様式例：課程博士用

2002 年 2 月 21 日

電子・情報工学専攻	学籍番号	943437
申請者氏名	山　本　英　子	

指導教官氏名	梅　村　恭　司 中　川　聖　一

論　文　要　旨(博士)

論文題目	事実集合とテキスト集合からの知識発見に関する研究

（要旨 1,200字程度）

近年，コンピュータネットワークが発展し，空間を越えてデータを低コストで取り寄せることができるようになってきた．これは，データの在処を知っていれば，誰でもデータを入手することができるということである．また，情報抽出に高性能なコンピュータを利用できるようになればなるほど，探そうとする対象の範囲が広がるので，いくら高性能であっても十分ということはない．このようにコンピュータを用いて，新しく意味のある情報を発見することはコンピュータサイエンスにおいて最も面白い側面を扱っている．なぜなら，データが多ければ多いほど，人間はデータを読んで解釈するだけの能力では新しく意味のある情報を得ることはできなくなり，機械的にデータを生成したり再生したりすることによって，戦略を考え，振るいにかけたりして，データを選択し，解釈するという機械的な方法を採らざるを得ないからである．このような背景から，データベースにおける知識発見 (Knowledge Discovery in Databases:KDD) の研究分野が確立してきた．KDD において最も問題となる工程はデータの中から情報を取り出す工程である．この工程はデータマイニングと呼ばれる．KDD は主に情報が整理された事実集合をデータベースの対象とするが，対象としてテキスト集合を扱う場合，テキストマイニングと呼ばれる．データマイニングやテキストマイニングを効率的かつ人間にとっての有用性を考慮して行うことは難しい．それゆえデータベースとして，事実集合やテキスト集合からの知識発見に関する研究が意味を持っている．

そこで本研究では，情報を取り出すための新しいマイニング手法を提案するために，データの選択から知識発掘までの処理を行う手続きを，論理的な側面からのアプローチと統計的な側面からのアプローチを用いて検討する．そして，この一連の手続きを想定した論理的手法と統計的手法それぞれを評価する．また，それらのマイニング工程の前処理にあたる情報抽出の対象となるデータを選択するための手法についても検討し，新しい手法を提案する．

具体的には，論理的手法として，事実集合から規則を論理導出手法である Skipping Ordered Linear resolution (SOL 導出) で導き出す手法を提案し，評価した．そして，導かれた規則が有用な知識であるかどうかを判定する尺度を思案し，検討した．もう一方で，統計的手法として，テキスト集合から注目した事柄をデータとして抽出し，そのデータ間にある一対多関係をパターン認識で用いられる補完類似度で推定する手法を提案し，評価した．また，知識発掘の前処理となるデータの選択手法として，有用な情報を得るために，発掘に有効な類似したテキトを集めるための表記の揺れに寛容な類似尺度を検討し，情報検索によって収集性能を評価した．さらに，より類似したテキストを集めるために，選択される対象となるテキスト集合に特化したシソーラスを構築するシステムを検討し，評価した．本論文では，これらの実験と評価結果を報告する．