

要求仕様書からのオブジェクト指向分析における モデリング要素の抽出

山本 健一, 山田 寛康, 島津 明

北陸先端科学技術大学院大学 情報科学研究科

{k-yamamo,h-yamada,shimazu}@jaist.ac.jp

1 はじめに

オブジェクト指向開発は、大規模なシステム開発に非常に有効であり、実際のシステム開発の現場でも取り入れられつつある。オブジェクト指向開発の過程は、オブジェクトの発見（オブジェクトの識別、オブジェクト間の関連の識別、オブジェクトの属性と操作の識別）、オブジェクトの整理（オブジェクトの抽象化、オブジェクトの特殊化、オブジェクトの組み立て、オブジェクトの要素分解）に分類される [5]。

このようなモデリングは、作成者の経験に依存する部分が大きく、対象とする世界が複雑な場合、間違いが起りやすい。一般にソフトウェア開発においてシステム開発の要求者はシステム開発の上流工程に参加することは難しいとされている。したがって、オブジェクト指向開発の初期段階において

- システム開発の上流工程での間違いを少なくする
- 要求者をシステム開発の上流工程に参加しやすくする

という2点が支援されることが期待される。

このような支援に関して、滝沢らは、要求仕様書から、オブジェクトモデル図を作成する手法を提案している [1]。このシステムの入力は、単文に限られている。名詞辞書には、各名詞に対し、クラス・インスタンス・属性のいずれかが記入されており、ある名詞がクラスになったり、インスタンスになったりする問題に対応できない。

原田らは、要求仕様書からオブジェクト指向設計図を自動生成するシステムを提案している [2]。クラスを抽出するルールにはシソーラス (EDR シソーラス) を用い、関連を抽出するルールは格フレームを用いている。

本報告では、「オブジェクトの発見」のうち「オブジェクトの識別」及び、「オブジェクト間の関連の識別」を支援するために、自然言語で記述された要求仕様書から、クラス、クラス間の関連を抽出する方法について述べる。方法はシソーラスによる抽出に加え、辞書の語釈文、動詞の係り受け情報を用いることで、精度の向上を目指している。

2 モデリング要素

2.1 クラス

クラス図の中で最も中心的なモデリング要素はクラスである。文献 [4] では、クラスを識別するのに役立つ質問がいくつか示されているが、そのうち次の4つの質問が、支援システムの参考になると考える。

- 保管または分析する必要のある情報があるか。保管、転送、分析などの処理が必要な情報があれば、その情報をクラスの候補となる。
- システムが処理しなければならない装置があるか。その装置は、クラスの候補となる。
- 組織に関する単位があるか。クラスを使うと組織はうまく表現できる。
- 人はどんな役割を演じるか。役割はクラスとして見なすことができる。

この4つの質問を参考に、提案手法はクラスを次の5つに分類して用いる。

- (イ) 情報を保管、転送、分析するもの
- (ロ) 工業的な装置
- (ハ) 組織に関する単位
- (ニ) ユーザー、システム・オペレーターなどの人に関連するもの
- (ホ) その他

2.2 関連

関連とは、クラス間の構造的関係や情報の流れ (メッセージ通信の経路) を表現するものであり、クラスと共に対象世界を表現するものである。要求仕様書においては、多くの場合、関連は動詞句に相当する。関連に使用される動詞は図1にあるように多岐にわたり、多様なドメインの要求仕様書に対応するためには、要求仕様書に現れる動詞から関連の候補を抽出する際に、シソーラスなどでフィルタリングすることは、難しい。したがって、関連についてはフィルタリングを行うことなく、できるだけ多数の関連を出力し、その中に正解を含むことで、開発者の支援を行うこととする。

を発行する, 交差する, で働く, 管理する, アクセス権をもつ, 上場する, 所有する, の委員である, を雇う, を確認する, を保持する

図 1: 関連の具体例

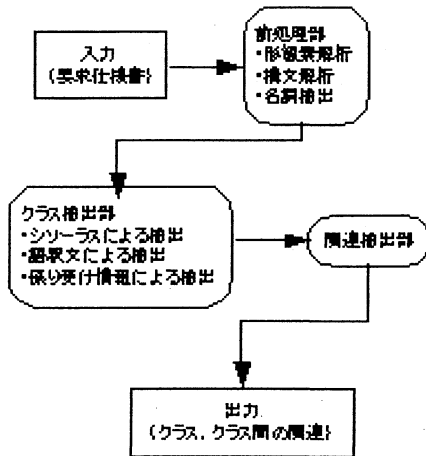


図 2: システム図

3 モデリング要素抽出法

システムは図 2 に示すように大きく前処理部, クラス抽出部, 関連抽出部の 3 部分からなる。前処理部では, 形態素解析, 構文解析, 及び名詞の抽出を行う。名詞の抽出は, 構文解析の結果を受けて行う。複合語に関しては, 最後の形態素を出力するものとする。これは最後の形態素に, 複合語全体の概念を表す形態素が多いためである。例えば, “現金自動支払機” は, “現金” + “自動” + “支払” + “機” と 4 つの形態素から構成され, 最後の形態素 “機” が, “現金自動支払機” 全体の概念を示している。

3.1 クラスの抽出

クラスの抽出は, シソーラスによる抽出, 辞書の語釈文による抽出, 動詞の係り受け情報による抽出, 不要候補をフィルタリングする部分からなる。

(1) シソーラスによる抽出

クラスのうち「工業的な装置」に分類されるものは, シソーラスにおいて, 意味クラス<機械>を上位概念にもちやすい。そこで, 前処理部で抽出された名詞のうち, 上位概念に意味クラス<機械>をもつものを, クラスとして抽出する。なお, ここで用いるシソーラスは, 日本語語彙大系 [8] である。

人, 者, 客, 一団, 団体, 組織, 法人, 機関, 部類

図 3: 語釈文中で着目する単語

学校で学業を修めている者。特に, 大学に学ぶ者

図 4: “学生” の語釈文

(2) 語釈文による抽出

クラスのうち, 「組織に関する単位」と「ユーザー, システム・オペレーターなど人に関連するもの」は, 語釈文中に図 3 にある単語を含みやすい。よって, 前処理部で抽出された名詞のうち, 語釈文の中心的概念として図 3 にある単語を含んでいるものをクラスとして抽出する。ここでは, 語釈文の中心的概念とは, 語釈文が体言止めになっている場合の最後の体言とする。例えば, “学生” の語釈文は図 4 であるが, 文末の体言は “者” であるので, “学生” がクラスの候補として抽出される。また, ここで使用する国語辞典は岩波国語辞典 [9] である。

(3) 動詞の係り受け情報による抽出

クラスのうち「情報を保管, 転送, 分析するもの」, 及び「その他」のなかにある部分全体の関係にあるものは, 情報を保管, 転送, 分析することを意味する動詞や, 部分全体の関係を意味する動詞に係っていると推察される。具体的には, 日本語語彙大系において, “保管する” や “部分全体” を意味するクラス<所有>, “転送する” を意味するクラス<物理的变化>, “分析する” を意味するクラス<思考動作>に含まれる動詞に係っている名詞をクラスとして抽出する。

(4) 不要候補のフィルタリング

次のような名詞は, クラス候補から排除する。

- 形式名詞, 副詞的名詞, または名詞形態指示詞のいずれか
- 名詞性名詞接尾辞であり, 直前の形態素が形式名詞, 副詞的名詞, または名詞形態指示詞のいずれか

ここで削除されるものは, “それ”, “それら”, “もの” などである。

3.2 関連の抽出

関連は, 要求仕様書において動詞句に相当する場合が多い。そこで, クラス抽出部で出力された 2 つのクラスが, 同じ用言に係っている場合, その用言をクラス間の関連とする。例えば “銀行は銀行出納端末を所有する” を構文解析すると図 5 のようになる。“銀行” と “銀行出納端末” は, “所有する” に係っているので,

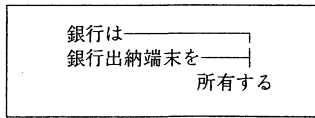


図 5: 構文解析結果

“所有する”が“銀行”と“銀行出納端末”の関連として抽出され，“所有する（銀行，銀行出納端末）”と出力される。以後，クラス A とクラス B の間の関連 L を“L(A,B)”と表記する。

4 モデリング要素の抽出実験

実験に用いた要求仕様書は，銀行ネットワークに関する要求仕様書 [3] で，図 6 に示す。正解としたクラス図は文献 [3] にあるものとした。また，形態素解析には JUMAN[6] を，構文解析には KNP[7] を利用した。

クラスの抽出に関する実験結果を表 1 に示す。ただし，本実験においては，“ATM”，“自動支払機”と“現金自動支払機”，及び“現金出納端末”と“銀行出納端末”は，表記のゆれと判断し，同じものとみなした。表 1 の見方は，クラスの列が正解のクラス名で，システムの列には，本システムが出力したクラスには“○”が記してある。また，動詞，シソーラス，語釈文の列にはそれぞれ，動詞の係り受け情報を用いた手法，シソーラスを用いた手法，辞書の語釈文を用いた手法が，それぞれ記してある。フィルタ前の列には，フィルタリングを施す前の結果を示している。比較対象として，要求仕様書から単純に名詞のみを抽出したものを，名詞の列に示す。誤抽出の行には，誤抽出した単語の数が書かれている。

表 1 より，要求仕様書から単純に名詞のみを抽出する方法と比較して，精度の向上が確認できる。再現率は名詞を抽出する方法が，要求仕様書からクラスを抽出する場合の最高値であり，システムもその値を示している。システムの最終出力の前のフィルタにより除去されたものは，形式名詞の“もの”であった。

システム出力の内訳をみると，動詞の係り受け情報を用いた手法が，抽出可能なクラス 9 個のうち，7 個を抽出しており，また誤抽出も 4 個と少ない。シソーラスを用いた手法は，動詞の係り受け情報を用いた手法では抽出できなかった“ATM”を抽出している。辞書の語釈文を用いた手法は，動詞の係り受け情報を用いた手法では抽出できなかった“顧客”を抽出している。

システムが誤って抽出したものを図 7 に示す。このうち動詞の係り受け情報を用いた手法だけが間違っ

コンピュータ化された銀行ネットワークを支援するソフトウェアシステムを設計せよ。このシステムには，現金出納係および銀行コンソーシアムによって共有される複数の自動支払機が含まれる。各銀行は，それぞれのコンピュータによって口座の管理と口座に対して発生するトランザクションを処理している。各銀行は自らのコンピュータに直接接続された複数の銀行出納端末を所有している。現金出納係は，口座とトランザクションを投入する。現金自動支払機は，キャッシュカードを受け付けて，ユーザと対話し，トランザクションを実行するために中央のシステムと通信し，現金を支払い，レシートをプリントする。このシステムは，適当な記録の保管設備とセキュリティ対策が必要である。また，同じ口座への並行アクセスを正確に扱わなければならない。各銀行は，自らのコンピュータに，それぞれのソフトウェアを用意するものとせよ。そして，あなたがこれらの自動支払機とネットワークのためのソフトウェアを開発するものとする。共有されたシステムのコストは，キャッシュカードを持っている顧客の数に応じて各銀行に割り当てられることになる。

図 6: 銀行ネットワークに関する要求仕様書

コンピュータ，システム，ソフトウェアシステム，管理，並行アクセス，保管設備

図 7: システムが誤って抽出したもの

誤抽出したものは，“保管設備”であり，動詞の係り受け情報を用いた手法とシソーラスを用いた手法が共に間違っ

次に関連についての実験結果を表 2 に示す。評価の列には，“1”から“3”の数値が示されているが，“1”が正解とするクラス図にあるもの，“2”がクラス図にはないものの，支援システムとしては，妥当と判断できるもの，“3”は，明らかな誤抽出を意味している。このように分類したのは，関連が直接には要求仕様書には表現されない場合が多いためである。“1”と評価したものは 2 つ，“2”と評価したものは 3 つであり，“1”または，“2”と評価したものは，合計 5 つであった。逆に“3”と評価したものは 6 つであり，関連の抽出は，精度が高いとは言えない。“3”と評価したもののうち，“処理する（管理，トランザクション）”と“投入する（口座，トランザクション）”は，関連として抽出することは不適當であり，表層格に関するルールを加えることで除去可能であると思われる。

表 1: クラスに関する実験結果

クラス	名詞	システム	システム出力の内訳			
			フィルタ前	動詞	シソーラス	語釈文
トランザクション	○	○	○	○		
入力端末						
現金出納係トランザクション						
遠隔トランザクション						
更新						
ATM	○	○	○		○	
現金出納係端末	○	○	○	○	○	
現金出納係	○	○	○	○		
顧客	○	○	○			○
カード認証						
コンソーシアム	○	○	○	○		
銀行	○	○	○	○		○
口座	○	○	○	○		
キャッシュカード	○	○	○	○		
誤抽出	40	6	7	4	3	1
再現率	0.64	0.64	0.64	0.57	0.14	0.14
精度	0.18	0.60	0.57	0.64	0.40	0.67
F 値	0.28	0.62	0.60	0.56	0.21	0.23

表 2: 関連に関する実験結果

システムの出力 (L(A,B))	評価
含む (システム, 自動支払機)	2
所有する (銀行出納端末, 銀行)	1
必要だ (システム, 保管設備)	3
する (コンピュータ, 銀行)	3
処理する (銀行, 管理)	3
処理する (銀行, トランザクション)	2
よる (現金出納係, 銀行コンソーシアム)	3
処理する (管理, トランザクション)	3
投入する (現金出納係, 口座)	2
投入する (現金出納係, トランザクション)	1
投入する (口座, トランザクション)	3

5 おわりに

オブジェクト指向分析の初期段階である「オブジェクトの発見」, 「オブジェクト間の関連の識別」を支援するために, 自然言語で記述された要求仕様書から, クラス, 及びクラス間の関連を抽出する手法を提案した. 実験の結果, クラスの抽出に関しては, 要求仕様書から名詞を抽出する手法と比較して, 精度の向上がみられた. また, 再現率に関しては, 要求仕様書に明記されているクラスは, すべて抽出することができた.

クラス間の関連に関しては, 支援システムを考慮した評価を行ったが, 十分な精度は得られなかった. しかし, 表層格に関するルールなどを追加することで, 精度の向上は期待できる.

人間が要求仕様書からクラス図を作成する場合, 要求仕様書には明記されていない様々な知識を使用し, クラス図を作成している. したがって, 要求仕様書に明記されないモデリング要素が多数存在し, それらに対応する必要がある. 例えば, 要求仕様書に表現されない関連に対応するためには, クラスと出力された単語と共起しやすい動詞を大規模コーパスから自動抽出する方法などが考えられる.

ラス図を作成している. したがって, 要求仕様書に明記されないモデリング要素が多数存在し, それらに対応する必要がある. 例えば, 要求仕様書に表現されない関連に対応するためには, クラスと出力された単語と共起しやすい動詞を大規模コーパスから自動抽出する方法などが考えられる.

参考文献

- [1] 滝沢 陽三, 上田 賀一, オブジェクト指向に基づく要求記述からの形式的仕様の導出手法, 情報処理学会ソフトウェア工学研報, Vol.93-SE-94, 1997.
- [2] 原田 実, 野村 佳秀, 山本 幸二, 大野 雅志, 田村 浩樹, 高橋 史郎, 自然言語要求仕様からオブジェクト指向設計図を自動生成するシステム CAMEO, 情報処理学会論文集, Vol.38-10, pp.2031-2039, 1997.
- [3] J. ランポー, M. プラハ, W. プレメラニ, F. エディ, W. ローレンセン (羽生田 栄一: 監訳), オブジェクト指向方法論 OMT, トッパン, 1992.
- [4] H.E. エリクソン, M. ペンカー (杉本 宣男, 落合 修, 武田 多美子: 監訳), UML ガイドブック, エスアイビー・アクセス, 1998.
- [5] 落水 浩一朗, 東 雅宏, オブジェクトモデリング, ジャストシステム, 1995.
- [6] 黒橋 禎夫, 長尾 真, 日本語形態素解析システム JUMAN version3.6, 1997.
- [7] 黒橋 禎夫, 日本語構文解析システム KNP version2.0b6, 1998.
- [8] NTT コミュニケーション科学研究所, 日本語語彙大系, 岩波書店, 1997.
- [9] 西尾 実, 岩淵 悦太郎, 水谷 静夫, 岩波国語辞典第六版, 岩波書店, 2000.