記や文法、意味レベルの校正のみならず、組織情報に基づいた校正も可能であることを示す。

第2の応用では、暗黙知を暗黙知のままに伝承する知識の「共同化」の問題として工作機械操作方法での知識共有化を取り上げる。ここでは、熟練者による加工条件の設定方法を模倣することにより暗黙知を抽出し、共有化することに着目する。まず、この模倣に必要な最小限の情報を検討し、さらに人間の主観的な判断を重視できるように AHP とエントロピーモデルを組み合わせた手法を提案する。これにより、熟練者の作業を妨害することなく、暗黙知を含んだ有益な情報を収集すること、及び収集した情報の中で試行錯誤中の作業が必要とする知識をマイニングすることが可能であることを示す。

## 2 データマイニングの有効性検証

最初にデータマイニングの有効性を検証するために、代表的な手法である相関ルールを用いて、居住者の活動認識、すなわちパターン認識で認識率の向上を目指す。ここでは、相関ルールの最も基本的なアルゴリズムである Apriori に正例と負例という概念を導入し、これに対応する正と負のルールを抽出できるようにした。ここで、正例は正しく認識できた事例であり、負例は認識を誤った事例を意味する。さらに、このルール抽出時に、通常用いられる最小支持度、最小確信度に加え、「正の確信度」と呼ぶパラメータを導入し、式(1)のように定義する。

$$\text{正の確信度} = \frac{\text{support}(X \Rightarrow Y(+))}{\text{support}(X \Rightarrow Y)} \quad (1)$$

ここで、 $X \Rightarrow Y(+)$  は、 $X$  と  $Y$  を同時に含み、かつ正例であることを表す。この正の確信度は、生成されたアイテム集合中の正例の割合を示す指標で、0～1までの値を取る。なお、この正の確信度は、1に近ければ抽出した相関ルールが正のルール、逆に0に近ければ負のルールであることを示し、0.5に近ければ正とも負とも判断できない曖昧なルールとなる。また、このパラメータも最小支持度、最小確信度と同様にユーザがその閾値を与えるものとする。

以上のように抽出した正と負のルールに観測したにおいセンサ出力をマッチさせることで、正のルールにマッチした事例は認識は正しく行われたものと考えられ、逆に負のルールにマッチした事例は認識が誤った事例と考えられる。よって、これらのルールにより認識結果を補正することができる。

提案手法の評価実験を行った結果、図1に示したように LBG による認識率と提案手法に

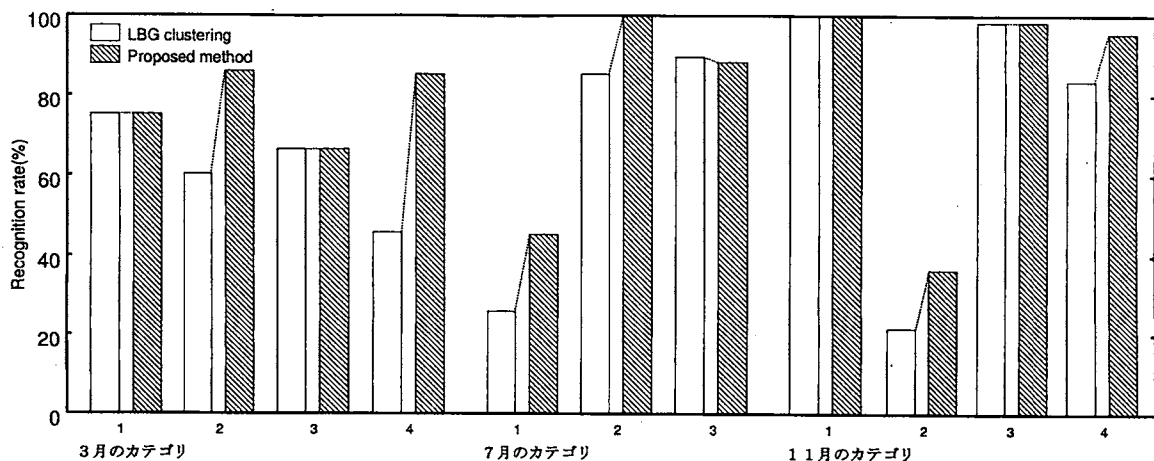


図1 LBG と提案手法による認識率

よる認識率を比較すると、11カテゴリ中で6カテゴリにおいて提案手法により認識率が向上することが分かった。また、認識率に変化が見られなかったカテゴリでは負例が不足しているため、負のルールが抽出できず、認識結果の補正が行えなかったことが分かった。

以上、パターン認識の問題においてデータマイニングの有効性を検証できた。

### 3 暗黙知から形式知への変換による知識共有化

暗黙知を形式知に変換する問題として、組織内で行われる文書校正作業の支援を取り上げた。まず、校正履歴を文書ファイルのバージョン管理により蓄積し、2つのバージョンのファイルに形態素解析を行い、単語に分割する。これらのファイルの差分から文書校正データを生成する。そのデータ構造は、置換、挿入や削除された単語と文書上でのその位置が分かる構造とした。

このような文書校正データに相関ルールを応用して校正ルールを抽出する。校正ルールは式(2)のように定義した。

$$\begin{aligned} & \{target\} \{item_1, item_2, \dots, item_n\} \\ & \quad \{rate_1, rate_2, \dots, rate_m\} \\ & \Rightarrow \{action\} \{revised\} \end{aligned} \quad (2)$$

1つのルールは、 $\Rightarrow$ を区切りとして左辺の条件部と右辺の行動部からなる。左辺の条件部では校正のターゲットとなる単語  $\{target\}$  と、校正された文書中でこの  $\{target\}$  付近で出現し、かつ相関ルールの手法でその組み合わせが有意であるとされた単語の集合  $\{item_1, item_2, \dots, item_n\}$  と、適用度  $\{rate_1, rate_2, \dots, rate_m\}$  で構成している。次に、行動部は置換、削除、挿入を表す  $\{action\}$  と、 $\{target\}$  を置換する単語または挿入する単語を表す  $\{revised\}$  で構成している。

ここで、相関ルールで用いられる支持度と確信度では、同じ校正が何度か出現しないと有意なルールとされない。ところが、実際の校正作業を考えると、出現頻度は低いながら、重要度の高い校正が存在する。このため、ルールの適用度と呼ぶ指標を導入し、相関ルールの抽出時にこれを利用することを提案する。 $rule_i$  のドメイン  $j$  におけるルール抽出時の適用度  $rate_j$  は、式(3)で与えられるものとする。

$$rate_j = \frac{\text{実際に校正が行われた事例数}}{\text{rule}_i \text{の条件部を満足する事例数}} \quad (3)$$

このような適用度  $rate_j$  を相関ルールの抽出時に支持度と同時に求めることで、たとえ支持度や確信度が設定した最小値より低くとも適用度が高ければ有意なルールとする。さらに、前述のように文書のドメインによってルールの有意性が変化するものと考えられるので、抽出したルールの適用度を文書のドメインごとに学習する。

以上のような提案手法により、ドメインを無作為に 430 個の文書ファイルを選択し、校正ルールを抽出した結果、文の表記や文法、意味レベルの校正を行うルールとともに、組織情報に基づいた校正を行うルールが抽出できることを確認できた。さらに、コンピュータ関連の文書で適用度を学習させた後、校正前の 50 個の文書ファイルで校正者が行った校正とルールをマッチさせた結果を比較した。その結果、図 2 に示したように適用度のしきい値が 0.5 以下のときにヒット率は 0.795 を示し、実際の校正者が行った校正の 80% は提案手法で抽出したルールによって校正可能であることが分かった。また、適用度のしきい値を 0.5 より大きく設定すると、適用できるルール数は減少するため、ヒット率も低下することが分かった。一方、対象とした文書にマッチしたルールの行動部のうち、校正者が行わなかった校正を表すものの件数を検討した。この評価も前述と同様に適用度のしきい値を変化させて行い、その結果、適用度のしきい値が 0.5 以下では、誤った校正件数は 174 件であり、しきい値が大きくなるに従って、この件数は減少することが分かった。以上のように適用度のしきい値は 0.5 以下の小さな値に設定した方が、実際の校正者が行う校正とのヒット率が高い校正が行えるが、その反面、誤った校正も多くなる。本評価実験

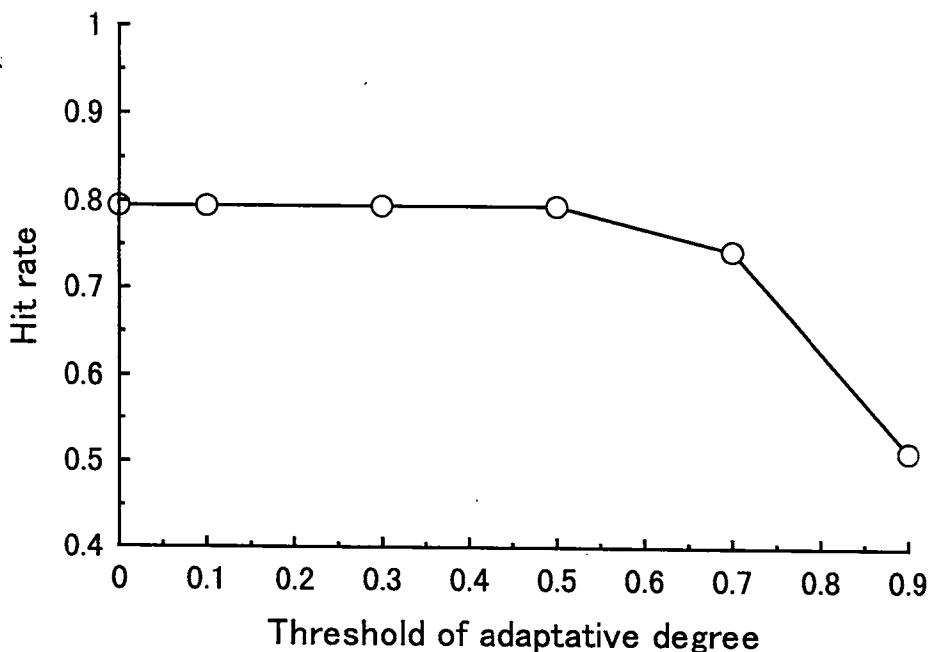


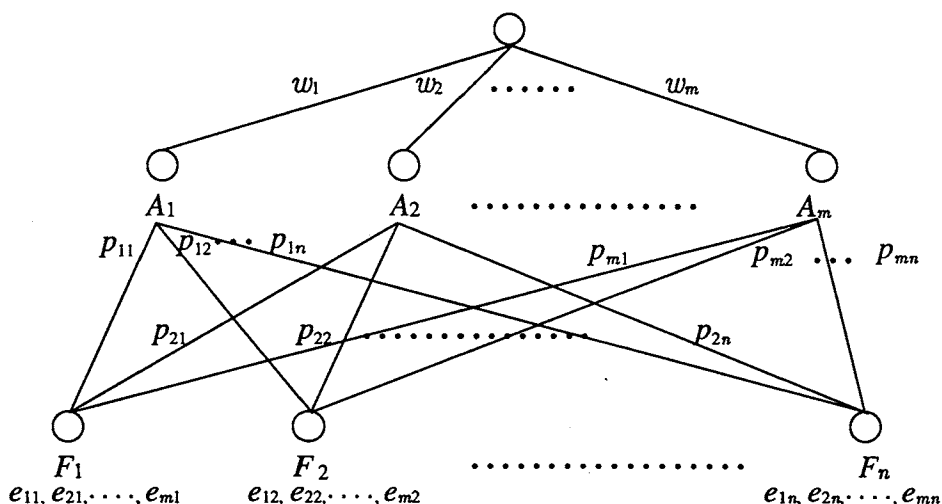
図 2 適用度のしきい値とヒット率の関係

では、この両方の側面を考慮して、適用度のしきい値は0.7とすることが妥当だと考える。

以上により、提案手法によって実際の校正者が行う校正とほぼ適合した校正ルールを抽出できるといえる。さらに適用度のしきい値を適切な値に設定することによって誤った校正の件数を減少させることができるといえ、提案手法の有効性を示すことができた。

#### 4 暗黙知のままでの抽出による知識共有化

暗黙知のまま知識を抽出し、共有する問題として、工作機械操作方法の伝承を取り上げた。ここでは、熟練者の装置の設定方法を模倣することに着目し、(1) 熟練者が暗黙知として持つ模倣情報を容易に収集できる手法、(2) 作業者が直面している問題と関連性の高い模倣情報を効率良く推定して提示する手法を検討した。これにより、熟練技能には達していない作業者の試行錯誤の回数を削減することを目指す。



$A_i$ : 評価項目,  $w_i$ :  $A_i$ の重み,  $F_j$ : 作業履歴データ,  
 $p_{ij}$ :  $F_j$ の選択確率,  $e_{ij}$ :  $A_i$ から見た $F_j$ の主観評価値

図3 模倣情報の推定

まず、熟練者の通常作業中に、その作業を妨害することなく、模倣に必要な最小限の情報を収集することとし、模倣情報を式(4)のような構造に定義した。

$$F_j = [c_{1j}, \dots, c_{pj} \mid s_{1j}, \dots, s_{qj} \mid e_{1j}, \dots, e_{mj}] \quad (4)$$

$c_{hj}$ : 環境条件,  $s_{kj}$ : 設定値の変動量,  $e_{ij}$ : 評価値

次に、図 3 に示したように、作業者の試行錯誤中の方針決定を階層化意志決定法 AHP を用いて、あらかじめ設定した評価項目  $A_i$  の重み  $w_i$  として推定する。また、熟練者から収集した模倣情報  $F_j$  に含まれる主観評価値  $e_{ij}$  にエントロピーモデルを応用して、評価項目  $A_i$  から見た  $F_j$  の選択確率  $p_{ij}$  を推定する。最後に、式 (5) で、試行錯誤中の作業者が  $F_j$  を選択する好ましさを表す指標  $p_j$  を求める。

$$p_j = \sum_{i=1}^m w_i \cdot p_{ij} \quad (5)$$

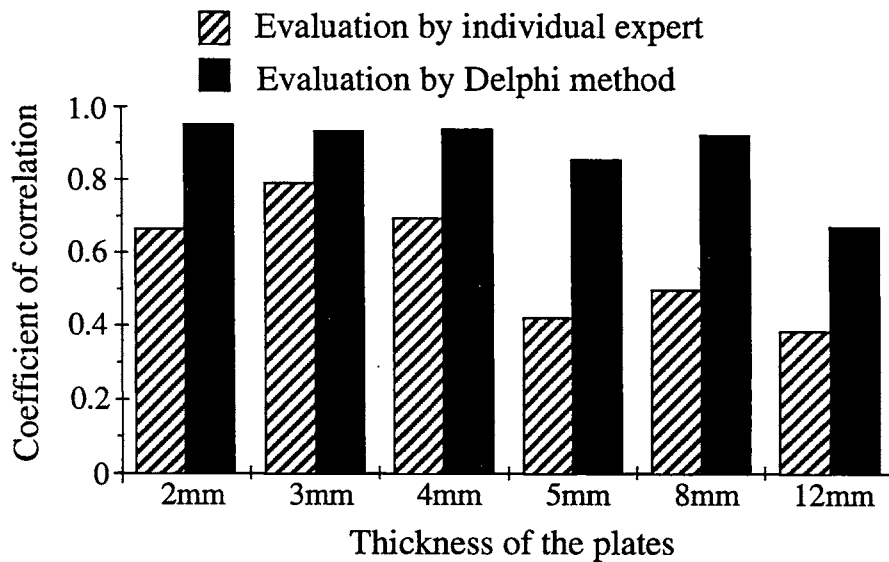


図 4 主観評価値と測定評価値の相関

なお、熟練者から模倣情報を収集するとき、主観評価値  $e_{ij}$  の信頼性を向上するために、デルファイ法を応用した。すなわち、複数の熟練者に同一の試行結果を主観評価してもらい、その評価値の分布を熟練者にフィードバックして、評価を繰り返すことにより、評価値が収れんすることが期待できる。

提案手法の有効性を評価するために、レーザ加工機の加工条件設定タスクに提案手法を実装したシステムを応用した。その結果、まず、デルファイ法により収集した熟練者の主観評価値が収れんできることを確認できた。さらに、図 4 に示したように、評価項目の 1 つであるドロスの幅を測定し、この測定評価値との相関分析を行った。熟練者単独の主観評価値とデルファイ法で収れんした主観評価値での相関係数を比較すると、実験を行った結果すべてについて、デルファイ法による主観評価値の方が、相関係数が高い値を示すことを確認した。よってデルファイ法により主観評価値の信頼性が向上できたといえる。



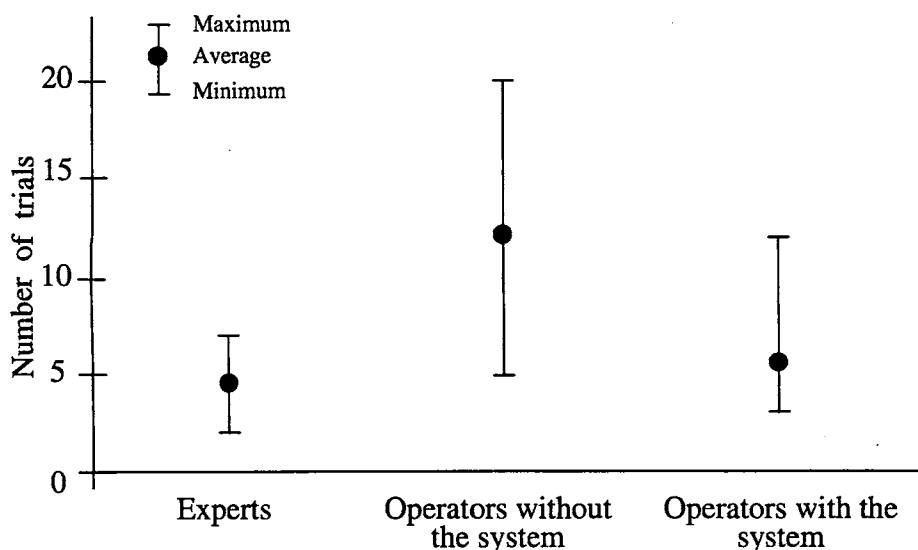


図5 平均試行回数

さらに、提案手法により推定した模倣情報について熟練者にアンケート調査を行ったところ、良好な結果を得ることができ、その妥当性を確認できた。最後に、熟練者でも数回の試行を必要とする加工問題を設定し、熟練者4名と作業員4名の試行回数を比較した。その結果、図4に示したように、熟練者は、平均4.8回の試行によって目標に到達した。これに対して、作業員4名が提案手法を実装したシステムを使用せずに作業を進めた場合は、平均12.3回の試行を必要とした。しかし、同じ作業員4名が本システムを使用すると、平均6.0回の試行で目標に到達でき、51%の回数削減が達成できた。さらに、図5において目標に到達するまでの最大試行回数と最小試行回数を比較すると、熟練者ではその差が5回であり、本システムを使用しなかった場合の作業員では15回で、3倍の開きがある。このことから熟練者の方が安定してレーザー加工を実施していると推察できる。作業員が本システムを使用した場合の最大と最小試行回数の差は、9回に減少し、作業の安定性も熟練者に近づけたと考えられる。以上により、本研究で提案した手法によって、レーザー加工機の加工条件設定における試行を支援できたといえる。

## 5 まとめ

本研究では、データマイニングを用いた知識共有化について考察した。

まず、データマイニングの代表的な手法である相関ルールが、居住者の活動認識に有用

であることを示した。ここでは、相関ルールの抽出アルゴリズムに正例と負例の概念を導入し、これに対応する正と負のルールを抽出できるようにして、認識結果を補正でき、認識率の向上が可能であることを示した。これにより、データマイニングの有効性を検証した。

次に、暗黙知から形式知へ変換することによる知識の「表出化」の問題として文書校正支援を取り上げ、校正者の暗黙知が表出した結果と考えられる文書校正データに相関ルールを応用し、校正ルールを抽出した。このとき、校正全体から見ると出現頻度の少ない校正も見逃さずに、ルールとして抽出するために、適用度と呼ぶ概念を導入し、さらに文書のドメインごとにこの適用度を学習することによって、ドメイン別に異なる校正ができることを示した。これにより、表記や文法、意味レベルの校正のみならず、組織情報に基づいた校正も可能であることを示した。

最後に、暗黙知を暗黙知のままに伝承する知識の「共同化」の問題として工作機械操作方法での知識共有化を取り上げ、熟練者による加工条件の設定方法を模倣することにより暗黙知を抽出し、共有化することに着目した。まず、この模倣に必要な最小限の情報を検討し、さらに人間の主観的な判断を重視できるように AHP とエントロピーモデルを組み合わせた手法を提案した。これにより、熟練者の作業を妨害することなく、暗黙知を含んだ有益な情報を収集すること、及び収集した情報の中で試行錯誤中の作業者が必要とする知識をマイニングすることが可能であることを示した。

以上、本研究では、まずデータマイニングの有効性をパターン認識の問題で検証し、次に知識共有化における「表出化」と「共同化」の2つのモードで、データマイニングを応用した新たな手法を提案し、さらに提案手法の有効性をオフィスワークと製造工程での問題において評価した。

## 学位論文審査結果の要旨

平成 13 年 7 月 31 日に第 1 回学位論文審査委員会を開催，8 月 2 日に口頭発表，その後第 2 回審査委員会を開催し，慎重審議の結果以下の通り判定した。なお，口頭発表における質疑を最終試験に代えるものとした。

データマイニングは，最近様々な分野においてその有効性が認識され始めている。本論文では，まずその代表的な手法である相関ルールが，パターン認識の問題である居住者の活動認識に有用であることを示した。次に，データマイニングを知識共有化の 2 つの問題に応用することを検討した。第 1 の応用では，暗黙知から形式知へ変換することによる知識の「表出化」の問題として文書校正データに相関ルールを応用し，校正ルールを抽出した。第 2 の応用では，暗黙知を暗黙知のままに伝承する知識の「共同化」の問題として工作機械操作方法での知識共有化を取り上げた。ここでは，熟練者による加工条件の認定方法を模倣することにより暗黙知を抽出し，共有化することに着目した。このように本論文では，データマイニングの有効性を検証し，知識共有化における「表出化」と「共同化」の 2 つのモードで，データマイニングを応用した新手法を提案し，さらに提案手法の有効性をオフィスワークと製造工程での問題において評価した。

以上の研究成果は，知識の共有化に大きく貢献するものであり，本論文は博士（工学）に値するものと判定した。