

シーンを構成する文書中の連続文と画像の対応付けの研究

板橋 美子[†] 増永 良文[‡]

[†]お茶の水女子大学大学院人間文化研究科博士前期過程 〒112-8610 東京都文京区大塚 2-1-1

[‡]お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†] itabashiy@db.is.ocha.ac.jp, [‡] masunaga@is.ocha.ac.jp

あらまし Web の普及とマルチメディア化の進展と共に、日記や伝記と写真、物語と絵画や映画などのように、同じ内容を違うメディアで表現したデータや、片方からもう一方が派生したメディアデータが増えている。このような内容的に同じだが、別のメディア形式のデータを、文書と画像について、自動または半自動的に対応付ける方法を提案する。対応付けは、文書中の連続した文の集合がシーンを形成していると考え、連続文と画像についてベクトル空間モデルの類似度を用いておこなう。提案方法による対応付けを、源氏物語の現代語訳と源氏物語を素材にした画像について実験した結果と、メディアの違いや人物の取り扱いに由来する困難さについて述べる。

キーワード テキストDB, マルチメディア処理, 情報検索, クロスメディア

Research on Finding Correspondence between a Scene as a Series of Document Sentences and an Image

Yoshiko ITABASHI[†] and Yoshifumi MASUNAGA[‡]

[†] Graduate Division of Mathematics and Computer Science, Ochanomizu University 2-1-1 Ootsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

[‡] Faculty of Science, Ochanomizu University 2-1-1 Ootsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [†] itabashiy@db.is.ocha.ac.jp, [‡] masunaga@is.ocha.ac.jp

Abstract With the popularization of the Internet and the development of multimedia technology, various data are produced and disseminated in different media such as documents, images, photographs, pictures, and movies. Therefore, there are a variety of data in different media which represent the same contents. However, the method of correlating data in different media having the same contents has not yet been established. This paper proposes a method of correlating a document with an image, both of which express the similar contents. This method bases upon the observation that a series of sentences in a document could constitute a scene, and by utilizing the similarity measure of the vector space model as a criterion, we could correlate a scene with an image if their similarity is high. An experimental result of the above-mentioned method is shown taking “The Tale of Genji” and the photographs taken on the subject of this tale, as an example. It is shown that the method works fairly good as it is intended, however there still remain difficult problems for improvement.

Keyword Text Database, Multi-Media Processing, Information Retrieval, Cross-Media

1. まえがき

1.1. 背景

Web の普及により公開されアクセス可能なデータが増加すると同時に、マルチメディア技術の普及により、テキストや数値だけでなく、画像、音声、動画等様々な形態のデータが多用されている。これらデータには、同じ内容を異なるメディア形式で表現しているものも多い。

画像とテキストが同じコンテンツにある HTML 文書に対して行なわれる Google の画像検索は、画像とテキストの関連づけられた状態で参照できるため、画像の

意味による検索や結果の閲覧が可能になっている。しかし、同じ内容を表現したものでも、まったく別々に存在するため、同時に閲覧するには人手による対応付けが必要なものも多い。

例えば、日記や伝記、Blog や電子メールのような文書は、記述された事項についての写真やビデオデータが別に存在していることがある。また、文学作品についても、物語を素材とした絵画、映画などが存在している。日記や伝記と写真は、同じ出来事を複数のメディアで記録したものである。文学作品の挿絵、絵画、映画などは、文学によって人間が描く想像を異なるメ

ディアで表現している。これらは、メディア形式は異なるが、同じ内容が派生しているか、一方から他方が派生している。

画像等の文字検索不能なデータを検索する方法としては、単語を用いてメタデータ空間に写像し、検索する方法が従来から用いられている[1][2]。これらの手法を援用して、同じ内容を表現した多量の文書の一部と対応する画像を同時に閲覧することはできないだろうか。

1.2. 関連研究

1.2.1. テキストの内容理解

同じ内容を表現した画像とテキストを関連付けるには、テキストの内容を理解して定式化できればもっとも正確である。この方法については、記号論のアプローチを用いた[3]や event calculus formulas を用いて深層解釈を試みた[4]、汎化されたイベント系列で索引付けを試みた[5]などがある。いずれも、童話や説話を題材としていて多量の文書进行处理することはできない。また、マニュアルによる物語の記述が必要になる。

多量の文書进行处理するには、実用的なレベルにある文書検索のような表層的な処理を使用するのが現実的な手段である。

1.2.2. テキストのパスページ検索

多量の文書から、画像に対応する部分を取り出すのは、検索語によるパスページ検索に似ている。パスページ検索には、章や節のような形式段落によるもの[6]や検索要求に対して語彙の連鎖によりパスページを検索する[7]のような方法がある。

画像に対してつけられた注釈の語を検索要求と考えると、画像に対してパスページを検索するのは、似ているといえるが、異なる側面も多い。注釈自体も次の2種類のものがある。

- 画像の製作者か編集者がつけたタイトルやコメント
- 画像上に存在するものを判別してつけた注釈

例えば、テキストから画像を製作するときには、製作者の意図とデザインというプロセスが加わっている。タイトルには製作者の意図が反映され、それはテキストの内容に近いが、テキストの内容を示すためにつけられた文ではない。画像上に存在するものは、製作者のデザインの結果であり、これもテキストの内容を直接示すものとは限らない。通常のパスページ検索と画像の注釈によるテキスト検索には、なんらかの違いがあると考えられる。

また、画像で表現されるパスページを考えると、そのなかに会話文が含まれることはある。先行する研究では問題とされていないが、シーンに会話などの文を含む文書を取り扱う時には、会話文でパスページが分

割されないようにする配慮が必要になる。

1.2.3. 画像検索

大量の画像を単語による検索をおこなうために、画像の自動的な注釈付けが盛んに研究されている[8][9][10]。ここでの注釈は、画像上に存在するものを判別してつけた注釈であり、学習により注釈付けしている。

これらの方法は、例えば、1枚の写真に「1本の木と1人の人間」が写っていることを判別するものである。物語や伝記のようなテキストと画像を対応付けるには不十分である。より表現する対象に依存した注釈付けが必要になる。ただし、タイトルなどの他の注釈を補完するためになら使用できる可能性がある。

本研究に最も近い研究に、英文で300語程度程度のテキストに近い画像を選択する[11]がある。[11]は、テキストから単語と特有の単語を抽出し、その語と画像の注釈の語の類似度を算出してランク付けしている。画像の注釈は製作者がつけたタイトルや文を利用している。しかし、複数のエピソードからなる通常の物語を対象としてはいない。

1.3. 本稿のアプローチ

本稿では、同じ内容を扱っていることが既知である文書群と画像群を対象に、文書群の一部のテキストと画像を対応付けることを考える。テキストの大きさは利用者が画像と同時に閲覧できる程度であることがのぞましい。また、利用者は、扱っている内容に関心を持っていて、より詳細な知識を望んでいる人を対象として想定している。

文書は、事実の記述が中心となるテキストだけでなく、会話や心情の描写などを含む物語形式のテキストを対象とする。物語の定義は、「現実であれ虚構であれ、時間連鎖のかたちでの状況・事象の再現表象」（ジェラルド・プリンス 物語論の位相）とする。どのような記憶形式でも、文書は基本的に順序をもつ文の集合である。特定時間・特定空間について記述される状況・事象ものを場面(シーン: scene)とし、それを表現する画像があると仮定すると、シーンは文書中の連続文に対応していると考えられる。シーンは、会話や独白などを含むこともある。

画像については、画像の特徴量ではなく、注釈をつけて、テキストとの関連付けを試みる。注釈は、画像製作者か編集者がつけたタイトルと、画像上に存在するものを判別してつけた注釈を併用する。後者の注釈については、対話的に手動でつける。

テキスト中に現れる事物を現す単語はテキストの内容を表現し、シーンを構成する連続文は、それを特徴付ける単語の集合で表現されると仮定する。この仮定に基づき、ベクトル空間モデルを用いて、連続する

文と画像の注釈の対応付けをおこなうベクトル空間モデルには、画像から連続文の対応付けと連続文と画像の対応付けを同じに扱うというメリットがあり、文書中の指定の位置にふさわしい画像、あるいは指定した画像にふさわしい連続文という形での検索が可能になる。また、連続する文についてのベクトルの算出が単純に加算で行えるため、取り扱いしやすいという長所がある。

テキストと画像が同じ内容を扱っているため、これらは同一の特有の単語の集合で特徴付けられると考えられる。小さなテキストを対象にすると、含まれる語数が少なくなるため、1語の重要性が増す。類義語を処理できるシソーラスを使用することによって、対応付けの精度を上げることができることは、容易に予想される。[11]は画像の注釈に対して、WordNetを使用し、おおまかな意味的な類似度を評価している。[7]はパッセージに対し、角川類語新辞典を使用している。本稿は同じ内容をあつかう画像と文書を対象にしているため、可能なら内容に関係するシソーラスを使用する。

1.4. 人物の取り扱い

文書と画像の対応付けをするのに、有用な情報は人物である。しかし、人物の情報は単純にベクトル空間モデルを適用できない。人物を特定するのは、固有名詞であるが、これを索引語とするには次の問題がある。

- (1). 特定の人物を表す固有名詞は複数あり、シーンにより異なる。
例) ヒラリー、クリントン夫人、ファーストレディ、クリントン上院議員
- (2). 文書においては、人物は代名詞で言及されることが多く、必ずしも固有名詞で存在が示されない。
- (3). 重要な人物については、文書中では存在ではなく状態が重要である。
- (4). 画像では、例えば、「高校生のチェルシー」と「赤ちゃんのチェルシー」では、まったく異なる。「チェルシー」を知っている人は、固有名詞で注釈がつけられるが、知らない人にとっては、「高校生」「赤ちゃん」でしかない。

このような問題を回避するためには、シソーラス以上の、分野や内容に依存する知識の記述が必要となる。このような未解決の大きな問題があるため、本稿では特定人物については扱わず、他の事物のみでシーンの特定を試みる。

2. 文書中の連続文と画像の対応付け方式

2.1. ベクトル空間モデル

ベクトル空間モデル(vector space model)は Salton らにより提案された技法である[12]。文書から内容に関

連する索引語を抽出し、索引語の重みを要素とするベクトルで文書を表し、文書ベクトル(document vector)と呼ぶ。

$$\mathbf{d}_j = \begin{bmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{mj} \end{bmatrix} \quad (1)$$

ここに文書ベクトル \mathbf{d}_j の要素 d_{ij} は索引語 w_i の文書 D_j における重みである。

索引語の重みは、局所的な重み付け (l_{ij}) と大域的な重み付け (g_i) の積で表現される。

代表的な局所的な重み付けは、索引語の出現頻度 (term frequency; TF) である。他には 2 進重み (binary weight) が用いられることがある。

大域的な重み付けとしては、文書頻度の逆数 (inverse document frequency; IDF) がよく知られている。

$$d_{ij} = l_{ij} \cdot g_i \quad d_{ij}: \text{索引の重み付け} \quad (2)$$

$$l_{ij} = f_{ij} \quad l_{ij}: \text{局所的重み付け (TF)} \quad (3)$$

$$g_i = \log \frac{n}{n_i} \quad g_i: \text{大域的重み付け (IDF)} \quad (4)$$

n : 総文書数

n_i : 索引語 w_i が含まれる文書数

情報検索では、検索質問文を文書ベクトルとして表現し、検索質問文と検索対象の文書ベクトルの類似度を計算する。類似度の定義としては次の 2 つがよく用いられる。

コサイン尺度

$$\text{sim}_n(\mathbf{d}_i, \mathbf{d}_j) = \cos(\mathbf{d}_i, \mathbf{d}_j) = \frac{\mathbf{d}_i \cdot \mathbf{d}_j}{\|\mathbf{d}_i\| \cdot \|\mathbf{d}_j\|} \quad (5)$$

内積

$$\text{sim}_p(\mathbf{d}_i \cdot \mathbf{d}_j) = \mathbf{d}_i \cdot \mathbf{d}_j \quad (6)$$

2.2. ベクトル空間モデルを用いた対応付け

2.2.1. 索引語の選定

ベクトル空間モデルでは、索引語の集合がベクトルの次元となる。索引語は、自由索引語 (free index term) でつける方法と、統制索引語 (controlled index term) でつける方法がある。本稿では、取り扱う対象が、同一分野のものに限られ、語彙を特定しやすいため、統制索引語を用いる。具体的には、特定の索引語の集合で文書ベクトルを計算し、画像の注釈 (annotation) 付けをおこなう。

索引語の多義語、類義語を処理するためにシソーラスを使用する。文献[13]では、索引語を類義語の集合

に置き換え、潜在的意味索引付け(Latent Semantic Indexing, LSI)によって次元を圧縮する方法を用いているが、本稿では単純に、類義の索引語を記号に縮退させる方法(文献[14])を使用する。

2.2.2. 会話文に起因する問題

ベクトル空間モデルを使用して、個々の文を一文書、連続する文を別文書として考えて、一般的なコサイン尺度で1つの画像と最も類似した連続文を算出すると、会話文のような語数の少ない文の類似度が不当に高く評価される。

図1に例によって説明する。この画像には、「阿弥陀仏」、「蓮」という注釈がつけられている。コサイン尺度で評価すると、この画像の場面を詳細に記述した文1より、「阿弥陀仏、阿弥陀仏」を唱えている他の場面の文2の方が類似度が高いと判定される。これは、文1は、画像に注釈されない詳細な事物の描写を含むため、文のベクトルの長さが大きくなり、コサイン尺度が小さくなるためである。

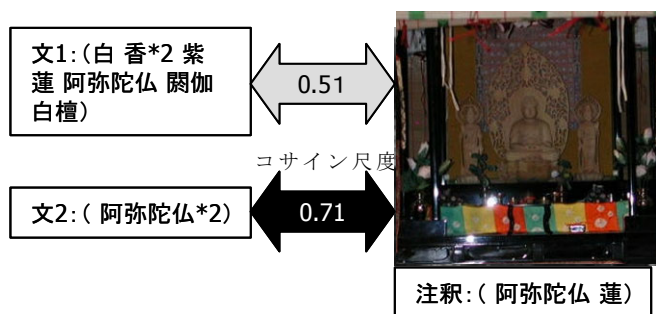


図 1. コサイン尺度による類似度の評価

Figure1. The Evaluation by Cosine

会話や心中描写は、索引語が現れてもわずかである。このような文は、物語には頻繁に出現し、シーンをなすシーンをなす連続文に含まれると考えられるが、そのみでシーンを構成することはない。

物語のような文書で特定のシーンを探すときは、コサイン尺度のみで類似度を評価するのは適切ではないと考えられる。

2.2.3. 連続文と画像の対応付け方式

コサイン尺度は、類似度を比較するとき、文書の長さの要素を正規化するために使用される。もし、文書の長さがほぼ均一なら正規化の必要性は薄れる。

そこで、本稿では、2段階で類似度を評価する。含まれる文数の同じベクトル同士の比較では、一致する語数を反映する内積を用いて類似度を評価し、文中の位置を特定する。位置の同じベクトル同士の比較では、語数の影響を小さくするコサイン尺度を用いて類似度を評価し、適切な範囲の連続文を検出する。

本稿の類似度の評価方式を、図2に単純化した例によって説明する。例の画像は、「関伽を供える若い尼姿の女房」のもので、(関伽 尼姿)が画像のベクトルであるとする。文毎のベクトルは図2のなかの表に示されているとする。

処理は、次のような順序でおこなう。

- (1). 画像のベクトルと連続文のベクトルを、集計文数ごとに、内積で評価して、解の候補を選択していく。この例では、6文まで評価している。
- (2). 中心となる文の位置毎にコサイン尺度で評価して、解の候補を選択する。
- (3). 最後に各位置の解をコサイン尺度でランキングして解を表示する。

この例では、文番号 9217 は登場人物の発言でその前の文は発言の背景の説明である。この方式での最も類似度が高い文は、文番号 9216 を中心とする5文となり、この場合は正解である。



関伽を供える
尼姿の若い女房

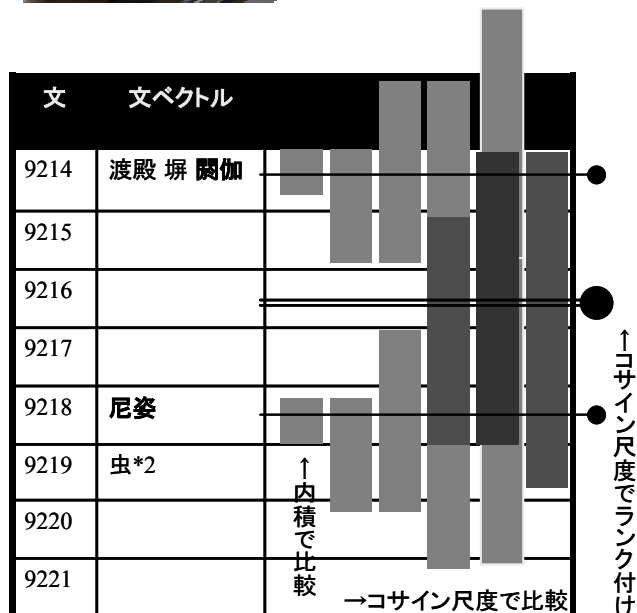


図 2. 連続文と画像の対応付け方式

Figure2. Method of Finding Correspondence

コサイン尺度のみで選択すると、文番号 9218 の1文の、「とお諫めになって、十何人かだけが尼姿になってお付きしている。」という文が高い順位で選択されてしまう。集計文数毎に、内積で評価してフィルタをか

けることによって、会話文のような文が単独で選択されず、会話を含むシーンを選択することができる。

集計行数を決める範囲特定は、コサイン尺度で比較する。1つの画像の注釈の索引語数は、文書全体に含まれる索引語数よりかなり小さい。同一の文を中心とする集計文数が異なる連続文については、集計文数が複数のシーンを含む範囲以上に増加すると、連続文のベクトルの長さが、連続文のベクトルと画像ベクトルの内積以上に早く増加するため、収束していくと予想される。

3. 画像と連続文の対応付けの実験

類似度の判定方式の検証、および人物の情報なしでどの程度の対応付けできるかを評価するため、同じ内容を表現し事前に対応づけを確認できるデータを用いて実験をおこなう。

比較のために、類似度の評価をすべてコサイン尺度でおこなった場合の結果も算出する。

実験は、画像からの対応する連続文の抽出を中心にこなう。文からの画像の対応付けについては、実験に使用する画像の数が少なく偏っているため、参考実験として一部実施する。

3.1. 実験環境

実験には源氏物語の現代語訳を使用した。選択した理由は、源氏物語は日本で最も長い小説の1つであり、多数のエピソードからなり、実験対象として十分な長さを持っていること、小説の特定の場面を取り上げた画像や分野に依存するシソーラスが得られることである。実験に使用するデータを以下にあげる。

- 対象文書
 - 源氏物語 現代語訳
 - ◇ Web上で公開され、オープンソース化されている渋谷栄一訳[15]
 - ◇ テキストの構造も渋谷による構造を使用
 - ◇ 宮脇文経氏によるCSV版を利用[16]
- 対象画像
 - 風俗博物館の展示写真を利用（鈴虫の帖）[17][18]
- シソーラス
 - 「源氏物語図典」秋山 虔，小町谷照彦編 小学館[19]

文書は、54帖→230章→1244段→14436行のように構造化され、最小単位の行がほぼ文に相当する。実験では、行を更に文に分け直さず、そのまま文ベクトルの算出に使用した。なお、テキスト中に旧字と新字が混在していたがすべて新字に統一した。

画像は「鈴虫」帖の再現模型の写真 78 枚を用いたが、文書中に直接描かれているシーンに対応している

のは 30 枚で、他の 48 枚は模型自体の説明や、文書に描かれていない舞台裏の場面になっている。表 1に、実験に使用した画像と源氏物語の対応を示す。

シソーラスについては、源氏物語の時代の風俗・習慣に関する図版を収録した、「源氏物語図典」の巻末の索引語 1528 語を索引語とし、図録の見出し 987 に集約する処理を実施した。

表 1. 画像と鈴虫帖の章・段
Table 1. Images and the Targeted Articles

章	段	行数	場面画像	他の画像
一章	持仏開眼供養の準備	7	6	7
	源氏と女三の宮、和歌を詠み交わす	17	8	14
	持仏開眼供養執り行われる	4	3	18
	三条宮邸を整備	4		
二章	女三の宮の前栽に虫を放つ	11	7	6
	八月十五夜、秋の虫の論	14		
	六条院の鈴虫の宴	8	6	3
	冷泉院より招請の和歌	10		
三章	冷泉院の月の宴	5		
	秋好中宮、出家を思う	9		
	母御息所の罪を思う	7		
	秋好中宮の仏道生活	3		
合計		99	30	48

3.2. 実験システム構成

実験システムを図 3 に示す。

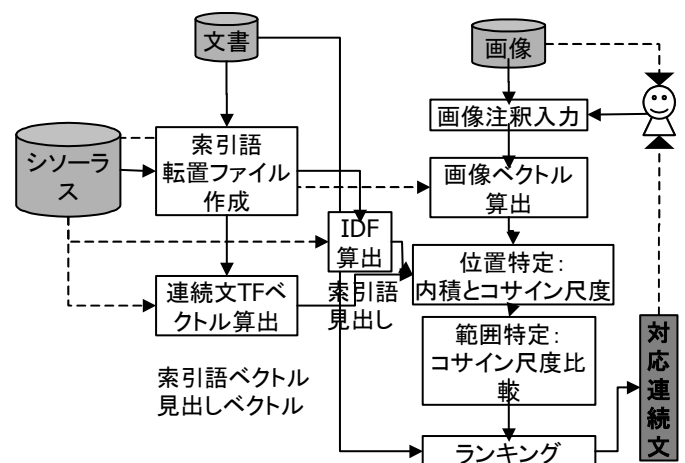


図 3. 実験システム
Figure3. Experiment System

シソーラスの索引語情報に基づいて、索引語転置フ

ファイル作成部が文書から転置ファイルを作成する。

転置ファイルを元に、IDF 算出部と連続文 TF ベクトル算出部が、IDF と文の TF ベクトル、連続文を集計した TF ベクトルを算出する。IDF と TF ベクトルは、索引語に基づくものと、シソーラスを参照して見出しに集約したものと 2 種算出する。

利用者は、画像を参照しシソーラスによるチェックをしながら注釈をつける。画像ベクトル算出部が、注釈から画像ベクトルを索引語に基づくものと見出しに集約したものと 2 種算出する。

位置特定部は、集計行数毎に、画像ベクトルと連続文のベクトルを内積とコサイン尺度で評価する。それぞれについて、類似度が高いものを解の候補とする。

範囲特定部は、中心位置が同じ解の候補をコサイン尺度で比較し、類似度が最も高いものを解とする。

ランキング部は、中心位置の異なる解をコサイン尺度でランクをつけ、文書の該当連続文と共に利用者に表示する。

3.3. 画像の注釈付け

画像の注釈は、画像タイトルと画像上で確認できる事物の 2 種つける。実験では、タイトルについては Web で取得したものを利用した。画像上で確認できる事物については、他の画像のタイトルで特定できるものと「源氏物語図典」で確認できるものをつけた。いずれも、索引語に存在しない場合は、索引語の範囲で言い換えをおこなった。

注釈付けは、源氏物語について一般的な知識を持っている 1 人がつけた 1 種類でおこなった。

表 2. 画像の注釈数
Table 2. Annotations with Each Image

注釈数	画像数
1	7
2	30
3	24
4	8
5	5
6	2
7	1
8	1
合計	78

3.4. 実験の範囲と評価

実験では、集計の行数を 20 行までとする。画像に対応する連続文の検索は、範囲特定類似度（コサイン尺度）でランク付けし、10 件まで選択し、その中の正解で適合率(precision)、再現率(recall)を評価する。

文書ベクトルの重み付けは、索引語によるベクトル、見出しによるベクトルそれぞれについて、TF・IDF に

よる重み付けをおこなう。

正解は、場面画像に対する解のうち、中心文が段に一致したものとする。但し、集計行数が対象となる段の行数より大きいことと、「鈴虫」巻の 1 章の 1～3 段は時間と場所が一致する連続した場面のため、連続行の中心がずれていても、人からみると納得できる解である場合もあるので、必要に応じて章の一致についても算出して評価する。

3.5. 画像から対応する連続文の検索の実験結果

3.5.1. 結果

類似度の比較の方式を検証するために、位置特定を内積で比較した場合と、コサイン尺度で比較した場合の結果の比較をおこなう。

比較は、10 位までの検索結果の平均適合率を用いて、シーンに対応する 30 の画像について平均した結果を表 3 に示す。ここでは、索引語から直接文書ベクトルを計算し、類似度を算出した場合を示す。

表 3. 位置特定方法による平均適合率の違い
Table 3. Difference of Average Precision for Position Identify Method

位置特定	段の一致	章の一致
内積	44.4%	50.0%
コサイン尺度	37.1%	39.5%

コサイン尺度のみでは、まったく正解が検出されない 4 件の画像で、内積を使用したとき正解が検出されている。文書上の位置を決めるための位置特定類似度の計算では内積を用いる方が有効であると考えられる。

また、人物の情報なしで、14436 行からなる文書を対象に 7 行～17 行のシーンを半分近く検出できている。

シーンに対応する 30 の画像ごとの平均適合率を図 4 に示す。

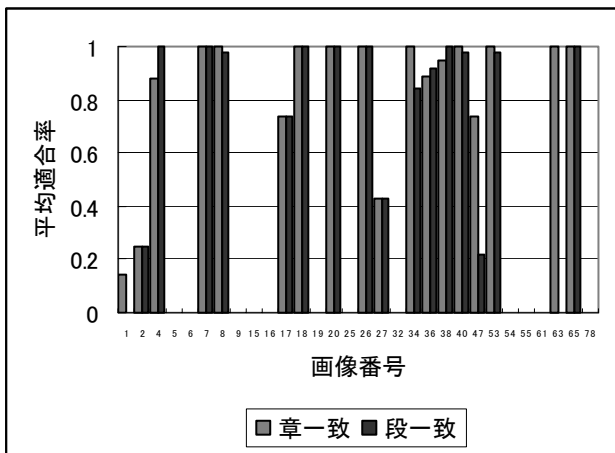


図 4. 画像ごとの平均適合率
Figure 4. Average Precision per Image

図 4で見るように、平均適合率の分布は 1.0 と 0.0 に偏って分布している。画像の製作者がシーンの特徴を捉えているため、適合率は 1.0 のものが多いと考えられる。

平均適合率が 0 になった画像については、原因を後に示す。

3.5.2. シソーラスの使用

ベクトル算出を索引語そのままでおこなう場合と、「源氏物語図典」[19]の見出しで類義語を集約した場合の結果を比較する。

表 4. シソーラスによる平均適合率の違い
Table 4. Difference of Average Precision by Using Thesaurus

ベクトル	段の一致	章の一致
索引語	44.4%	50.0%
見出し	36.9%	40.8%

見出しによって集約したほうが、注釈の不正確さを避け、意味的に正しく処理できると予想したが、結果は逆になった。原因は、「源氏物語図典」[19]の索引語の意味関係が、例えば「月の宴」が「宴」に縮退され、他の宴との一致が高くなるなど、視覚的に重要な要素が縮退により失われたためである。

3.5.3. 適合率が低い画像

10 位までの結果で、章レベルでも 1 件も一致しなかった 10 の画像について、原因を調査した。

不一致になっている画像は、人物が主になっていて有効な注釈付けができず、建物や道具について注釈付けしたものである。特に、表 5 にしめすように、「柱」や「廂」、「御簾」などは、注釈に使用されると適合率を下げた。

これは、画像上で目立つ建物や道具を注釈としていれると、文書上はそれらの事物について特別なことのあるシーンが選択されるためである。例えば、女三宮が数珠を手にして柱に寄りかかっている画像は、本稿の実験では、「女三宮」という注釈ができない。そのため、「柱」という索引語が多く使われる「真木柱」帖が検索される結果になっている。「鈴虫」帖では、「御帳台」と「廂」については、仏教関係の索引語と共起すると正解になるが、画像に仏教関係と分かるものがないと不正解になっている。

表 5. 建物等の索引語と平均適合率

Table 5. Index Terms for Building Structure and Average Precision

索引語	画像数	段一致	章一致	補足
御帳台	4	54%	65%	仏教関係と共起すると正解

索引語	画像数	段一致	章一致	補足
廂	9	36%	49%	「行道」と共起すると正解
御簾	4	12%	25%	
柱	4	0%	0%	

3.6. 文書から対応画像の検索の実験結果

「鈴虫」帖の 1 章について、対応する画像の検出をおこなう。但し、画像の対象が鈴虫巻の場面に直接対応するものか、間接的に対応するものしかないため、参考実験として実施する。

対応する画像のない段も含まれるため、集計行と画像の組み合わせのうち、類似度の高い組み合わせだけを解候補とし、各行を中心とする解の集合の中で最も一致度の高い画像を解とする。なお、ここでの評価は索引語によるベクトルでおこなう。

どの類似度の評価方法を用いても、対応する画像のない 4 段と画像は存在するが行数のすくない 3 段については、対応する画像は検出されなかった。3 段については、2 段の文の集計行数が大きい解に内容が含まれている。

表 6. 鈴虫帖の 12 行と対応させられた画像数

Table 6. Number of Images corresponding to Targeted 12 Lines (Suzumushi)

位置特定類似度	内積	コサイン尺度
画像の種類	3	5
段の一致	4	8
章の一致	12	12
場面以外の画像	1	2

どの類似度評価方法でも、対応画像が検出できた 1 段と 2 段の 12 行についての検出画像を比較する。この 12 行に対しては、全て正しい章に対応する画像が検出されている。

3.7. 性能

78 件の画像について対応付けを実施し、1 件あたりの処理時間を表 7 に示す。測定条件は次のとおりである。

- CPU : Pentium4 2.6GHz, 2.59GHz
- Memory : 1.00GByte
- OS : Windows XP Professional
- DataBase : Microsoft Access 2000

表 7. 最大集計行数と性能

Table 7. Performance and Maximum Count of Lines

最大集計行数	5行	10行	15行	20行
検索時間	1.5秒	4.2秒	8.6秒	11.0秒

4. まとめと今後の課題

ベクトル空間モデルを用いて、文書中の連続文と画像を対応付けた。本稿の実験は限られたものであったが、物語中の登場人物という重要な要素を除いておこなったことを考慮すると、画像から連続文への対応で、平均適合率で、段レベルで40%以上、章レベルで50%以上の正解を検出したのは、一定の評価ができる。

今後の課題には、次のようなものがある。

(1). 他の分野の文書への適用

今回の実験では、源氏物語を使用した。歴史的な作品であり、索引語の取り扱いに特別な点があった。今後、他の文書に適用する際には、文書から索引語を生成し、分野に固有の語を追加するプロセスを検討する必要がある。

(2). 人物の取り扱い

(1)とも関連するが、テキスト中の人物と画像上の人物をどのように対応付けるか、そのための知識をどのように与えるかを検討する必要がある。

(3). 索引語の分野別やメディア別の重み付け

視覚的に目立つ「柱」は文章中では「柱」に関する事件がなければ言及されず、画像の注釈で使用されると対応付けが失敗する。文書側では、そのような語としては、本稿の実験では、画像上に注釈できない「香」、「虫の音」などがある。これは、単語の分野によっては、画像ベクトルと文書ベクトルで重み付けを変える必要があるかもしれないことを示唆している。今後の検討が必要である。

本稿の実験対象である「源氏物語」についていえば、今回使用したソーラスの「源氏物語図典」[19]は、古典読者のための参考書であり、現代語にも存在する「扇」や「供物」のような索引語が除かれていて、注釈に使用できる索引語が制限されてしまった。意味的な類縁についても、実験結果では有効な結果が得られなかった。索引語の候補については、他のソーラスを使用するか、あるいは文書側からなんらかの形で自動生成するなどの手段をとるべきであると思われる。

文 献

[1] 清木康, 金子昌史, 北川高嗣, "意味の数学モ

デルによる画像データベース探索方式とその学習機構", 電子情報通信学会論文誌, Vol. J79-D-II, No. 4, pp. 509-519, Apr. 1996.

- [2] Kiyoki, Y., Kitagawa, T. and Hayama, T.: "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning," ACM SIGMOD Record, vol. 23, no. 4, pp. 34-41, 1994.
- [3] 増永良文, "オブジェクト指向マルチメディアデータベース : 特にテキストデータベースにおける内容検索を考える", 研究報告「データベースシステム」, No.1991-DBS-084, pp239-248, July.1991.
- [4] Erik T. Mueller, "Story understanding through multi-representation model construction", In Graeme Hirst & Sergei Nirenburg (Eds.), Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop (pp. 46-53), 2003
- [5] M. Haraguchi, S. Nakano and M. Yoshioka, "Discovery of Maximal Analogies between Stories", Springer LNAI 2534 (Proc. 5th Int'l. Conf. Discovery Science), 324-331, Oct.2002
- [6] G. Salton, J. Allan and C. Buckley: "Approaches to Passage Retrieval in Full Text Information Systems", Proc. of 16th ACM SIGIR, pp.49-56, 1993.
- [7] 望月源, 岩山真, 奥村学: "語彙的連鎖に基づくパッセージ検索", 自然言語処理, vol.6, No.3, pp.101-126, 1999.
- [8] 北研二, 任福継, 福見稔, "クロス・メディア・コンテンツ検索に関する研究", 放送文化基金「研究報告」平成13年度助成・援助分
- [9] J. Jeon, V. Lavrenko, R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models", Annual ACM Conference on Research and Development in Information Retrieval Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp.119-126, July.2003.
- [10] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, M. Jordan, "Matching Words and Pictures", Journal of Machine Learning Research, Special Issue on Machine Learning Methods for Text and Images, vol. 3, pp. 1107-1135, 2003.
- [11] D. Joshi, J.Z. Wang and J. Li, "The Story Picturing Engine: Finding Elite Images to Illustrate a Story Using Mutual Reinforcement", Proc. 6th International Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia, pp. 119-126, New York, NY, ACM, October 2004.
- [12] 北 研二, 津田和彦, 獅子堀正幹, 情報検索アルゴリズム, 共立出版株式会社, 2002
- [13] 大内浩仁, 三浦孝夫, 塩谷勇, 多義性を考慮した文書検索, DEWS2003 1-P-01, 2003
- [14] 徳永健伸, 言語と計算 5 情報検索と言語処理, 東京大学出版会, 1999
- [15] 渋谷 栄一 : 源氏物語の世界, <http://www.sainet.or.jp/~eshibuya/>, 1996-2004
- [16] 宮脇文経 : 源氏物語の世界 再編集版 第二サイト, <http://genji.nce.butobi.net/>
- [17] 風俗博物館～よみがえる源氏物語の世界～, <http://www.iz2.or.jp/index.htm>
- [18] 風俗博物館を10倍楽しむ!, <http://evagenji.hp.infoseek.co.jp/costume.htm>
- [19] 秋山 虔, 小町谷照彦, 源氏物語図典, 小学館, 1997