

会計データからの会計自動仕訳用述語論理式学習方式

A Knowledge Discovery Method for Accounting Entry Classification Service

永井義明^{*1}
Yoshiaki Nagai

高橋道郎^{*1}
Mitio Takahashi

佐々木浩二^{*2}
Koji Sasaki

^{*1} 中部大学
Chubu University

^{*2} (株)アドイン研究所
AdIn Research, Inc.

Accounting Processing is necessary for any company. Especially classifying entry data is critical but not trivial task for many small companies without accounting experts. And so we proposed a new knowledge discovery method for accounting entry data classification using "rough set theory". The special features of this method are clustering with the tuple values of decision attribute, and identifying the lower approximation and the upper approximation. This method can efficiently find the solutions because focusing decision attribute values and clustering individual data. Also this method can measure the usefulness of discovered knowledge using the accuracy of approximation in rough set theory. The outputs of this method are predicate logic clauses, and then we can classify entry data with predicate logic inference system based on the discovered knowledge.

1. はじめに

会計業務のような間接業務は企業運営にとって必須のものではあるが小規模企業では専任の担当者を置かず税理士事務所などの専門家に委託して記帳代行や税務申告処理を行うケースが多い。このようなケースでは当該小規模企業のみならず税理士事務所側もコスト削減や利便性の向上への要望が強く抜本的な改善が望まれている。そこでインターネットを用いて税理士事務所などの専門家の指導の下、小規模企業の会計処理をデータセンターで行う方式が考えられる。しかし、会計処理パッケージをインターネット VPN(Virtual Private Network)接続で小規模企業へ情報サービスを提供する方式では、会計システムへの入力となる仕訳処理が会計専門家の知識を必要とするので、小規模企業では困難であって利用できない。結局、従来どおり税理士事務所などの全面的な記帳代行に頼ることになり、インターネットによる会計情報サービスの有効活用がされていない。そこで、今までに会計専門家が行ってきた仕訳済みデータと小規模企業の個別特性と会計専門家の背景知識とから当該小規模企業の仕訳処理を行うルール(述語論理形式の知識)を学習する方式の開発を行った。

2. 小規模企業向け会計自動仕訳の課題

小規模企業で会計仕訳を行う場合、自社内の経理知識がある担当者か、あるいは、税理士事務所の会計専門家が行う場合が多い。会計仕訳を行うときは、出金であれば支払い先からの領収書、小口の支払いであれば購入物のレシート等、あるいは、銀行口座からの振り込みであれば、入出金明細書などの原始伝票を基にして、取引の特性を考慮して適切な会計仕訳を行ってきた。市販されている会計処理ソフトウェアパッケージは仕訳処理が適切に行われておれば税務申告や銀行への財務状態の説明資料等のための帳票出力を容易に行うことができるため、仕訳を適切に行えるかどうか重要となる。一般に、適切に仕訳を行なうためには、簿記を勉強し実務経験を積むことにより能力が獲得されるとして来た。

仕訳には種々のものが存在するが、比較的簡単な仕訳が大部分であることに着目し、自動的に仕訳を行うための自動仕訳パターンを作成して自動仕訳を行う方式が考案されている。しかし、ここで利用される自動仕訳パターンは会計専門家が個別に人手で入力する必要があるものであり、会計知識を持つ担当者がいない小規模企業では自動仕訳パターン(知識)を作成することが困難なため、十分に利用できていないと言えなかった。なお、本方式においても、少数の会計専門家でも迷う仕訳は別途会計専門家が人手でおこなうものとする。

3. 仕訳知識学習方式

3.1 会計仕訳の方式

決算仕訳以外の通常の企業活動で発生する仕訳を実際に調べてみると、既存仕訳の77%は前月に発生したと同じ種類の摘要をもっていた。また、摘要の他に取引の相手先、その他の企業の固有特性を含めると98%は自動仕訳ルールで仕訳できる可能性を確認した。そこで、既存仕訳データや企業固有の特性情報から得た会計取引の摘要情報、取引の相手先情報と、企業固有の特性情報とを条件属性として用いて、自動仕訳知識を自動的に生成できると考えた。更に、税理士からも小規模企業向けの会計仕訳のノウハウ集が出版されており、これらのノウハウを使って既存仕訳データにはまだ現れていない仕訳のパターンにも摘要できる方式を作成することを狙った。

3.2 ラフ集合論の決定表の作成と述語論理式の生成

直接入手できる会計取引のための情報には、過去に税理士事務所などの会計専門家が行った適切な仕訳結果、当該小規模企業固有の特性、会計専門家が持っている一般的/共通的な仕訳に関する知識(背景知識)、がある。会計専門家は、これらの情報を用いて適切な仕訳を行うが、取引の中には会計専門家でも迷うような仕訳や、どちらでもよい仕訳など、機械的に仕訳することが難しい仕訳も存在する。このため、必ずしも十分な情報が得られていない中で、どの程度分類がうまくできているかの指標が得られるようにするために、ラフ集合の考え方をを用いた自動仕訳知識の学習方式を作成した。また、学習により作成する知識は、表現/利用するために広く使用されている述語論理形式の知識とし、我々が現在まで研究してきた述

連絡先: 永井義明, 中部大学 経営情報学部, 〒487-8501
愛知県春日井市松本町 1200 番地, nagai@isc.chubu.ac.jp

語論理ベースの推論処理機構 (KAUS) と結び付けて利用することとした。

会計仕訳のために利用できる情報から、既存仕訳データ、対象企業の固有特性、とから、既存の仕訳結果である借方勘定科目と貸方勘定科目の組を決定属性として、既存仕訳の摘要情報、相手先情報、入手金の区別 (現金入手金、あるいは、銀行口座での入手金など)、企業固有の特性 (相手先との情報提供契約の有無など) の属性を条件属性として、ラフ集合の決定表を作成する。

決定表から分類規則を導き出す場合、条件属性の行列から識別表を作成して分類決定ルールを生成する方式を多くの研究では採用されている。本方式では、会計専門家によって判定された仕訳結果である借方勘定科目と貸方勘定科目との組の値により、決定表の各行であるデータベクトルを個体要素とするクラスターを作成して、分類決定ルールを探索する方式を採用した。本方式の学習のための手順を、図1に示す。

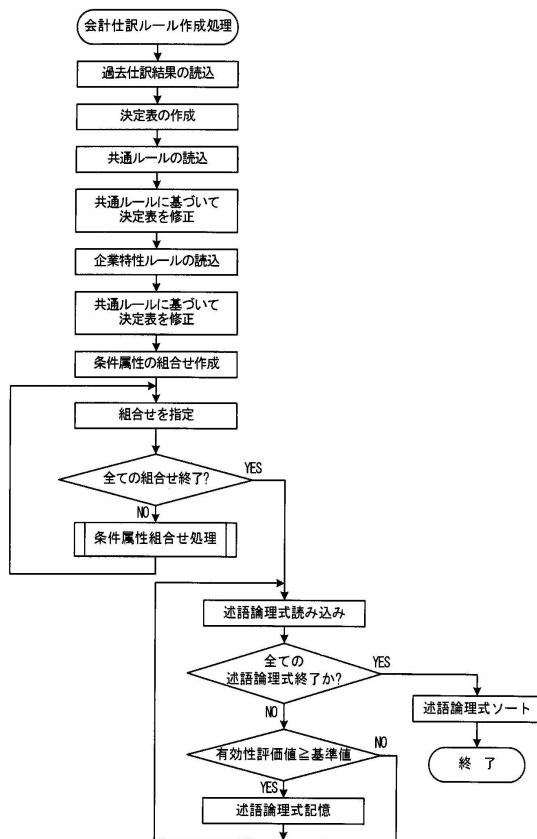


図1 学習のための動作手順

対象問題領域が会計仕訳ということから、次のような特徴を持つ方式とした。

- (1) 会計専門家が仕訳に対する持つ共通の知識 (背景知識) から、作成した決定表の中で知識の生成に関係しない列を削除する。これにより学習計算の効率を向上させる。
- (2) 学習する述語論理式の条件部分 (Body 部分) となる条件属性のすべての組合せを生成し、個々の条件属性の組合せについて述語論理式の生成を試行する。
- (3) 決定属性の組の値により条件属性の各データ行を個体要素としたクラスター化を行ない、取り上げた条件属性の組合せに含まれる条件属性列だけで構成したサ

ブ行列の中で、列のすべての値が未定義の場合、その列の条件属性は当該クラスターの決定属性の組の値の分類に寄与しないと考えるサブ行列から削除する。このサブ行列の縮約により計算量を削減する。

- (4) クラスター化によって最初の近似ルールを導出する。このルールが偶然ラフ集合論の上近似集合でかつ下近似集合であれば、あいまい性を持たない分類決定ルールとなる。この場合、条件属性のデータ型が非数値であれば、その非数値の条件属性の値が分類決定ルールの条件部となる。条件属性のデータ型が数値であれば、クラスターを構成している個体要素の数値の上限値と下限値とから分類決定ルールの条件部を作成する。
- (5) 上近似集合の条件属性の範囲は、上記(4)で作成した範囲となる。上近似集合を構成する要素は、上記(3)のサブ行列を構成する個体要素 (データ行) の内、上近似集合の条件属性の値の範囲に入る個体要素の数を計算しておき、後で、近似精度の計算のために使用する。

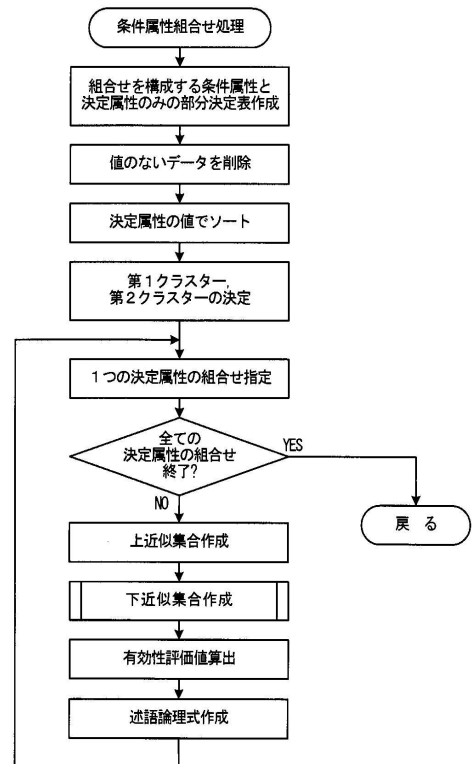


図2: 条件属性の組合せから述語論理式生成手順

- (6) 下近似集合の条件属性の範囲は、当該決定属性の組の値のクラスターに含まれる個体要素に対し、他の決定属性の組の値の持つクラスターに含まれる個体要素の包含関係を次の4つに分類する。
 - 第1関係: 重なりを有しない
 - 第2関係: 重なりを有するが互いに含まれない
 - 第3関係: 当該クラスターが他のクラスターを完全に包含している
 - 第4関係: 当該クラスターが他のクラスターに完全に包含されている
 それぞれの場合、図3の下近似集合作成の手順に従って下近似集合を作成する。

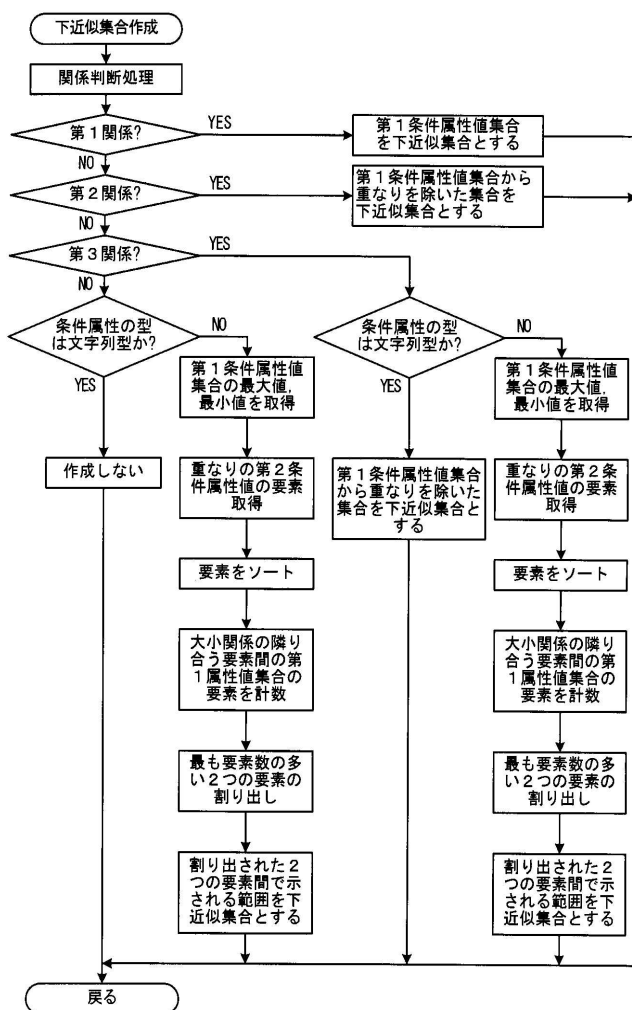


図4：下近似集合の作成手順

- (7) 生成する述語論理式は次のような形式とする。
 決定属性名 (決定属性値 1, 決定属性値 2):-
 条件属性₁ (属性値 x₁ の条件), ...,
 条件属性_n (属性値 x_n の条件).
 ここで、条件属性が数値型の場合は次のような形式とする。
 決定属性 (決定属性値 1, 決定属性値 2):-
 条件属性 a(A),
 greaterEqualInt(A, s), lessEqualInt(A, t).
 条件属性が数値型でない場合は次のような形式とする。
 決定属性 (決定属性値 1, 決定属性値 2):-
 条件属性 a(A), member(A, [...]).

- (8) 上近似集合の範囲及びそれに含まれる個体要素数 (決定表のデータ行数) と、下近似集合の範囲及びそれに含まれる個体要素数とを計算する。求めた値から、次の式により学習した仕訳知識の有効性評価値を求める。

$$e = \frac{n}{m}$$

ここで、

- e: 有効性評価値
- n: ラフ集合の近似精度
- m: 条件属性の組合せの条件属性数のパラメータ

: パラメータ (例として 2 を用いる)

m: 全データ数

n: 下近似集合に属する要素をもつデータ数

- (9) 有効性評価値により、学習した仕訳知識のランク付けを行い、一定値以上の知識を抽出する。

4. 自動仕訳知識学習方式の実験

実仕訳データを元に作成した実験データを使用して、自動仕訳知識の学習方式の実験を行った。実験の結果を動作例を用いて説明する。

仕訳データと企業の固有特性データから図5に示すような決定表を作成した。この表のデータはカラム1からカラムiまでと、仕訳結果を示すカラムと、i+1、カラムi+2とに分けられる。ここで、決定すべき仕訳結果 (決定属性と呼ぶ) であるカラムi+1、カラムi+2の組と、それらの組の値を決める条件属性であるカラム1からカラムiまでとに分け、条件属性の値の組合せから決定属性の値の組を求める問題を解く。

- 第1ステップ:** 背景知識の利用による決定表の縮退を行う。
 背景知識から日付データ列を外す。

- 第2ステップ:** 決定属性の組み合わせによる条件属性データのクラスター化を行う。
 決定属性の組合せの値の組で、決定表の行をソートし、同じ決定属性値を持つデータ行の要素のクラスターを作成する。

- 第3ステップ:** 条件属性のすべての組合せについて述語論理式が作成できるか探索する。
 条件属性の組合せとして、[相手先、情報提供契約]を説明のため採り上げる。

- 第4ステップ:** 条件属性の組合せと決定属性だけのサブ決定行列を作成する。

- 第5ステップ:** 未定義列を含むデータ行の縮約を行う。
 上記サブ行列中で、条件属性の値が未定義の行があれば、そのデータ行を削除してサブ行列を作成する (図6)。

- 第6ステップ:** 条件属性値の範囲を計算する。
 同じ決定属性の値 (個々の例では、(交際費, 当座預金) と (販売手数料, 当座預金)) の組を持つデータ行の条件属性の値を当該クラスターとそれ以外 (第2クラスター) に分けて、条件属性値の範囲を計算する (図7)。

- 第7ステップ:** 下近似集合と上近似集合を計算する。
 下近似集合と上近似集合を計算し、下近似集合に属する個体要素数と下近似集合に属する個体要素数を計算し、ラフ集合の近似精度を計算する。近似精度の値に対し、全体のデータ数に対する述語論理式を生成するために直接使用したデータ数との比率による補正と、使用した条件属性の数による補正を行ない、当該述語論理式生成の有効性評価値とする (図8)。

- 第8ステップ:** 述語論理式を生成する。
 述語論理ベースの推論機構が利用できるようにテキスト形式の述語論理式を出力する。

カラム No. 1

i i+1 i+2

条件属性										決定属性		
日付	摘要	相手先	現金入金	現金出金	当座預金入金	当座預金出金	普通預金入金	普通預金出金	対象社員比率 (%)	情報提供契約	借方勘定科目	貸方勘定科目
4月2日	本日売上		86080								現金	売上高
4月2日	仕入れ	A豆腐店		37972							商品仕入高	現金
4月2日	社会保険料	T社会保険事務所						129204			法定福利費	普通預金
4月3日	本日売上		136650								現金	売上高
4月3日	仕入れ	A豆腐店		9713							商品仕入高	現金
4月3日	仕入れ	K酒店		3675							商品仕入高	現金
4月3日	ホワイトボード マーカー	S事務用品店		850							備品・消耗品費	現金
4月3日		東京ガス				20405					水道光熱費	当座預金
4月3日		Cカード会社				4200					備品・消耗品費	当座預金
4月7日	ごみ処理券	Kスーパー		1080							雑費	現金
4月10日	電話代	NTT				6074					通信交通費	当座預金
4月10日	電話代	TTネット				1320					通信交通費	当座預金
4月14日	香典代	U様		10000							交際費	現金
4月15日	社員旅行			300000						5	事務員給与	現金
4月15日	情報提供料	M調査会社				200000				有り	販売手数料	当座預金
4月17日	仕入れ	N商店				478155					商品仕入高	当座預金
4月25日	電話代	KDDI				623					通信交通費	当座預金
4月26日	顧問料	Y会計事務所				47500					管理諸費	当座預金
4月26日	家賃	Tビル				300000					地代家賃	当座預金
4月27日	給与	A専務		450000							役員報酬	現金
4月27日	給与	従業員2名		779420							事務員給与	現金
5月22日	情報提供料	P氏				150000				無し	交際費	当座預金
5月25日	社員旅行			2500000						95	福利厚生費	現金
6月1日	お茶代(社内)			10000						100	福利厚生費	現金
6月7日	情報提供料	Q大学R教授				160000				無し	交際費	当座預金
8月29日	情報提供料	S氏				120000				無し	交際費	当座預金
10月15日	社員旅行			240000						4	事務員給与	現金
10月15日	情報提供料	M調査会社				150000				有り	販売手数料	当座預金
12月1日	お茶代(社内)			10000						100	福利厚生費	現金
12月25日	社員旅行			1000000						90	福利厚生費	現金

図5：決定表

条件属性		決定属性		
相手先	情報提供契約	借方勘定科目	貸方勘定科目	組合せ
P氏	無し	交際費	当座預金	第1
Q大学R教授	無し	交際費	当座預金	
S氏	無し	交際費	当座預金	第2
M調査会社	有り	販売手数料	当座預金	
M調査会社	有り	販売手数料	当座預金	

図6：縮約後の決定表

決定属性の組合せ	条件属性名		第1クラスター		第2クラスター	
	条件属性名	データ型	値	第1条件属性値集合	値	第2条件属性値集合
第1 (交際費, 当座預金)	相手先	文字列型	[P氏,無し], [Q大学R教授,無し], [S氏,無し]	[P氏,Q大学R教授,S氏]	[M調査会社,有り], [M調査会社,有り]	[M調査会社]
	情報提供契約	文字列型	[無し]	[無し]	[有り]	[有り]
第2 (販売手数料, 当座預金)	相手先	文字列型	[M調査会社,有り], [M調査会社,有り]	[M調査会社]	[P氏,無し], [Q大学R教授,無し], [S氏,無し]	[P氏,Q大学R教授,S氏]
	情報提供契約	文字列型	[有り]	[有り]	[無し]	[無し]

図7：条件属性値の範囲計算

決定属性の組合せ	条件属性		重なり 関係	下近似集合		上近似集合	有効性 評価値	述語論理式 (prolog言語形式)
	属性名	値		値	値			
第1 (交際費, 当座預金)	相手先	[]	1	[P氏,Q大学R教授,S氏]	[P氏,Q大学R教授,S氏]	0.0707	仕訳(交際費,当座預金):- 相手先(A),member(A,[P氏,Q大学R教授,S氏]), 情報提供契約(B),member(B,[無し]).	
	情報提供契約	[]	1	[無し]	[無し]			
第2 (販売手数料, 当座預金)	相手先	[]	1	[M調査会社]	[M調査会社]	0.0471	仕訳(販売手数料,当座預金):- 相手先(A),member(A,[M調査会社]), 情報提供契約(B),member(B,[有り]).	
	情報提供契約	[]	1	[有り]	[有り]			

図8：下近似集合と上近似集合の計算、及び、有効性表価値の計算

5. おわりに

実際に発生している仕訳データを基に実験データを作成して実験を行ったが、税理士事務所の協力を得て実用性の高い方式とするために現場で発生している多種の会計仕訳データを入力し、本研究で開発した自動仕訳知識学習方式の実用性の検証を行い方式の改良を進めている。また、Webサービスのインタフェース標準に従い、インターネットから仕訳に必要な取引データを入力して、本研究の学習方式を使って仕訳を行い、仕訳結果をXBRL-GLに準拠したXML形式でWebサービスのリクエスト元へ配信する方式のプロトタイプシステムを開発中である。更に、ラフ集合論の決定表から述語論理形式知識を学習する方式として、会計仕訳以外の分野にも応用することを検討していきたいと考えている。

参考文献

[Inuiguchi 03] Masahiro Inuiguchi, Shoji Hirano, Shusaku Tsunmoto (eds.), "Rough Set Theory and Granular Computing", Physica-Verlag, 2003
 [大須賀 81] 大須賀節雄,"知識ベースシステム KAUS - その知識表現と推論機構",情報処理学会 人工知能と対話技法 20-5,1981
 [Yamauchi 99] Hiroyuki Yamauchi,"KAUS User's Manual",1999
 [鍾 00] 鍾寧,"確率的ラフ帰納学習システム GDT-RS による共通データからの知識発見",人工知能学会,Vol.15 No.5,2000
 [稲村 00] 稲村榮典,"仕訳らくら大辞典",あさ出版,2000
 [永井 04] 永井義明,佐々木浩二,"述語論理式生成方式",特願 2004-378628, 2004