

観光の形態に基づいた 旅行ブログエントリの自動分類と可視化

¹柴田 有基, ¹篠田 広人, ²難波 英嗣, ³石野 亜耶, ¹竹澤 寿幸
¹広島市立大学大学院, ²中央大学, ³広島経済大学

①研究の目的

大量の旅行ブログを対象に旅行者（ブログ著者）の観光形態に基づいて自動分類し、この分類結果を地図上に可視化するシステムを構築。

インフラツーリズムの例 Infrastructure Tourism

- Title : **Detroit Dam**
- Text : On the drive home from the Northwest Cashmere Field Day and Fleece Show, we stopped to enjoy the view of the **Detroit Dam** and also walked across it.
- Image title: **Detroit Dam**



ヘルスツーリズムの例 Health Tourism

- Title : The Volcanic Island of Kyushu
- Text : I've had a very enjoyable couple of days here, exploring the sights of a typical Japanese **onsen** town.
- Image title: Tannoy Local **Onsen** Bath



カルチュラルツーリズムの例 Cultural Tourism

- Title : First **Matsuri**, karage and cotton candy ramen
- Text : **Matsuri** are **festival** in Japan to celebrate gods, food and good fortune during the year.
- Image title: Pushing up the **Danjiri**



②分類方法

- テキスト解析
- 画像解析（物体検出）
- Wikification

これらの解析結果から観光の形態を自動分類。

③それぞれの分類の例

The Shirakawa-Go village is not a very big place as I soon found out after bussing my way to the viewing point for an overview of the place.
Even though this is a **UNESCO Heritage Site**, I liked that the authorities have not overly exploited the site for tourist dollars.



ヘリテージツーリズム



Google Cloud Vision APIの結果

- footwear
- **telemark skiing**
- **piste**
- **snowboard**
- **ski**
- geological phenomenon
- **ski equipment**
- **snowshoe**
- **snow**
- **winter sport**

スポーツ
ツーリズム

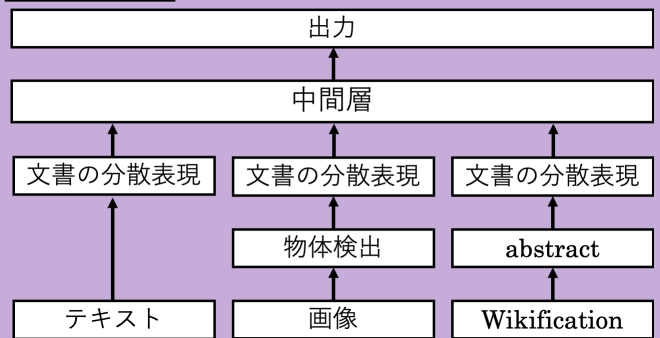
With perfect weather before the impending storm, I jumped onto the opportunity for a whirlwind Shinkansen excursion to **Himeji** and **Kyoto** for the day. Home to the magnificent white castle, the beautiful **Himeji-Jo** was evident as soon as I made my way out of the train station.

Wikipedia
姫路城は世界遺産

ヘリテージ
ツーリズム



④分類器



⑤実験

観光の形態に基づいた旅行ブログエントリの自動分類の実験を行った。

「インフラ、ハードツーリズム」168件、「ヘルスツーリズム」125件、「スポーツツーリズム」57件、「グリーンツーリズム」453件、「ヘリテージツーリズム」198件、「カルチュラルツーリズム」49件。

手法(素性, モデル)	精度	再現率	F値
Ensemble (提案手法)	0.807	0.179	0.293
Ensemble (ベースライン手法)	0.747	0.216	0.335
SCDV [txt+img+wiki (abst)] (提案手法)	0.752	0.218	0.338
SCDV (txt)	0.639	0.169	0.268
SVM (txt)	0.654	0.272	0.385
SCDV (img)	0.725	0.140	0.235
SVM (img)	0.788	0.170	0.279
SCDV [wiki (abst)]	0.528	0.116	0.191

⑥今後の課題

- アンサンブル学習の工夫
(例) スタッキング
- Wikipediaの全文を考慮する

参考文献

D. Mekala. "SCDV: Sparse Composite Document Vectors using soft clustering over distributional representations".