SD式意味モデルを利用したテキストベースステガノグラフィ

新見道治[†] 峯脇さやか[†,☆]
野田秀樹[†] 河口英二[†]

自然言語文の意味情報を利用したステガノグラフィを提案する．本手法では，著者らが提案しているSD式意味モデルを利用して自然言語文を意味記述し，その記述されたデータに対して埋込み抽出処理を行う．SD式意味モデルでは，文の意味情報をSD式と呼ばれる記号列で記述し，それらに意味量と呼ばれる情報量が数値として対応する．本手法での埋込み処理とは，意味量と秘密情報の値を一致させることである．すなわち，与えられた自然言語文に対応するSD式の構造を変化させることによってそのSD式の意味量を埋め込もうとする秘密情報の二進数値に一致させるものである．意味量を増加させる場合は，SD式内の抽象的な概念を具体的な概念で置き換え，逆に意味量を減少させる場合は，具体的な概念を抽象的な概念で置き換える．これにより，秘密情報が埋め込まれた文にも意味的な矛盾は生じない．秘密情報を抽出する場合，SD式の意味量を計算することにより，秘密情報を取り出すことができる．実験では，新聞記事と招待状の2つのテキストデータに対してデータ埋込み実験を行い，提案法の妥当性を検討した．

Linguistic Steganography Using SD-Form Semantics Model

Michiharu Niimi,[†] Sayaka Minewaki,[†] Hideki Noda[†]
and Eiji Kawaguchi[†]

This paper proposes a method for linguistic steganography in consideration of the meaning of natural language sentences. To describe the meaning of the sentences, this method uses SD-Forms, which is a meaning description form consisting of symbols, developed by the authors. An SD-Form is assigned an amount of semantic information of a sentence. The amount of the meaning of sentences is used to carry secret information on text data. In embedding secret information, firstly the sentences are transformed to SD-Forms and then the amount of semantic information of SD-Forms is decreased or increased to coincide with the value of the secret information. In decreasing the amount of semantic information, concepts are substituted with its upper concepts, on the other hand, concepts are substituted with its lower concepts in increasing the amount. We can expect the sentence with secret information embedded is consistent in the meaning. We report the feasibility of the proposed method by experiments in which texts data in newspaper and invitation are used.

1. はじめに

ブロードバンドネットワークが目覚ましい勢いで普及しており，今後インターネットを利用したコミュニケーションがさらに身近になることは間違いない．インターネットを流れる情報の中には，プライバシーに関わるものも多く，その情報保護が重要である．近年，セキュリティ技術の1つとしてステガノグラフィが注目されている．ステガノグラフィとは，秘密データの存在そのものを第三者に気づかせないようにする技術である．これは，秘密データを他のデータに隠すことで実現される．

なんらかの情報を他のメディアに埋め込む技術は，総称してインフォメーションハイディングと呼ばれている．数年前より，その一応用分野である電子透かしが脚光を浴びている．ステガノグラフィはインフォメーションハイディングの一応用分野として考えられるが，電子透かしと比較すると大容量の秘密情報を埋め込めることが必要であるという点が大きく異なる．ステガノグラフィの先行研究では，秘密情報を埋め込むデータとしてディジタル画像や音声データがよく用いられてきた．本研究では，従来あまり取り上げられなかっ

† 九州工業大学工学部
Faculty of Engineering, Kyushu Institute of Technology
☆ 現在，九州工業大学大学院情報工学研究科
Presently with Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

たテキストデータに対して秘密データを隠蔽する．

この分野における従来研究に関しては，文献 1)〜3) が詳しい．テキストに情報を埋め込む技術は，文章の字面情報は変更せずにフォーマットを変更する，文章の字面自体を変更する，それ以外の情報を変更する方式に分類できる．フォーマット変更方式では，埋め込める情報量はわずかであり，機械的処理により秘密情報の存在が簡単に第三者に知られてしまう可能性が高い．文書の内容そのものを書き換える手法は，辞書変換法と呼ばれており，あらかじめ用意した辞書に従って，元の文書を書き換える．その技術の 1 つである文献 4) の手法は，秘密情報と辞書から秘密情報を埋め込んだ文章を生成している．しかしながら辞書には文法知識が組み込まれておらず，秘密情報を埋め込んだ後の文章の意味は理解できない．一般に，辞書変換法では単語置換による意味情報の変化が問題にされており，この点を解決するために，文献 1), 2) では，意味的情報を保存する辞書ベースの情報ハイディング方法を提案している．これらの方法では，テキストデータがあらかじめ与えられており，その文章中の単語等を辞書を参照しながら置換することで情報を埋め込む．

本稿では，テキストデータに対するステガノグラフィの一手法を提案する．提案法の特徴は，テキストの意味的情報量を利用する点にある．具体的には，SD 式意味モデル[5),6)] を利用する．SD 式（Semantic-structure Description Form）[5),6)] とは，著者らの研究グループが提案している自然言語概念の意味表現形式である．SD 式は，自然言語における個々の概念，陳述表現，感情表現，あるいはシステムに与える知識データ等を記述するための一種の中間言語である．SD 式は，何かの概念を表現するだけでなく，その概念の意味的な情報量の大小も表すことができる．それぞれの SD 式が持つ意味的な情報量のことを “意味量” と呼ぶ．一般に簡単な構造の SD 式の意味量は小さく，複雑なものは大きい．自然言語文を SD 式に変換することにより，文の意味を定量的に扱うことができる．すなわち，本稿は自然言語文と SD 式間の変換が機械的に可能であるとの前提に立ち，自然言語文（日本語）を利用したステガノグラフィを検討するものである．本稿では，SD 式意味モデルに概念の階層構造を導入し，情報ハイディングに適するような SD 式意味モデルへの改良も行う．

提案法では，意味的情報量に秘密情報を埋め込むが，文に対する意味的情報量の与え方にはいくつかの方法が存在する．たとえば，潜在的意味解析[7)] と呼ばれる手法では，文章中に隠された単語の意味的あるいは概念的な相関関係を用いて，次元数を削減することにより単語の意味的な特徴を自動的に抽出している．テキスト検索においてはその特徴量の有効性が報告されている[8)]．このような手法により得られた数値データを，意味的情報量として文に与えれば，その情報量に秘密情報を対応させることは可能であると思われる．SD 式意味モデルを利用する利点は，意味的情報が記号列で記述されているので，知的処理が容易に実現できる点にある．つまり本研究では，“情報埋め込みにより発生する意味変化を，意味記述された記号列（SD 式）に対する知的処理によりなるべく小さくする” ことを最終的な目標としており，そのために SD 式意味モデルを利用し，そのモデル中で提案されている意味的情報量を利用することにした．

以下，2 章では，SD 式による概念の意味記述について述べる．3 章では，SD 式を利用した言語ステガノグラフィの手法について述べる．ここでは，概念の階層構造について言及する．4 章で実験例を示し，提案法について検討する．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[5),6)]

自然言語概念を SD 式で記述し，その記述データを基にして意味処理を行うモデルを「SD 式意味モデル」と呼んでいる．

2.1 SD 式による意味記述

SD 式は，自然言語における概念の意味構造を記述するための中間言語の一種であり,「SD 式記号」と呼ばれる概念ラベル，修飾子，規定子，結合子等から構成される記号列である．本稿で必要となる SD 式記号を簡単に説明する．

- 変数概念ラベル（変数ラベル）
  変数ラベルは，“何か”，“或るもの” のような意味を記述する場合に用い，たとえば,「X」のように記述する．
- 単純概念ラベル（単純ラベル）
  単純概念ラベルとは，自然言語における単語に相当する概念を記述するために用い，たとえば，“車” という概念を「車」で記述する．
- 修飾形式
  記号「/」を修飾子と呼ぶ．この記号の右側で左側を修飾している．たとえば,「本/料理」は “料理の本” を表す．
- 規定子形式
  SD 式に特別な役割，たとえば，否定，受け身等を記述するために用いる．“買わない” という概

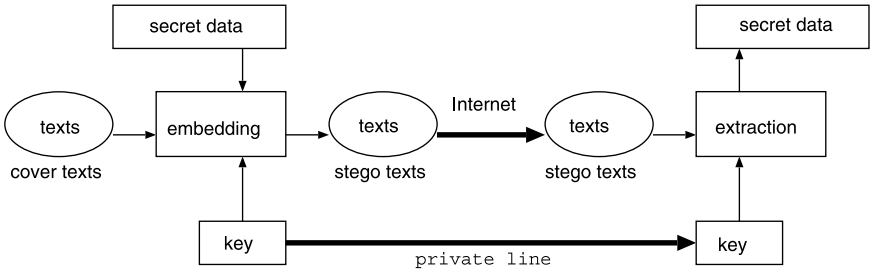

図 1 テキストベースステガノグラフィシステム
Fig. 1 System for text-based steganography.

念は,「*nega*(買う)」と記述できる.

- 結合子形式
  2 つの概念を何らかの関係で結合して新しい概念を作るために利用する.たとえば,“老人と子供”は,「(老人)*plus*(子供)」と記述できる.
- 陳述形式
  陳述 SD 式とは,陳述的な概念構造を記述するために利用する.以下の 6 種類の機能項目記号と呼ばれる意味表現上の機能を指定するために導入された記号を利用して,記述する.
  - 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

  陳述 SD 式形式は英語の基本的な文型に模して定義されている.たとえば,“私は,毎日テニスをする.” は,「[s(自分),v(テニス/時/毎日)]」と記述される.
- 区切り記号
  「[ ]」,「( )」,「,」は区切り記号と呼ばれており,Prolog 形式で記述するために用いられる.

SD 式では,これらの記号を組み合わせた記号列により,自然言語文の意味情報を記述する.記号列は,SDG(SD-form Generative Grammar)と名付けられた文脈自由文法により生成される.

2.2 SD 式の意味的情報量

SD 式意味モデルでは,意味を定量的に処理するために,各 SD 式記号に対して意味素量と呼ばれる値が設定されている.ある SD 式に対して,その SD 式を構成する SD 式記号の意味素量をすべて足し合わせた値を,その SD 式の意味量と定義する.SD 式を D とすると,その意味量を

$$si(D) = n$$

と表記し,単位を *semit* と呼んでいる.SD 式意味モデルでは,SD 式記号の意味素量の値は SD 式意味モデルの利用者が独自に定めることができる,と規定している.たとえば,文献 6) では以下のように意味素量を設定している.

(1) 変数ラベル「X」,「Y」,「Z」, $\cdots$:1 [*semit*]
(2) 単純ラベル「車」,「買う」, $\cdots$:10 [*semit*]
(3) 修飾子「/」:1 [*semit*]
(4) 規定子「*nega*」,「*only*」, $\cdots$:2 [*semit*]
(5) 結合子「*para*」,「*equa*」, $\cdots$:1 [*semit*]
(6) 機能項目記号「s」,「v」,「o」, $\cdots$:1 [*semit*]
(7) 区切り記号「[ ]」:1 [*semit*]
(8) 区切り記号「( )」,「,」:0 [*semit*]

この例の意味素量に従う場合,「D=[s(相手),v(読む/過去),o(本/当該)]」の意味量は,

$$si(D) = 56$$

と計算できる.

3. 提 案 法

3.1 テキストベースステガノグラフィシステム

本稿で取り扱うテキストベースステガノグラフィシステムを**図 1** に示す.本システムは,ステガノグラフィを秘匿通信手段として利用するものである.秘密情報を送信する者と,受信する者が存在する.送信者は,あらかじめ秘密情報を埋め込むためのテキストデータを用意し(以後,カバーテキストと呼ぶ),カバーテキストに秘密情報を埋め込む.カバーテキストに秘密情報が埋め込まれたテキストデータをステゴテキストと呼ぶ.送信者はステゴテキストをインターネットを通して受信者に送る.鍵は安全な通信路を利用して受信者に送る.ステゴテキストの受信者は,鍵とステゴテキストから埋め込まれた秘密情報を抽出する.

3.2 ステゴテキスト生成の概略

ステゴテキスト生成までの概略を**図 2** に示す.本稿では,カバーテキストを $\mathrm{CT}_i(i = 0, 1, \cdots, N-1)$,ステゴテキストを $\mathrm{ST}_i(i = 0, 1, \cdots, N-1)$ と表記し,またカバーテキストの SD 式を $\mathrm{CS}_i(i = 0, 1, \cdots, N-1)$,

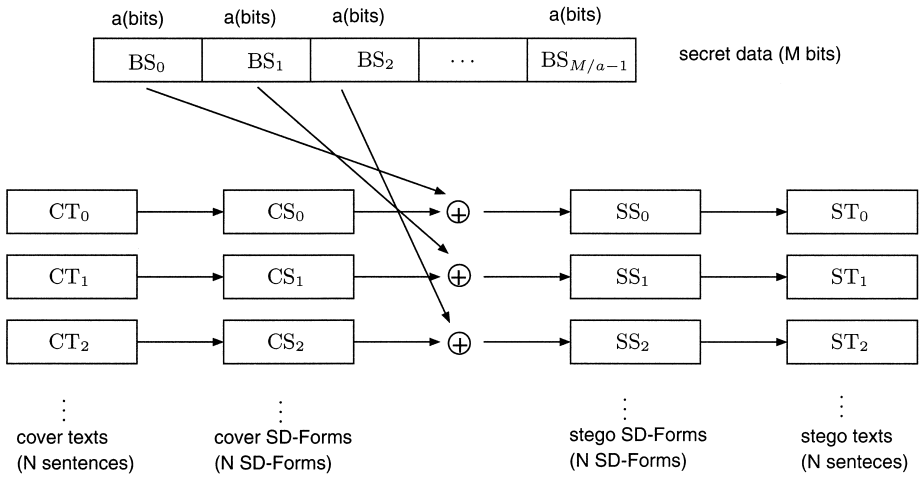

図 2 提案法におけるステゴテキスト生成の概略
Fig. 2 Outline of generating stego texts in proposed method.

ステゴテキストの SD 式を $\mathrm{SS}_i(i = 0, 1, \cdots, N-1)$ と表記する．ここで，CS をカバー SD 式，SS をステゴ SD 式と呼ぶ．コンピュータ処理では記号や文字データ等もすべて数値データとして扱うことができるので，秘密情報はすべて二進数値として考えることができる．よって，秘密データを M ビットと仮定し，そのビット系列の先頭から a ビットごとにブロック化したビット系列を $\mathrm{BS}_j(j = 0, 1, \cdots, M/a-1))$ と表記する．なお本稿では，

$$N = M/a$$

と仮定している．また，二進数 b の十進数値を $de(b)$ として表記する．すなわち，

$$0 \leq de(\mathrm{BS}_j) \leq 2^a - 1 \quad (j = 0, 1, \cdots, M/a-1)$$

である．

提案法における**埋込み**とは，$si(\mathrm{CS}_i)$ を $de(\mathrm{BS}_j)$ に一致させる処理のことである．SD 式の意味量を増減して，秘密データを構成する部分的な値に一致させる操作を “SD 式意味量と秘密データを対応付ける” と呼ぶことにする．対応付けるためには，SD 式の構造を操作し，意味量を増減させなければならない．本稿では，**SD 式の構成要素に着目した意味量操作**と**単純ラベルの階層構造による意味量操作**により意味量と秘密データを対応付ける．

3.3 SD 式の構成要素に着目した意味量操作

3.3.1 修飾子の操作

修飾子と修飾 SD 式を削除することにより，意味量を減少させることができる．ただし，主 SD 式と修飾 SD 式は単純ラベルとする．たとえば，「D_1 = 車/赤い」の場合「D_2 = 車」となる．この場合，意味量の変化量は，

$$|si(D_1) - si(D_2)| = |21 - 10| = 11$$

となる．

3.3.2 結合子の操作

結合子と結合子の右側の SD 式を削除することにより，意味量を減少させることができる．たとえば，「D_1 = (ビール)$plus$(焼酎)」の場合「D_2 = ビール」となる．この場合，意味量の変化量は，

$$|si(D_1) - si(D_2)| = |21 - 10| = 11$$

となる．

3.3.3 規定子の操作

規定子を削除することにより，意味量を減少させることができる．たとえば，「D_1 = [s(太郎),v(聞く),o($only$(意見/花子))]」の場合，「D_2 = [s(太郎),v(聞く),o(意見/花子)]」となる．この場合，意味量の変化量は，

$$|si(\mathrm{D}_1) - si(\mathrm{D}_2)| = |47 - 45| = 2$$

となる．

3.4 単純ラベルの階層構造による意味量操作

単純ラベルの置換により意味量操作を試みる．従来の SD 式意味モデル[5),6)] では，単純ラベルの意味素量がすべて同じ値なので，単純ラベルを置換することによる意味量増減は発生しない．しかしながら，単純ラベルの中にも抽象化具体化した概念の上位下位の関係を持つラベルもある．提案法では，この概念の上位下位関係も利用して意味量の増減を試みる．そこで，単純ラベルに階層構造を導入し，異なった意味素量を割り当てることにする．具体的には，分類語彙表[9)] を用いて単純ラベルの階層構造を SD 式意味モデル中に構築した．

```
2．用の類
2.1　抽象的関係

2.1110　関係
2.1120　異同
2.1130　相対
2.1200　存在
2.1210　出没
2.1220　成立・発生 復活
2.1230　成立・失敗
2.1240　保有・残存
2.1250　消滅・除去
2.1300　整備
2.1310　でき・利き
2.1320　はずれ・損じ
2.1330　取合せ・つりあい
```

図 3 分類語彙表の項目例
Fig. 3 A part of topics in Japanese thesaurus.

```
かかわりあう，かかわりあう，2.1110,1,1,3 かかわる，
かかわる，2.1110,1,1,2
かかりあう，かかりあう，2.1110,1,2,2
かかる，かかる，2.1110,1,2,1
相互作用，そうごさよう，2.1110,1,3,1
あう，あう，2.1110,1,3,2
なすりあう･とりあう･愛し合う，なすりあう，2.1110,1,3,3
縁がある，えんがある，2.1110,1,4,3
関係する，かんけいする，2.1110,1,4,1
関する，かんする，2.1110,1,4,2
関連する，かんれんする，2.1110,1,5,1
相関する，そうかんする，2.1110,1,5,3
連関する，れんかんする，2.1110,1,5,2
あずかる，あずかる，2.1110,1,6,1
かかずらう，かかずらう，2.1110,1,6,2
```

図 4 分類語彙表の索引例
Fig. 4 A part of indexies in Japanese thesaurus.

3.4.1 単純ラベル階層構造の構築

分類語彙表のフロッピー版[9]には，本表，項目，索引というまとまりごとにデータが収録されている．本表には，語が分類番号順に収められ，分類番号には意味上の分類を示す分類項目名が与えてある．項目には，木構造での葉以外の節に対応する語が収められている．索引には，木構造での葉に対応する各語がそのコードとともに収められている．

項目と索引の一部をそれぞれ**図 3**，**図 4** に示す．項目には，「“分類番号”，“語”」という形式でデータが収められている．また，索引には，「“表記”，“読み”，“分類番号”，“段落番号”，“段落内行番号”，“行内語番号”」という形式でデータが収められている．項目と索引に共通する分類番号は，*a.bcde* という形で与えられている．ここで，a〜e は整数で，$1 \leq a \leq 4$，$0 \leq b, c, d, e \leq 9$ である．a は品詞を，*bcde* は意味分類を示している．

分類語彙表を利用して単純ラベルを 4 つの階層に分割する．その階層レベルを，$\mathrm{LV}_i (i = 1, \cdots, 4)$ と表記する．LV_1 が最も抽象的な概念を表し，LV_4 が最も具体的な概念を表す．具体的には以下の方法に従い，単純ラベルの階層構造を構築した．今，処理対象としている単純ラベルを $word_1$ とし，$word_1$ の分類番号を *a.bcde* とする．さらに，$word_1$ の親に相当する概念ラベルを $word_2$ とする．

(1) $word_1$ が索引内の語である場合
$word_1$ は葉になるので，LV_4 階層とし，$word_1$ と等しい分類番号を持つ項目内の語を $word_2$ とする．$word_2$ は LV_3 階層に属する．

(2) $word_1$ が項目内の語である場合
(2-1) $e \neq 0$ の場合
*a.bcd*0 の分類番号を持つ項目内の語を $word_2$ とし，$word_1$ を LV_3 階層，$word_2$ を LV_2 階層とする．
(2-2) $d \neq 0$ の場合
*a.bc*00 の分類番号を持つ項目内語を $word_2$ とし，$word_1$ を LV_2 階層，$word_2$ を LV_1 階層とする．

3.4.2 単純ラベルの意味素量の割当て

単純ラベルの階層構造では，上位階層は抽象的な単純ラベルの集合，下位階層は具体的な単純ラベルの集合となる．本稿で述べる手法では，各階層に属する単純ラベルの意味素量について，階層ごとに異なる値を割り当てる．

ユーザが任意に決定した単純ラベルの意味素量を t [*semit*] とし，レベルが 1 段変わるたびに x [*semit*] 意味量が増減するとする．このとき，レベル k（k=1,2,3）階層に属する概念ラベルの意味素量を次式で与える．

$$t + x \times k \qquad (1)$$

たとえば，「D = 話」は LV_2 階層なので，$t = 10$，$x = 1$ とすると，

$$si(D) = 10 + 1 \times 2 = 12$$

となる．

LV_4 階層は，すべてが索引内の語の集合となる最下位階層である．この階層において，同じ分類番号と段落番号を持つ語は，意味的に非常に類似している．このような語の集合について，互いに違う意味素量を持つよう，概念ラベルの値を設定する．具体的には以下の式で計算する値を，LV_4 階層に属する概念ラベルの意味素量とする．

$$t + 4 \times x + (m - 1) \times l \quad (m = 1, 2, \ldots) \qquad (2)$$

ここで，l はユーザが任意に定める定数であり，m は出現順である．たとえば，「D=学園」は意味分類上 2

番目に出現するので，$t = 10$，$x = 1$，$l = 1$ とすると，

$$si(D) = 10 + 4 \times 1 + ((2-1) \times 1) = 15$$

となる．

3.5 $si(\mathbf{CS}_i)$ と $de(\mathbf{BS}_j)$ に関する考察

カバー SD 式の意味量と，埋め込む値に関して考察し，埋め込む値を加工する必要があることを明らかにし，その解決方法を示す．

カバー SD 式に対する操作量が大きくなると，ステゴテキストとカバーテキスト間の意味が大きく変化すると考えられる．つまり，少ない操作量が望ましい．送信者は，1 つの SD 式に埋め込む情報量を a ビットと設定しているので，

$$si(\mathrm{CS}_i) > 2^a$$

の場合，大幅な意味量操作が必要になる場合がある．

また，前述したように SD 式は SDG と呼ばれる文脈自由文法により生成されるので生成された SD 式は規則的な形式になっている．さらに SD 式では，意味素量により意味量が決定されてしまう．つまり，1 つの SD 式が持つ意味量の下限値が存在することになる．その値を SD_{min} とすると，

$$de(\mathrm{BS}_j) < \mathrm{SD}_{min}$$

であるような値は埋め込めないことになる．

これら 2 つの問題を解決するために BS_j の値をもとにして，都合の良い埋め込む値を以下のように決定する．0 以上の整数の集合を I とする．$z \in I$ とし，z の関数 $h(z)$ を以下の式で定義する．

$$h(z) := de(\mathrm{BS}_j) + 2^a \times z$$

カバー SD 式の意味量と $h(z)$ との差の絶対値を $F(z)$ とする．すなわち，

$$F(z) = |si(\mathrm{CS}_j) - h(z)|$$

である．$F(z)$ に最小値を与える z_{min} を，

$$h(z) \geq \mathrm{SD}_{min}$$

の条件のもとで求める．つまり，

$$z_{min} = \arg\min_z F(z), \quad h(z) \geq \mathrm{SD}_{min}$$

を求めることになる．提案法では $si(\mathrm{CS}_i)$ を $h(z_{min})$ に一致させることにする．

3.6 意味量操作のアルゴリズム

CS_i に対する，具体的な秘密データ埋込み手順を以下に示す．

(1) $si(\mathrm{CS}_i)$ を計算する．
(2) 3.5 節により BS_j と $si(\mathrm{CS}_i)$ から，実際に埋め込む値 $h(z_{min})$ を求める．
(3) $h(z_{min})$ と $si(\mathrm{CS}_i)$ の差を求める．
(4) $h(z_{min}) > si(\mathrm{CS}_i)$ の場合 (4-1) の意味量増加処理を，$h(z_{min}) < si(\mathrm{CS}_i)$ 場合 (4-2) の意味量減少処理を行う．$h(z_{min}) = si(\mathrm{CS}_i)$ の場合，秘密情報が埋め込まれた SD 式は CS_i となる．

(4-1) 入力 SD 式中のすべての単純ラベルについて，$h(z_{min})$ と $si(\mathrm{CS}_i)$ が一致するまで，順番に以下の操作を行う．
- LV_1 に属する単純ラベルを LV_2 の単純ラベルに置換する．
- 同様に $\mathrm{LV}_2 \rightarrow \mathrm{LV}_3$ を行う．
- 同様に $\mathrm{LV}_3 \rightarrow \mathrm{LV}_4$ を行う．
- LV_4 に属する単純ラベルについて，より大きい意味素量の単純ラベルに置換する．

(4-2) $h(z_{min})$ と $si(\mathrm{CS}_i)$ 一致するまで，順番に以下の操作を行う．
- 修飾子の操作
- 結合子の操作
- 規定子の操作
- LV_4 に属する単純ラベルについて，より小さい意味素量の単純ラベルに置換する．
- LV_4 に属する単純ラベルを LV_3 の単純ラベルに置換する．
- 同様に $\mathrm{LV}_3 \rightarrow \mathrm{LV}_2$ を行う．
- 同様に $\mathrm{LV}_2 \rightarrow \mathrm{LV}_1$ を行う．

3.7 秘密情報の抽出

受信したテキストから秘密データを抽出する場合，まず受信したテキスト (ST_i) を SD 式に変換 (SS_i) し，その意味量 $(si(\mathrm{SS}_i))$ を求める．このとき，テキストに意味的な変化がなければ，

$$si(\mathrm{SS}_i) = h(z_{min})$$

の関係が成り立つ．この $si(\mathrm{SS}_i)$ から秘密情報 $de(\mathrm{BS}_j)$ を求めるには，式 (3) を用いる．

$$de(\mathrm{BS}_j) = si(\mathrm{SS}_i) \pmod{2^a} \tag{3}$$

この式から分かるように，ステゴテキストの受信者は a を知っていれば，埋め込まれた情報を正確に復元できる．送信者受信者間で a は秘密鍵となる．

4. 埋込み実験と考察

3 章で示した手法を計算機上に実装し，秘密データの埋め込み実験を行い，提案法について検討した．

4.1 実 験 方 法

提案法を，Perl v.5.6.1 for MSWin32 を用いて計算機上に実装した．カバーテキストとして，新聞記事と招待状を利用した．それぞれ**図 5** と**図 6** に示す．それらのテキストデータを SD 式に変換したものを，それぞれ**図 7** と**図 8** に示す．本実験では，自然言語文から SD 式および SD 式から自然言語文への変換は人手により行った．1 つの SD 式に埋め込む秘密情報

```
D1：米国は，2 日，2012 年夏季五輪の招致候補都市を
                                ニューヨークに決定した．
D2：米国内の最終選考にはニューヨークと
                                サンフランシスコが残っていた．
D3：米国オリンピック委員会は理事会の投票で
                                ニューヨークを選出した．
D4：国際オリンピック委員会は 2005 年に開催都市を
                                決定する．
D5：ニューヨーク市の招致委員会はオリンピックの
        開催のために総予算 50 億ドルの計画を発表した．
D6：マイケル・ブルームバーグ市長は会見で招致実現
                                に向け強い決意を示した．
```

図 5 カバーテキスト（新聞記事）
Fig. 5 Cover texts (newspaper).

```
D1：こんにちは，九工大の太郎です．
D2：私たちの結婚式について皆様にお知らせします．
D3：私と花子は，8 月 6 日に結婚式をあげるつもりです．
D4：そこで，恩師の先生には結婚式に出席して
                                もらいたい思っています．
D5：場所は，大阪か神戸の式場にしようと考えています．
D6：詳細が決まり次第，改めてご連絡いたします．
```

図 6 カバーテキスト（招待状）
Fig. 6 Cover texts (invitation).

```
D1：[s(米国),v(決定/(過去)para(時/日 (2))),
        o(招致候補都市/夏季オリンピック/年 (2012)),
                                c(ニューヨーク)]
D2：[s(選考/最終/米国内),v(残る/(過去)para(進行)),
         c((ニューヨーク)plus(サンフランシスコ))]
D3：[s(委員会/オリンピック/米国),v(選出/(過去)
          para(手段/投票/理事会)),o(ニューヨーク)]
D4：[s(委員会/オリンピック/国際),v(決定/(未来)
                    para(時/年 (2005))),o(開催都市)]
D5：[s(招致委員会/ニューヨーク市),v(発表/
      (過去)para(目的/開催/オリンピック)),o(計画/
             [s(総予算),v(である),c(50(億ドル))])]
D6：[s(マイケル・ブルームバーグ (市長)),v(示す/
        (過去)para(対象/招致実現)para(場所/会見)),
                                o(決意/強い)]
```

図 7 カバー SD 式（新聞記事）
Fig. 7 Cover SD-Forms (newspaper).

の最大ビット長を $a = 7$ [ビット] とし，図 7 と図 8 の SD 式に対して，それぞれ 42 ビットの秘密情報を埋め込んだ．秘密情報には乱数で発生させた値を用いた．なお，本実験では文献 6) で示されている意味素量を用い，単純ラベルを SD 式の最小構成要素と考え，$\mathrm{SD}_{min} = 10$ とした．

4.2 実験結果

表 1 と**表 2** に埋込み前後の各 SD 式の意味量，秘密データの値，必要な意味量の増減を示す．秘密情報を埋め込んだ SD 式を**図 9** と**図 10** に示す．また，それらの SD 式を自然言語文に変換したものを**図 11** と**図 12** に示す．以下では，埋込み処理の例として，新聞

```
D1：([a(挨拶/時/昼)])pseq([s(自分),v(である),
                                c(太郎/九工大)])
D2：[s(自分),v(知らせる),i(皆様),o(結婚式/我々)]
D3：[s((自分)plus(花子)),v(あげる/(未来)
                para(時/月 (8)/日 (6))),o(結婚式)]
D4：([s(自分),v(願う),o([s(先生 ($1)/[s($1),
   v(である),c(恩師)]),v(出席/対象/結婚式)])])
                                bcau(前述)
D5：[s(自分),v(考える/状態),o([s(場所),v(する),
                        c(式場/(大阪)orxx(神戸))])]
D6：[s(自分),v(連絡/(改めて)para(終点/[s(詳細),
                                v(決まる)]))]
```

図 8 カバー SD 式（招待状）
Fig. 8 Cover SD-Forms (invitation).

表 1 埋込みの前後における各 SD 式の意味量と秘密データ（新聞記事）
Table 1 Semantic information of each SD-Form in before and after embedding and secret data (newspaper).

	埋込み前	秘密データ	埋込み後	増減
D1	179	6	134	−45
D2	130	1	129	−1
D3	211	25	153	−58
D4	133	8	136	+3
D5	155	28	156	+1
D6	117	1	129	+12

表 2 埋込みの前後における各 SD 式の意味量と秘密データ（招待状）
Table 2 Semantic information of each SD-Form in before and after embedding and secret data (invitation).

	埋込み前	秘密データ	埋込み後	増減
D1	98	101	101	+3
D2	100	113	113	+13
D3	125	89	89	−36
D4	146	126	126	−20
D5	131	116	116	−15
D6	121	119	119	−2

記事の D6 と招待記事の D3 に関して具体的に述べる．

まず，新聞記事の D6 に対する埋込み処理を説明する．表 1 より，D6 に埋め込む秘密データの値は “1” である．このとき $h(z)$ は，

$$h(z) = 1 + 2^7 \times z$$

と表現でき，$F(z)$ は，

$$F(z) = |117 - h(z)|$$

と表現できる．$h(z) \geq 10$ のもとで $F(z)$ に最小値を与える z_{min} は，

$$z_{min} = 1$$

と求まるので，

$$h(z_{min}) = 129$$

となり，si(D6) を “129” に変更することにより秘密情報 “1” を埋め込む．つまり，D6 に関しては意味量

```
D1：[s(米国),v(確定/(過去),o(招致候補都市),
                                    c(ニューヨーク)]
D2：[s(選考/最終/米国内),v(残す/(過去)para(進行)),
        c((ニューヨーク)plus(サンフランシスコ))]
D3：[s(委員会/オリンピック/米国),v(簡抜/過去),
                                    o(ニューヨーク)]
D4：[s(委員会/オリンピック/国際),v(既定/
          (未来)para(時/年 (2005))),o(開催都市)]
D5：[s(招致委員会/ニューヨーク市),v(発表/
            (過去)para(目的/開催/JOC)),o(計画/
          [s(総予算),v(である),c(50(億ドル))])]
D6：[s(マイケル・ブルームバーグ (市長)),v(示す/
       (過去)para(対象/招致実現)para(場所/会見)),
                                    o(見切り/根強い)]
```

図 9 ステゴ SD 式（新聞記事）
Fig. 9 Stego SD-Forms (newspaper).

```
D1：([a(挨拶/時/昼)])pseq([s(自ら),v(である),
                                    c(太郎/九工大)])
D2：[s(自分),v(知らせる),i(各々方),
                                    o(銅婚式/我々)]
D3：[s((自分)plus(花子)),
                    v(あげる/(未来)), o(三三九度)]
D4：[s(自己),v(志望・反省),o([s(先生 ($1)/
          [s($1),v(である),c(師匠)]),
             v(出欠/対象/祭儀・式・宗教的行為)])]
D5：[s(自分),v(存じる/状態),o([s(場所),
                              v(する),c(宴席/大阪)])]
D6：[s(自分),v(連絡/(改めて)
                  para(終点/[s(委細),v(決まる)]))]
```

図 10 ステゴ SD 式（招待状）
Fig. 10 Stego SD-Forms (invitation).

を “+12(= 129 − 117)” する必要がある．3.6 節 (4-1) に従うと，まず上位概念の単純ラベルを下位概念の単純ラベルで置換しなければならないが，D6 には LV_1 から LV_3 に属する単純ラベルは存在しない．よって “LV_4 に属する単純ラベルを，より大きい意味素量の単純ラベルに置換” する．D6 には LV_4 に属するいくつかの単純ラベルが存在するので，その中からランダムに 1 つの単純ラベルを選択する．この場合，まず「決意」が選ばれた．意味量を最も増加させる場合は「見切り」との置換になり，意味量は “+3” となる．さらに意味量の増加が必要なので，再度ランダムに単純ラベルを選択すると「強い」が選ばれた．意味量を最も増加させる場合は「根強い」との置換になり，“+9” となる．先ほどの増加値と合わせると，意味量は “+12” となり目的の SD 式が得られた．

続いて，招待状の D3 に対する埋込み処理を説明する．表 2 より，D3 に対して埋め込む秘密データの値は “89” である．このとき $h(z)$ は，

$$h(z) = 89 + 2^7 \times z$$

```
D1：米国は，招致候補都市をニューヨークに確定した．
D2：米国内の最終選考にはニューヨークと
                    サンフランシスコを残していた．
D3：米国オリンピック委員会は,ニューヨークを簡抜した．
D4：国際オリンピック委員会は，
          2005 年に開催都市を既定する予定だ．
D5：ニューヨーク市の招致委員会は，JOC 開催の
        ために総予算 50 億ドルの計画を発表した．
D6：マイケル・ブルームバーグ市長は，
    会見で招致実現に向け根強い見切りを示した．
```

図 11 ステゴテキスト（新聞記事）
Fig. 11 Stego texts (newspaper).

```
D1：こんにちは，私が九工大の太郎です．
D2：各々方に，我々の銅婚式を知らせます．
D3：私と花子は，三三九度をあげる予定です．
D4：師匠である先生には祭儀・式・宗教的行為に出欠
                    することを志望・反省します．
D5：場所は大阪の宴席にすることを存じています．
D6：委細が決まり次第改めて連絡します．
```

図 12 ステゴテキスト（招待状）
Fig. 12 Stego texts (invitation).

と表現でき，$F(z)$ は，

$$F(z) = |125 - h(z)|$$

と表現できる．$h(z) \geq 10$ のもとで $F(z)$ に最小値を与える z_{min} は，

$$z_{min} = 0$$

と求まるので，

$$h(z_{min}) = 89$$

となり，si(D3) を “89” に変更することにより秘密情報 “89” を埋め込む．つまり，D3 に関しては意味量を “−36(= 89 − 125)” する必要がある．3.6 節 (4-2) に従い，まず修飾子に関する処理を行う．D3 には 2 つの修飾子が含まれている．それらを削除した場合，意味量は “−22” となる．さらに意味量の減少が必要なので，結合子を削除する．D3 には結合子として $para$ が含まれているので，それを削除すると “−11” となり，先ほどの減少値と合わせると “−33” になる．このときの SD 式は，

$$[s((自分)plus(花子)),v(あげる/(未来)),o(結婚式)]$$

となる．さらに意味量の減少が必要なので，“LV_4 に属する単純ラベルを，より小さい意味素量の単純ラベルに置換” する．該当する単純ラベルは複数あるので，ランダムに選び，この場合「結婚式」が選択された．意味量が “−3” となるような単純ラベルは「三三九度」であり,「結婚式」を「三三九度」に置換することにより，目的の SD 式が得られた．

4.3 考　　察

新聞データの場合は，ほぼ意味的に問題はないと思

われるが，若干置換した単語に不自然さを感じる．たとえば「簡抜」したという単語はこの文脈中では不自然である．また，招待状の場合は，文の構造としては問題ないと思われるが，かなりの単語の置換は適当でないように思われる．提案法では，意味的な分類により階層化された概念を置換しているが，それでもなお不自然な置換が発生している．これは同じ意味のような単語であっても文脈によっては，不適当な場合が発生することを示している．単純ラベルのみに注目して，単純に置換するだけでは意味的な変化が生じてしまう．そこで，格辞書のようなものを用意し，主語や動詞あるいは文脈に対して共起できる単語群をさらに細かく分類した辞書を作成することが必要である．

提案法では，意味量を増加させる処理として単純ラベルの置換のみを用いた．しかし，修飾子，結合子，規定子を操作して意味量を増加させることは可能である．この操作は，概念を具体化させる処理として考えることができる．ただしこの場合，文章や文脈を考慮しなければ意味的な不自然さが発生してしまうことはいうまでもない．SD 式意味モデルでは，意味データと同等に知識データが利用できるので，このような問題に対しては容易に対処できる．

文献 1) では，意味的に矛盾を感じない程度の埋込みで，カバーテキスト 1,024 バイトあたり，約 32 ビットのデータ埋込みを行ったと報告している．一方提案法では，399 バイト（新聞記事）と 272 バイト（招待状）のテキストデータに対して，それぞれ 42 ビットのデータを埋め込んだ．提案法は，1 文に対して埋め込めるデータ量が一定であるという特徴を持つ．

本手法の特徴の 1 つは，生成される文の品質は，意味量と秘密データの差に依存し，埋込み量には依存しない点である．表 1，表 2 から分かるように，意味的に違和感のある文章では，意味量の増減値が大きいが，1 文への埋込み量はそれぞれ一律 7 ビットである．この性質は従来法とはまったく異なる．

第三者からのステゴテキストに対するアタックについて検討してみる．一般に，フォーマットを変更することにより意味情報は変化しないので，本手法はフォーマット変更に対しては完全にロバストである．今，提案法が公開されており，自然言語文と SD 式が 1 対 1 に変換でき，SD 式の意味素量も公開されているとする．そして，二者間の通信をチェックしている第三者がすべての通信を傍受し解析するとしよう．この場合，第三者はステゴ SD 式の意味量を計算することができる．秘密鍵は a（1 つの SD 式に埋め込む秘密データビット数）のみなので，第三者が秘密情報を復元できる確率は，$1/a$ となる．上記の条件の下，第三者がテキストを変更した場合，意味的な変化が生じれば SD 式は変化し，その結果ステゴ SD 式の意味量も変化するので，埋め込まれた情報は抽出できなくなる．SD 式意味モデルでは，意味素量の値の設定は利用者に任せられているので，概念，単語間の意味的な差異を保存しつつ，a の値をある程度大きくすることは可能であるが，現在の暗号システムに利用されているような 128 ビット程度で表現される大きな数値を a に対応させることは困難であると思われる．ステガノグラフィは，それだけでセキュリティを高める技術としてだけではなく，暗号技術と融合してセキュリティを高めるための手段としても利用できる．つまり，埋め込むデータを暗号化しておけば，提案手法でも，よりいっそうセキュリティを高めることは可能である．

なお，本研究の前提である “自然言語文と SD 式の機械的な相互変換” は，著者らの研究グループにおける重要課題[10),11)] であり，限定された話題を対象とすれば実現可能である．

5. おわりに

本稿では，SD 式意味モデルを利用したテキストベースのステガノグラフィを提案した．文の意味的な情報を抽象化具体化させることにより秘密情報を埋込み，ステゴテキストに意味的な矛盾を発生させないよう試みた．実験の結果，構文的な意味矛盾は感じられないが，文脈によっては置換された単語が不自然な文章もあった．これは，利用した同義語の中には文によっては置換すると不自然なる単語が存在した点と，文章全体を対象とした意味量操作をしていない点が原因であると考えられる．今後の課題は，知識データを利用して文脈に矛盾しない概念置換を実現することである．

謝辞 本研究に関連する予備実験的研究に携わった本学伊藤友和君（現在，住友重機械工業（株））に感謝する．

参考文献

1) 中川裕志，木村浩康，三瓶光司，松本 勉：辞書変換法に基づく日本語テキストへの情報ハイディング，情報処理学会論文誌，Vol.41, No.8, pp.2272–2280 (2000).
2) 中川裕志，三瓶光司，松本 勉，柏木健志，川口修司，牧野京子，村瀬一郎：意味保存型の情報ハイディング—日本語文書への適用，情報処理学会論文誌，Vol.42, No.9, pp.2339–2350 (2001).
3) 滝澤 修，牧野京子，村瀬一郎，松本 勉，中川裕志：改行位置の調整による自然言語テキスト

への情報ハイディング，SCIS2002 予稿集，Vol.2, pp.997–1002 (2002).
4) Chapman, M. and Davida, G.: Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text, *Information and Communications Security*, Han, Y., Okamoto, T. and Qing, S.(Eds.), pp.335–345, Springer (1997).
5) Shao, G., Nozaki, K., Kamata, S. and Kawaguchi, E.: SD-Forms as interlingua and a prototype of a conversational-text retrieving system, 人工知能学会誌，Vol.9, No.5, pp.684–693 (1994).
6) Wakiyama, M., Noda, H., Nozaki, K. and Kawaguchi, E.: Computation Algorithm of Semantic Difference Measure in the SD-Form Semantics Model, 情報処理学会論文誌，Vol.40, No.3, pp.1065–1079 (1990).
7) Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.: Indexing by latent semantic analysis, *Journal of American Society for Information Science*, Vol.41, No.6, pp.391–407 (1990).
8) Letsche, T.A. and Berry, M.W.: Large-scale information retrieval with latent semantic indexing, *Information Science*, Vol.100, No.1–4, pp.105–137 (1997).
9) 国立国語研究所：分類語彙表（フロッピー版）(1994).
10) 榊原正典：自然言語文からの SD 式生成，修士論文，九州工業大学大学院工学研究科 (2001).
11) 林 勝仁：SD 式からの自然言語文生成システムに関する研究，修士論文，九州工業大学大学院工学研究科 (2002).

(平成 14 年 11 月 27 日受付)
(平成 15 年 6 月 3 日採録)

新見 道治（正会員）

平成 4 年九州工業大学工学部電気・計算機工学コース卒業．平成 6 年同大学大学院博士前期課程修了．同年長崎総合科学大学助手．平成 8 年九州工業大学工学部助手，現在に至る．ステガノグラフィ，画像解析，自然言語理解の研究に従事．電子情報通信学会，IEEE 各会員．

峯脇さやか（学生会員）

平成 13 年九州工業大学工学部電気工学科卒業．平成 15 年同大学大学院工学研究科博士前期課程修了．この間，自然言語の意味処理に関する研究に従事．現在，同大学大学院情報工学研究科博士後期過程に在学中．

野田 秀樹

昭和 48 年九州大学工学部電子工学科卒業．昭和 50 年同大学大学院修士課程修了．同年第二精工舎（現セイコーインスツルメンツ）入社．昭和 53 年警察庁科学警察研究所入所．平成元年郵政省通信総合研究所入所．平成 7 年九州工業大学工学部助教授，現在に至る．パターン認識，話者認識，画像処理等の研究に従事．博士（工学）．日本音響学会会員．

河口 英二（正会員）

昭和 39 年九州大学工学部通信工学科卒業．昭和 44 年同大学大学院博士課程修了．同年九州産業大学講師．昭和 48 年九州大学工学部情報工学科助教授．昭和 54 年同大学院総合理工学科情報システム学専攻助教授．昭和 63 年九州工業大学工学部教授，現在に至る．昭和 59 年〜60 年テネシー大学訪問教授．工学博士．画像理解，情報圧縮，Steganography，自然言語理解等の研究に従事．電子情報通信学会，人工知能学会各会員．