自然言語処理 135－15
（2000．1．28）

SD式意味モデルによる文書の自動要約

小柳 貴　新見 道治　河口 英二

九州工業大学工学部電気工学科
〒804-8550 北九州市戸畑区仙水町 1-1

E-mail:{koyanagi,niimi,kawaguch}@know.comp.kyutech.ac.jp

概要

文間の意味差を利用した重要文抽出による自動要約の手法を提案する．著者らの提案しているSD式意味モデルの下では，自然言語をSD式という意味データによって記述することで，文間の意味差を定量的に扱うことが可能となる．本論文では，まず従来のSD式意味モデルでは取り上げていなかった結合子形式のSD式に関する意味差を新たに定義し，さらに意味差の正規化手法を提案する．このうえで，文間の意味差を用いた自動要約手法を実現する．本論文では，各段落における最重要文は多くの文において意味を共有していると仮定する．この仮定に基づき，各段落から最重要文を決定し，その後，順次重要文を抽出していく．実験では童話を要約の対象とした．原文の各文をSD式へ変換し，関連する知識データを作成・登録し結果を求めた．また，人手によって抽出された文と比較することで結果を考察した．

キーワード　自動要約, 重要文抽出, 意味差, 意味モデル, 知識データ

Text Summarization by SD-Form Semantics Model

Takashi Koyanagi　Michiharu Niimi　Eiji Kawaguchi

Kyushu Institute of Technology,
Department of Electrical, Electronic and Computer Engineering
1-1 Sensui-cho, Tobata, Kitakyushu, 804-8550 Japan

E-mail:{koyanagi,niimi,kawaguch}@know.comp.kyutech.ac.jp

Abstruct

This paper presents a summarization method by extracting sentences using semantic difference measure between sentences. The SD-Form semantics model, which was proposed by the authors, deals with the difference of meaning in a quantitative way between two sentences in terms of SD-Form. In this paper, we define a compute method of semantic difference measure with connected SD-Forms that is no definition in former SD-Form semantics model, and we propose a normalization method of semantic difference measure. Then we materialize automatic summarization using semantic difference measure between sentences. We assume that the most important sentence in a paragraph shares meaning in many sentences. Under this assumption, the summarization method decides the most important sentence in each paragraph, and then extracts important sentences one after another. This paper applies the method to a fairy tale. We converted each sentence of original text into SD-Form, made out and registered a knowledge data related to the original text, and obtained a result. Also, we evaluated the result by comparing with sentences which were extracted by the examinee.

key words　automatic summarization, sentence extraction, semantic difference measure, semantic model, knowledge data

1 はじめに

これまでに提案されてきた文書の自動要約手法は，タイトルの情報や単語の出現頻度など表層表現に注目したものが多い [1]．文や段落間の類似度を利用した方法も提案されているが，この類似度も共通な単語の出現度合などによって計算されている [2]．他には，文間の関係に着目し重要文を抽出することによって，要約文に飛躍を生じさせない手法も提案されており，スクリプトのような知識構造も利用されている [3]．しかし，人間が行う文の理解・要約の過程になるべく近い処理の実現を目的とすれば，より多くの知識と意味を扱うことができるシステムが必要となってくる．

このように知識を基にした意味処理を可能にするため，著者らは自然言語の意味を記述する SD 式 (Semantic-structure Description Form) と呼ぶ中間言語を提案し，さらにこれを意味データとする SD 式意味モデルにおいて，自然言語の意味を定量的に扱うことができる枠組みを提唱している [4]．この SD 式意味モデルの最大の特長として，概念間の意味的距離である意味差を求めることができる．また，我々の持つ知識や規則を一定の形式で記述しておけば，その情報は意味差の計算時に用いられ，より人間の感覚に近い計算結果が得られる．

本論文では，この SD 式意味モデルによる文間の意味差に基づく要約の一手法を提案する．この手法では，要約文とは重要な文の集合であるとの立場を取っている．これは，重要である文は他の文と意味を多く共有しているという仮定に基づいている．要約文は，まず原文の各段落における最重要文を抽出した後，得られた文と最も意味差の小さい文を順次抽出していくことで求まる．実験では，要約の対象文書としてグリム童話 [5] の物語を扱った．物語の各文を SD 式として記述し，関連する知識データを登録し，重要文の抽出を行った．このように，他の多くと類似する文の集合を要約文とする手法は以前にも報告例が見られるが [1]，文間の類似度に知識を用いている点で本論文の手法は他と異なる．

以下，2 章では SD 式意味モデルについて述べる．特に，本論文で定義した結合子形式の SD 式に関する意味差の計算方法や，意味差の正規化について説明する．3 章で要約の手順を，4 章で実験例を示し，結果を考察する．最後に，5 章でまとめと今後の課題について述べる．

2 SD 式意味モデル

2.1 SD 式による意味記述

SD 式は曖昧さを持たない文脈自由言語の一種であり，「SD 式記号」と呼ばれる概念ラベル，修飾子，規定子，結合子，機能項目記号，区切り記号などから構成される記号列である．SD 式は SDG(SD-Form Generative Grammar) という文法の生成規則によって定義され，以下の 8 種類に分類できる．

(1) 変数概念ラベル
(2) 単純概念ラベル
(3) パラメータ付き概念ラベル
(4) 修飾形式
(5) 規定子形式
(6) 結合子形式
(7) 陳述形式
(8) 感情形式

以下，具体的な 2 つの SD 式を例に説明する．

- $[s(\text{彼}), v(\text{借りる/過去}), o(\text{参考書}(\$)/[s(\text{兄}), v(\text{買う/過去}), o(\$)])]$
  : 彼は，兄が買った参考書を借りた

これは陳述形式の例である．陳述形式は，英語の文型を模した 9 つの形式を用いて記述する．この形式の中で用いられる陳述機能項目記号には，以下のものがある．

s : 主語項目, v : 述語項目, o : 目的語項目
i : 間接目的語項目, c : 補語項目, b : 行為者項目

上の例における" 彼" や" 借りる" は単純概念ラベルであり，通常は自然言語の単語をそのまま用いる．また，" 買う/過去" のような修飾形式では，左側の概念を右側の概念で修飾している．目的語項目でも同じ修飾形式が使われているが，ここでは先行詞" $" を用いた関係代名詞的用法によって" 兄が買った参考書" を表現している．

- $(assu([s(\text{雨}), v(\text{降る})]))\ caus([s(\text{サッカー}), v(\text{中止になる})])$
  : もし雨が降れば，サッカーは中止になる

この例では，規定子"$assu$" を用いて仮定を表現した規定子形式の SD 式を，さらに結合子"$caus$" を用いて結合子形式とし，因果関係を記述している．

2.2 SD式の意味的情報量

SD式では，記号列の構造で何か固有の概念を表現しようとするだけでなく，その概念の意味量の大小も表すこととしている．一般には複雑な構造をしたSD式ほど多量の意味を持つ．しかし意味を定量的に扱う場合，意味の絶対量は本質的に重要ではなく，他との相対量が重要となってくる．

任意のSD式を d とするとき，その意味量を，

$$si(d) = n[semit]$$

と表す．単位は $semit$ と名付けている．

SD式意味モデル実験システム「SDENEV-3」では，各SD式記号の意味素量を次のように設定している．

(1) 変数ラベル	$1[semit]$
(2) 単純ラベル	$10[semit]$
(3) 修飾子	$1[semit]$
(4) 規定子	$2[semit]$
(5) 結合子	$1[semit]$
(6) 機能項目記号	$1[semit]$
(7) 区切り記号"[ ]"	$1[semit]$
(8) 区切り記号"( )"および ","	$0[semit]$

SD式の意味量はこれらの総和となる．例を示す．

- si(テキスト/新しい) $= 21$
- $si([s$(彼)$, v$(投げる/過去)$, o$(ボール)$]) = 45$

2.3 知識の記述

SD式意味モデルで用いられる知識データには，対象とする世界の中で常に真となる事実を記述した「事実知識」と，推論の根拠となる規則を記述した「規則知識」とがある．それぞれ，$fact$(事実知識)，$rule$(規則知識) という形式で登録する．例を示す．

[事実知識]

- (日本/首都)$equa$(東京)
  : "日本の首都"と"東京"は等しい
- (野菜)$incl$([人参, 大根])
  : "野菜"は"人参，大根"を含む

[規則知識]

- $(assu([s(X), v$(食べる/程度/特大)$, o(Y)]))$
  $caus([s(X), v$(好む)$, o(Y)])$
  : もし「X が Y をよく食べる」ならば，
  「X は Y を好きである」

2.4 概念間の詳述関係

2つの概念間の「詳述関係」は，SD式意味モデルにおける最も基本的な枠組みである．今，2つの概念 d_1，d_2 に関して d_1 の意味をより詳しくしたものが d_2 であるとき，d_1，d_2 間には詳述関係があるといい，

$$elab(d_1, d_2) = n \quad \text{または} \quad elab(d_1, d_2, n)$$

と表す．ただし，n は詳述量 (単位：$semit$) であり，

$$0 \leq n < \infty$$

とする．このような d_1 を d_2 の先祖，d_2 を d_1 の子孫と定義する．

SD式意味モデルにおける詳述関係には，d_1，d_2 そのものの構造による「構文的な詳述関係」と，システムが利用できる知識データによる「知識に基づく詳述関係」の2種類がある．それぞれ，

$$elab_{synt}(d_1, d_2) = n, \quad elab_{synt}(d_1, d_2, n)$$
$$elab_{know}(d_1, d_2) = n, \quad elab_{know}(d_1, d_2, n)$$

と表す．

2.4.1 構文的詳述関係

2つのSD式 d_1，d_2 に関して，構文的詳述関係を以下のように定義する．

(1) $d_1 = X$(変数ラベル)，d_2 が一般のSD式のとき，

$$elab_{synt}(d_1, d_2) = si(d_2) - 1$$

(2) d_1 が d_2 の一部分となっているとき (修飾形式)，これは必要条件である

$$elab_{synt}(d_1, d_2) = si(d_2) - si(d_1)$$

(3) d_1 と d_2 が同一の陳述形式同士または感情形式同士のとき，

$$elab_{synt}([s(d_{1s}), v(d_{1v})], [s(d_{2s}), v(d_{2v})])$$
$$= elab(d_{1s}, d_{2s}) + elab(d_{1v}, d_{2v})$$

$$elab_{synt}([s(d_{1s}), v(d_{1v}), c(d_{1c})],$$
$$[s(d_{2s}), v(d_{2v}), c(d_{2c})])$$
$$= elab(d_{1s}, d_{2s}) + elab(d_{1v}, d_{2v}) + elab(d_{1c}, d_{2c})$$

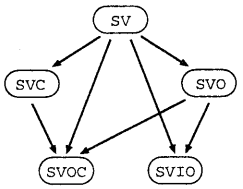
以下同様．

(4) 陳述形式の SD 式 d_1 と d_2 について，記号列 d_1 が記号列 d_2 の一部であるとき，

$$elab_{synt}([s(d_{1s}), v(d_{1v})], [s(d_{2s}), v(d_{2v}), c(d_{2c})])$$
$$= elab(d_{1s}, d_{2s}) + elab(d_{1v}, d_{2v}) + si(d_{2c}) + 1$$

$$elab_{synt}([s(d_{1s}), v(d_{1v})],$$
$$[s(d_{2s}), v(d_{2v}), i(d_{2i}), o(d_{2o})])$$
$$= elab(d_{1s}, d_{2s}) + elab(d_{1v}, d_{2v})$$
$$+ si(d_{2i}) + si(d_{2o}) + 2$$

以下同様．これらの関係を図 1 に示す．矢印の始点が先祖 (d_1) で，終点が子孫 (d_2) である．



S : $s(d_s)$
V : $v(d_v)$
O : $o(d_o)$
I : $i(d_i)$
C : $c(d_c)$

図 1: 陳述形式間の詳述関係

(5) d_1，d_2 間で詳述関係が成り立たないとき，

$$elab_{synt}(d_1, d_2) = \infty$$

2.4.2 知識に基づく詳述関係

2 つの SD 式 d_1，d_2 を特別な関係で結びつける知識データが存在するとき，両者の間には詳述関係が生じる．知識に基づく詳述関係には，

- 個別の知識に基づく詳述関係
- 一般の知識に基づく詳述関係

の 2 種類がある．知識に基づく詳述関係が成り立たないときは，

$$elab_{know}(d_1, d_2) = \infty$$

としている．

(1) 個別の知識に基づく詳述関係

個別の知識は，概念同士の関係を表す結合形式の SD 式で記述される．以下に定義の一部を示す．

- $(d_1)equa(d_2)$ (d_1 と d_2 は等しい)

$$elab_{know}(d_1, d_2) = 0$$

- $(assu(d_2))caus(d_1)$ (もし d_2 ならば d_1 である)

$$elab_{know}(d_1, d_2) = 2$$

- $(d_1)incl(d_2)$ (d_1 は d_2 を含む)

$$elab_{know}(d_1, d_2) = 3$$

(2) 一般の知識に基づく詳述関係

一般の知識に基づく詳述関係は，経験上一般に成り立つと考えられる知識によって定義される．これらは述語論理における限量記号の扱いを一般化したものと考えてよい．代表的なものを以下に示す．

- $d_1 = [s(d_a), v(d_b), \cdots]$,
  $d_2 = [s(d_a), v(nega(d_b)), \cdots]$
  (ただし，$\cdots$ 部分は同一) のとき，

$$elab_{know}(nega(d_1), d_2) = 2$$

- d_1, d_2, d_3 が名詞ラベルで，

  $elab_{know}(d_1, d_2) = 2$, $elab_{know}(d_1, d_3) = 3$

  を満たすとき，

$$elab_{know}(d_1/SOME, d_2) = 1$$
$$elab_{know}(d_1/SOME, d_3) = 1$$
$$elab_{know}(d_1/SOME, d_1/MOST) = 3$$

  などと定義する．

2.4.3 一般的な詳述関係

これまで述べたように，詳述関係には構文的なものと知識に基づくものとがある．もし 2 つの概念 d_1，d_2 間に，このどちらによっても詳述関係が成り立つならば，次のように定義する．

$$elab(d_1, d_2)$$
$$= min\{elab_{synt}(d_1, d_2), elab_{know}(d_1, d_2)\} \quad (1)$$

構文的な詳述関係については，与えられた 2 つの概念を構成するそれぞれの部分同士の詳述関係に分解できることがある．例えば，d_1，d_2 がいずれも同じ形をした陳述形式の場合は，2.4.1 の (3) の定義から次のようになる．

$$elab_{synt}([s(d_{1s}), v(d_{1v}), o(d_{1o})],$$
$$[s(d_{2s}), v(d_{2v}), o(d_{2o})])$$
$$= elab(d_{1s}, d_{2s}) + elab(d_{1v}, d_{2v}) + elab(d_{1o}, d_{2o})$$

ただし，右辺の $elab$ はいずれも (1) 式で定義されるものである．したがって，一般の詳述関係は再帰的な関係となる．

ここで，(1) 式で定義される詳述関係は 1 段の関係である．そして，この関係が連結して多段の詳述関係となることもあり得る．

2.4.4 *spec* 関係

これまで述べてきた詳述関係における厳密な因果関係だけでなく，自然言語によく現れるやや曖昧な因果関係をモデルに取り入れるために，*spec* 関係を次のように定義する．ここで，$spec_{synt}(d_1, d_2)$ を「構文的 *spec* 関係」，$spec_{know}(d_1, d_2)$ を「知識に基づく *spec* 関係」と呼ぶ．

(1) $spec_{synt}(d_1, d_2) = elab_{synt}(d_1, d_2)$

(2) $spec_{know}(d_1, d_2) = elab_{know}(d_1, d_2)$

(3) $(assu(d_2))indu(d_1)$ が与えられているとき，

$$spec_{know}(d_1, d_2) = 3$$

(4) $elab_{know}(d_1, d_2) = 2$ または
$elab_{know}(d_1, d_3) = 3$ が成り立つとき，

$$spec_{know}(d_2, d_1/MOST) = 4$$
$$spec_{know}(d_3, d_1/MOST) = 4$$
$$spec_{know}(d_1/d, d_1/MOST) = 4$$

spec 関係は詳述関係を拡張したものであり，この関係に基づいた意味の詳しさの程度を「*spec* 量 (単位：*semit*)」と呼ぶ．

2.5 意味差の定義

与えられた 2 つの概念の意味的な差異を定量的に捉えることができれば有益である．例えば，何かの事例検索において完全一致の事例が存在しないときでも，質問に類似した事例の検索が可能となる．SD 式意味モデルではそのような概念間の「意味差の尺度」を明確に定義している．

2.5.1 最近共通先祖と意味差

2 つの概念 d_1，d_2 について，*spec* 関係も含んだ詳述関係 $spec(d, d_1)$，$spec(d, d_2)$ が同時に成立するとき，この d を d_1，d_2 の「共通先祖」と呼ぶ．変数ラベル X も含めれば，共通先祖は必ず存在する．そして特に，

$$spec(d_0, d_1) + spec(d_0, d_2)$$
$$= min_d\{spec(d, d_1) + spec(d, d_2)\} = n_0$$

である d_0 を d_1，d_2 の「最近共通先祖」，n_0 を意味差と定義する．この関係を，

$$ncoa_{spec}(d_1, d_0, d_2, n_1, n_0, n_2) \quad \text{または}$$
$$ncoa_{spec}(d_1, d_0, d_2) = n_0,$$
$$diff_{spec}(d_1, d_2) = n_0$$

と表す．特別な場合として，

$$d_0 = d_1 \quad \text{または} \quad d_0 = d_2$$

となることがある．

最近共通先祖は，「与えられた 2 つの概念を包含する最も詳細な概念」である．つまり，「2 つの概念から帰納推論される最も意味情報の多い概念」であるということがいえる．そして，意味差は与えられた d_1，d_2 の最近共通先祖を探索することにより求まる．以下では，

$$ncoa(d_1, d_0, d_2, n_1, n_0, n_2)$$
$$= ncoa_{spec}(d_1, d_0, d_2, n_1, n_0, n_2),$$
$$diff(d_1, d_2) = diff_{spec}(d_1, d_2) = n_0$$

とする．

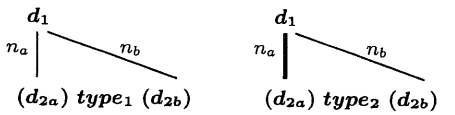
2.5.2 結合子形式に関する意味差の定義

従来の SD 式意味モデルでは，2 つの概念の一方，または双方が結合子形式の SD 式であった場合の意味差の求め方が定められていなかった．しかし要約される原文には，文節同士を様々な接続詞 (SD 式における結合子) によって結合したものが多く含まれており，当然このような文に対しても適用できる意味差の計算方法の定義が必要となってくる．本論文では以下に示す定義を新たに導入する．

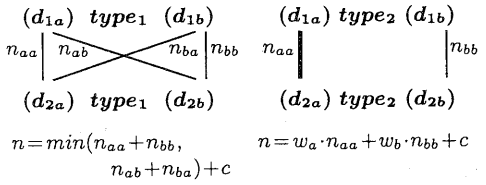
結合子形式の SD 式に関する意味差は，結合子の種類によって求め方が異なる．ここでは，結合子の左辺と右辺の概念を意味的に同じ重みで扱うことができる場合とそうでない場合があるという考えから，下の 2 つのタイプに分類する．

- $(d_1)type_1(d_2)$ <例> $plus$(結合), $pseq$(順接)
  : d_1 と d_2 の 2 つを同じ扱いができるもの
- $(d_1)type_2(d_2)$ <例> $bcau$(理由), $nseq$(逆接)
  : d_1，d_2 が主節，従属節の関係にあるもの

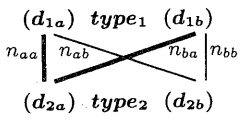
図 2 に，それぞれの組み合わせにおける意味差の求め方を示す．結合子が $type_2$ の場合，結合子によって主節，従属節が異なるが，図 2 では d_{na} を主節，d_{nb} を従属節としている．



(a): 一方が結合子形式



(b): 同じタイプの結合子形式



(c): 異なるタイプの結合子形式

図 2: 結合子形式との意味差

ここで，w_a，w_b はそれぞれ主節，従属節への重みで，上に述べた理由から $w_a > w_b$ である．本論文では一例として $w_a = 1.4$，$w_b = 0.6$ とした．また，図 2(b) では結合子が同じ場合 $c = 0$ とし，異なる場合 $c = 2$ とする．

2.5.3 意味差の正規化

2.5.1 の定義からわかるとおり，意味的になんら関係のない概念間の意味差は，2 つの SD 式の意味量が大きい程大きくなる．このままでは，互いに意味量の大きい SD 式間の意味差が関連する知識データによって少し小さくなっても，互いに意味量が小さく意味的に全く関係のない SD 式間のそれよりも大きくなってしまう．これでは両者を比較したとき，期待する結果が得られないといったことが考えられる．そこで，次式によって意味差の正規化を行う．

$$diff_{norm}(d_1, d_2) = \frac{diff(d_1, d_2)}{diff_{max}(d_1, d_2)}$$

ただし $diff_{max}(d_1, d_2)$ は，最近共通先祖と d_1，d_2 のそれぞれで構文的にも知識によっても詳述関係が成り立たなかった場合にとる最大の意味差で，例えば d_1，d_2 がともに結合子形式でなければ，$diff_{max}(d_1, d_2) = si(d_1) + si(d_2) - 2$ となる．こうして得られる正規化した意味差は，

$$0 \leq diff_{norm}(d_1, d_2) \leq 1$$

を満たす．以下の 3 つの文を用いて計算例を示す．ここでは，知識データは全くないものとして考える．

(d_a) [s(自分), v(書く/時/昨日), o(手紙)]
: 私は昨日手紙を書いた

(d_b) [s(彼), v(買う/未来), o(雑誌)]
: 彼は雑誌を買うつもりだ

(d_c) ([s(自分 ($)), v(書く/時/先週), o(手紙)])$pseq$ ([s($), v(もらう/(過去))$para$(時/今日)), o(返事)])
: 私は先週手紙を書き，今日返事をもらった

このとき，d_a と d_b，d_a と d_c から求めた意味差はそれぞれ次のようになった．

$$diff(d_a, d_b) = 87, \quad diff_{norm}(d_a, d_b) = 0.879$$
$$diff(d_a, d_c) = 110, \quad diff_{norm}(d_a, d_c) = 0.474$$

計算結果からは，正規化した意味差の方が我々の感覚に近い値を得られていることがわかる．以降は概念間の意味差として正規化した意味差を用い，$diff(d_1, d_2)$ で表すことにする．つまり，以下のように再定義する．

$$diff(d_1, d_2) = diff_{norm}(d_1, d_2)$$

3 要約の手順

3.1 概略

本論文で提案する自動要約の手順を以下に示す．

(1) 自然言語で記述された原文の SD 式への変換
(2) 知識データの作成
(3) 重要文の抽出
(4) 得られた SD 式の自然言語への変換

要約の対象となる文書中の文を原文と呼ぶことにする．まず，原文からそれに対応する SD 式を得る (1)．そして，意味差の計算時に必要となる知識データを作成する (2)．(1) で得られた SD 式と (2) で作成した知識データを利用して，重要文に対応する SD 式を抽出する (3)．その SD 式を自然言語に変換することにより要約文を得る (4)．この中で，本論文では特に (2)，(3) について詳しく説明する．

3.2 知識データの作成

原文中の各概念について，2.3 で示した方法で知識データを作成するが，この際注意しなければならないことが 2 点ある．

まず，知識データが互いに矛盾しないようにすることである．次に，各原文に特化しない，一般的な知識データの作成に努めることである．これは，本

実験システムをより汎用性の高いものにするために必要である．このことは特に，因果関係を記述した規則知識に求められる．

3.3 重要文の抽出

通常，要約文には原文の内容を決められた字数で表現することが求められる．そして，要約文により多くの情報を持たせるため，言い換えや省略が行われる．しかし，要約作業を計算機によって自動的に実現しようとする場合，まだ人間と同等の処理は望めないという考え方が一般的である．そこで，文や段落を単位とし，要約文に必要であるものを集めることで要約を実現する手法が多く提案されている．

本論文では上と同じ考えに基づき，要約に必要となる重要文の抽出を目的としている．先にも述べたように，本手法における重要文とは他の文と意味を多く共有しているものであるとする．重要文を抽出するときに必要となる意味の大小関係を測る尺度として，SD 式意味モデルの意味差を用いる．意味を多く共有しているということは，意味差の定義により，その値が小さいということに等しい．

要約処理では，まず段落に関係なくすべての文 (SD 式) 間の意味差を求める．次に各段落における最重要文を抽出する．これは各段落で，他のすべての文との意味差の平均が 1 番小さいものを最重要文とすることで実現できる．このことで要約文が局所的になることを防ぐ．次に，上で得られた重要文との間で意味差の最も小さいものを順次抽出していく．原文の文数を N とすると，要約文の文数 N' は $N' < N$ である．

4 実験例

4.1 実験方法

要約の対象として，グリム童話集の中の 1 つの物語を扱った．表 1 に物語のサイズを示す．

表 1: 実験に使用した物語

物語名	段落数	文数	文字数
指物師とろくろ職人	4	20	761

この物語の各文から人手によって SD 式を作成し，関連する知識データとして事実知識を 1 個，規則知識を 4 個登録した．下に例を示す．このような知識データの下で重要となる SD 式の抽出を行い，対応する自然言語文から要約文を得た．要約文の文数は原文の半数の 10 とした．

- $fact(([s(A), v(\text{おふれを出す}), c(B)])$
  $equa([s(A), v(\text{知らせる}), o(\text{みんな}), c(B)]))$
  : 「A が B とおふれを出す」ことは，
  「A がみんなに B と知らせる」ことに等しい
- $rule((assu([s(A), v(\text{依頼}), o(B),$
  $c([s(B), v(\text{貸す}), o(C)])]))$
  $caus([s(A), v(\text{手に入れる}), o(C)]))$
  : もし「A が B に"C を貸してくれ"と頼む」
  ならば，「A は C を手に入れる」

4.2 結果と考察

重要文抽出の評価として，被験者 (学生 10 人) による結果との比較を行った．被験者には重要と考えられる文を原文の半数選んでもらい，これを正解文とした．そして，各被験者から得られた正解文ごとに以下に示す再現率を求め，評価した．

$$\text{再現率} = \frac{\text{自動要約の結果に含まれる正解文数}}{\text{正解文数}}$$

表 2 に結果を示す．また，例として図 3 に物語の 2 段落目の原文，作成した SD 式，抽出された文を示す．各文，各 SD 式の先頭に番号を付しているが，この段落では d_8 が最重要文として選択された．

表 2: 再現率による評価

物語名	再現率 (%)		
	平均	最大	最小
指物師とろくろ職人	60	70	50

結果が完全には一致しなかった主な理由として，以下の 2 つが考えられる．まず正解文が抽出されなかった理由としては，適用できる知識データが存在しなかった，つまり，他の文との間に因果関係を見出すことができなかったためである．逆に重要でない文が抽出された理由としては，意味的には関係がなくとも，2.4.1，2.5.2 で定義したように構文的に，かつ結合子による文の構造においても非常に近い関係にあり，意味差が小さくなったためと考えられる．このような，特に前者の理由からなる結果の不一致を解決するためには，より多くの知識データが必要となる．今回登録した知識データのほとんどは原文中の 2 つの文に直接関係するものであったが，スクリプトのように幾段かの知識構造を持つ知識データを作成することができれば，結果の改善が期待できる．

d_7. その国に若い王子がいました.
d_8. 王子はろくろ職人が飛んでいるのを見て，どうかその2枚の翼を貸してくれないか，たっぷりお礼はするから，とたのみました.
d_9. そこで王子は翼を手に入れて飛んでいくと，ほかの王国にやってきました.
d_{10}. そこにはたくさんの明かりに照らされた塔がありました.
d_{11}. そこで王子は地面に降り立ち，そのわけをたずねました.
d_{12}. そして，そこには世界で一番美しいお姫さまが住んでいる，と聞かされました.
d_{13}. 王子はどうにもお姫さまのことが知りたくなって，夜になると，開いている窓のところまで飛んでいき，そこから中に入りました.
d_{14}. ところがふたりきりになっていくらもたたないうちに，ことの次第がばれてしまい，王子はお姫さまといっしょに積まれた薪の上で焼き殺されることになりました.

(a): 物語の原文

d_7. [s(王子/若い),v(存在/(過去)para(場所/国/当該))].
d_8. ([s(王子($1)),v(見る/過去),o([s(ろくろ職人),v(飛ぶ/状態)])])pseq(([s($1),v(依頼/過去),o(ろくろ職人($2)),c([s($2),v(貸す),o(翼/(枚(2))para(当該))])])bcau([s($1),v(お礼をする/程度/大)])).
d_9. ((([s(王子($)),v(手に入れる/過去),o(翼)])pseq([s($),v(飛んでいく/過去)]))pseq([s($),v(やってくる/(過去)para(場所/終点/王国/他))]))bcau(前述).
d_{10}. [s(塔($)/[s($),v(pass(照らす)),b(明かり/数/多い)]),v(存在/(過去)para(場所/当該))].
d_{11}. (([s(王子($)),v(降り立つ/(過去)para(場所/終点/地面))])pseq([s($),v(たずねる/過去),c(わけ/当該)]))bcau(前述).
d_{12}. (前述)pseq([s(王子),v(pass(聞く/過去)),o([s(お姫さま($)/[s(($)ofal(世界)),v(美しい)]),v(住む/(状態)para(場所/当該))])]).
d_{13}. ([s(王子($1)),v(知る/mood/欲求/程度/大),o(事柄/お姫さま)])pseq(([s($1),v(飛ぶ/(場所/終点/窓($2)/[s($2),v(開く/状態)])para(時/夜))])pseq([s($1),v(入る/(過去)para(場所/始点/当該)para(場所/終点/中))])).
d_{14}. (前述)nseq(([s((王子)plus(お姫さま)),v(ばれる/(過去)para(時/直後/[s((王子)plus(お姫さま)),v(ふたりきりになる/過去)])),o(ことの次第)])pseq([s((王子)plus(お姫さま)),v(pass(焼き殺す/(未来)para(場所/上/薪($)/[s($),v(pass(積む))])])])).

(b): 作成したSD式

d_8. 王子はろくろ職人が飛んでいるのを見て，どうかその2枚の翼を貸してくれないか，たっぷりお礼はするから，とたのみました.
d_9. そこで王子は翼を手に入れて飛んでいくと，ほかの王国にやってきました.
d_{11}. そこで王子は地面に降り立ち，そのわけをたずねました.
d_{12}. そして，そこには世界で一番美しいお姫さまが住んでいる，と聞かされました.
d_{13}. 王子はどうにもお姫さまのことが知りたくなって，夜になると，開いている窓のところまで飛んでいき，そこから中に入りました.

(c): 抽出された文

図 3: 要約例

5 おわりに

本論文では，SD式意味モデルによる文間の意味差に注目した要約手法を提案した．まだ対象とした物語の数が少なく十分とはいえないが，文書の自動要約の一手法として本手法が有効であることが確認できた.

今後は，より多くの知識データの蓄積を目指し，更に，抽出された各文の簡略化を行うことができれば要約率の向上が図れる．また，他の要約手法による結果との比較を行い，各手法による違いを検討し本手法の改良に役立てたい.

参考文献

[1] 奥村学, 難波英嗣: " テキスト自動要約に関する研究動向", 言語処理学会論文誌, Vol. 6, No. 6, pp. 1-26(1999)

[2] G. Salton, et al.: "Automatic text decomposition using text segments and text themes", Proc. of the 7th ACM Conference on Hypertext, pp. 53-65(1996)

[3] 中澤俊哉, 重永実: " エピソードネットワークを用いた物語のあらすじ生成", 情報処理学会論文誌, Vol. 32, No. 10, pp. 1215-1224(1991)

[4] Eiji Kawaguchi, et al.: "The Semantic Metric Computation Scheme in the SD-Form Semantics Model", Proc.PRICAI, pp. 623-629 (1993)

[5] Grimm Jacob, Grimm Wilhelm 共著, 吉原高志, 吉原素子 共訳: " 初版グリム童話集", 白水社 (1997)