

# 地物のカテゴリ階層と配置に基づく略地図自動生成に関する検討 A Study on Automatic Generation of Abbreviated Maps Based on Category Hierarchy and Arrangement of Geographic Features

芝原 拓人\*

Takuto Shibahra

福地 湧†

Yu Fukuchi

牛尼 剛聡‡

Taketoshi Ushiana

角谷 和俊\*

Kazutoshi Sumiya

## 1. はじめに

ユーザが地理情報を取得する情報源の一つとして略地図がある。例えば、観光地の案内図、災害避難経路図、カフェマップなどがその典型である。これらの略地図は、ユーザが特定の場所を見つけやすくするために、重要なランドマークや道路を強調して描かれることが多く、略地図を作成する目的によって表示する地物は異なる。しかし、手動で略地図を新たに作成する際に、作成する目的に適した表示する地物を選択するためには多くの時間と労力がかかる。

そこで、ユーザの目的と同じ目的で生成された略地図における地物のカテゴリ階層と配置を学習することによって新たに略地図を自動生成することを目的とし、本稿では略地図の自動生成に向けてユーザの入力から略地図作成の目的を抽出し、ユーザの目的に適した略地図を生成する際に参考となる地図画像を検索する手法を検討する。

本研究では、略地図の自動生成に向けて、テキストとの関連度の高い地図画像を提示するために、本手法では、図1で示すように、略地図と福地ら[1]が提案した地理オブジェクトのカテゴリを色で表現した「色付きカテゴリ地図画像 (Colored-Category Map 画像: CCM 画像)」をもとに、地物の強調を行った地図画像である「略地図 Colored-Category Map 画像: 略地図 CCM 画像)」を使用する。これに加えて略地図とそれが掲載されている Web サイト内の略地図に関するテキストをペアにして Encoder によってベクトル化し、機械学習モデルを訓練する。この機械学習モデルを活用して略地図を生成する例としては、「金沢市で美術館巡りがしたい」と考えるユーザがいたとき、「美術館巡り」などの略地図作成の目的と、「金沢市」という略地図作成対象地域名をテキストデータとして機械学習モデルに入力する。そして、入力したテキストデータと、データベース内に格納された全ての略地図 CCM 画像との類似度を計算し、その結果をもとに類似度の高い略地図 CCM 画像を複数枚検索す

る。なお、検索された略地図 CCM 画像に基づいて略地図を自動生成する手法については今後の課題である。

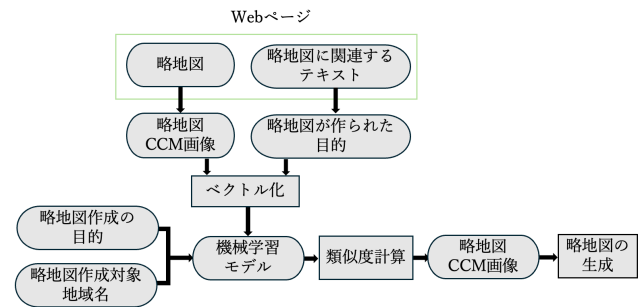


図1: 略地図生成の手順

## 2. テキストと略地図 CCM 画像とのマッチング

### 2.1 CLIP とは

CLIP[2] は OpenAI によって公開された言語と画像のマルチモーダルモデルである。学習時には言語データを Transformer により、画像データを Vision Transformer により埋め込み表現に変換し、正しいペアに関しては画像とテキストのペアの埋め込み表現のコサイン類似度が高くなるように、誤ったペアについては類似度が低くなるように Image Encoder と Text Encoder のパラメータを調整しながら学習を行う。この手法によって、ゼロショット転移において高い有効性を示している。

Radford ら [3] は、CLIP を事前学習モデルとして利用し、衛星画像に対して擬似ラベルを生成した。これによってゼロショット分類タスクにおいて高い精度を達成している。

一方、CLIP を略地図の自動生成に利用する研究は行われていない。そこで、本研究では CLIP を略地図に拡張することで、略地図 CCM 画像とテキストの類似度を予測する。

### 2.2 CLIP の略地図への拡張

CLIP は、画像とテキストを学習することによって、テキストと類似度の高い画像の提示などを行うことができるが、既存の CLIP モデルはインターネット上で公開さ

\*関西学院大学総合政策学部 School of Policy Studies, Kwansei Gakuin University

†関西学院大学総合政策学研究所 School of Policy Studies, Kwansei Gakuin University

‡九州大学大学院芸術工学研究院 Faculty of Design, Kyusyu University



画像：©国立西洋美術館

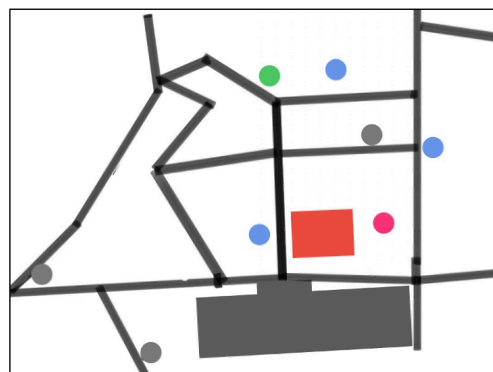


図 2: 略地図から略地図 CCM 画像への変換(国立西洋美術館)

れている多種多様な画像とそれに対応するテキストのペアで訓練されているため、略地図 CCM 画像とユーザが入力したテキストとの類似度を十分に予測することができない。

そこで、本手法においては、略地図をもとにして地物のカテゴリ階層と配置を強調した略地図 CCM 画像と、略地図が掲載されている Web サイト内の略地図に関するテキストを使用して CLIP と同様に、画像とテキストをペアにした事前学習を行うことにより、略地図 CCM 画像とユーザが入力したテキストの類似度の予測精度を向上させる。

### 2.3 学習モデルの作成

本節では、略地図 CCM 画像とテキストの類似度を予測するための学習モデルの作成に用いるテキストデータと画像データについて説明する。

まず、略地図と共にその略地図を説明するテキストが掲載されている Web ページから、テキストデータとして Web ページ上にある略地図を説明する文や、略地図のタイトルとして表示されているテキストを取得する。このとき、同様にして画像データとして略地図画像を取得する。ここで取得したテキストデータは、そのままモデルの作成に用いる。一方、画像データについては、図 2 に示すように、略地図画像の特徴を表現するために、地理オブジェクトのカテゴリを色で表現し、地物のカテゴリ階層と配置だけを画像上に表示して強調した略地図 CCM 画像を用いる。

次に、図 3 に示すように Text Encoder により、Web ページから取得したテキストデータをベクトルに変換し、Image Encoder により、作成した略地図 CCM 画像をベクトルに変換する。変換されたテキストベクトルと画像ベクトルは、共通の埋め込み空間にマッピングされ、テキストと画像のペアの類似度を計算することが可能とな

る。ここで、画像とテキストの組み合わせが正しいペア（例：国立西洋美術館周辺の略地図 CCM 画像と、美術館へのアクセス）の類似度を最大化し、誤ったペアの類似度を最小化するために、コントラスト学習を適用する。これにより、CLIP のモデルはテキストと画像の関連性を高精度で予測することが可能となる。

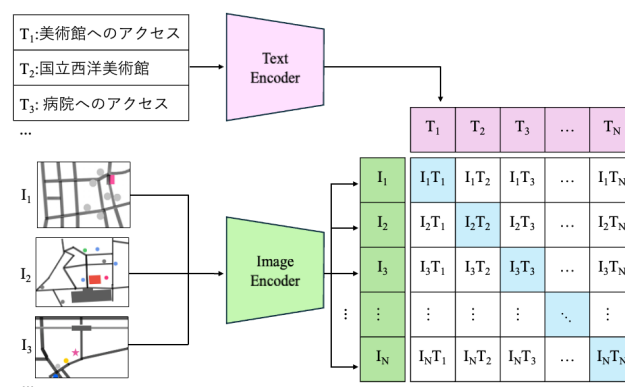


図 3: テキストと略地図 CCM 画像による学習モデルの作成

### 2.4 テキストと略地図 CCM 画像のマッチング手法

本節では、テキストと略地図 CCM 画像のマッチング手法について述べる。ここでのテキストは、対象とする地域で特定の目的に沿った略地図を生成したいユーザによる入力を指し、略地図 CCM 画像はユーザが求める略地図を作成するための参考となるものである。例えば、図 4 で示すように、特定の地域で「美術館巡り」がしたいユーザがいたとき、このユーザが入力したテキストと、略地図 CCM 画像のマッチングを行うために、入力

を Text Encoder によりベクトルに変換し、学習モデルにより、ユーザが入力したテキストと、データベース内に格納された全ての略地図 CCM 画像との類似度を計算する。このとき、入力テキストと類似度が高い略地図 CCM 画像を出力する。このプロセスにより、ユーザが入力したテキストとマッチする略地図 CCM 画像を得ることができ、これらを略地図の生成に利用する。

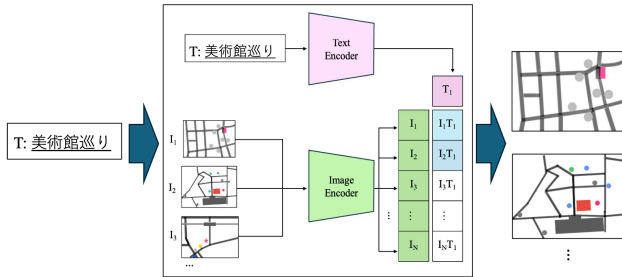


図 4: 目的に合致した略地図 CCM 画像の抽出

### 3. おわりに

本稿では略地図の自動生成に向けて、画像とテキストのコントラスト学習によって構成した学習モデルを用いたテキストと略地図 CCM 画像のマッチング手法を提案した。提案手法では、まず、Web ページからテキストと略地図画像を収集し、テキストと、略地図画像から作成した略地図 CCM 画像を用いて CLIP モデルによる事前学習を行い、ユーザが入力したテキストと類似度の高い略地図 CCM 画像を提示することを目的としている。

今後は、テキストと地図画像のマッチングについて、精度の高い学習を行う方法の検証を行う。また、提示された略地図 CCM 画像を基に略地図に表示する地物や、強調方法を学習し、略地図を自動生成する手法について検討する予定である。

### 参考文献

- [1] 福地 湧, 牛尼 剛聡, 田中克己, and 角谷和俊. 地理オブジェクトの位置関係と地物カテゴリに基づく地域特性推定. 第 16 回データ工学と情報マネジメントに関するフォーラム (DEIM 2024) T5-B-9-04, 2023.
- [2] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*, 2021.
- [3] Kürşat Kömürcü and Linas Petkevičius. Zero shot classification for change detection in satellite imagery. In *2024 IEEE 11th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, pages 1–6, 2024.