

# マルチモーダル埋込みを利用した 鳴き声からの野鳥検索システムの開発

## Development of A Search System through Multimodal Embeddings for Bird Song Species Identification

寺本 優香<sup>1\*</sup>

小島 諒介<sup>2†</sup>

TERAMOTO Yuka<sup>1</sup>

KOJIMA Ryosuke<sup>2</sup>

<sup>1</sup> 同志社大学<sup>1</sup> Doshisha University <sup>2</sup> 京都大学<sup>2</sup> Kyoto University

### 概要

野鳥検索システムは、自然環境理解や環境モニタリングの補助に有用なツールである [1]。見通しの悪いフィールドでの野鳥の検索においてはとりわけ音声による検索の需要が高い。既存の音声検索システムの多くは単一モダリティによる検索であるが、鳥種間の階層性や類似度を多面的に表現する必要がある [2]。本研究では知識グラフ・自然言語によるキーワード・音響の観点からの3種類のマルチモーダルな埋込みを用いた野鳥音声検索システムを提案する。さらに各モダリティごとに獲得した埋込みベクトルの分析を行う。

**知識グラフ埋込み** node2vecにより、Wikidata から得た知識グラフからグラフ埋込みを取得

**単語埋込み** 英語および日本語 BERT により、鳥類の名称ごとの単語埋込みを取得

**音声埋込み** wav2vec2.0により、日本野鳥大鑑・バードリサーチの音声データから音声埋込みを取得

システム内部では、鳥種ごとに3種類のモダリティを統合した埋込みを利用し、入力情報と最も類似した鳥種を順に検索結果として返す。また ChatGPT API による自然言語での解説出力を行う。

鳥種概念を異なるモダリティから表現するこれらの特徴量を用い、鳥種間の類似度行列を作成した。音声埋込みは他の類似度と比べ、最も類似度が高くなる傾向にあり、wav2vec の教師なしでの鳥種音声分類の困難さを反映している。

また、鳥種をノードとする巨大なマルチモーダル埋込

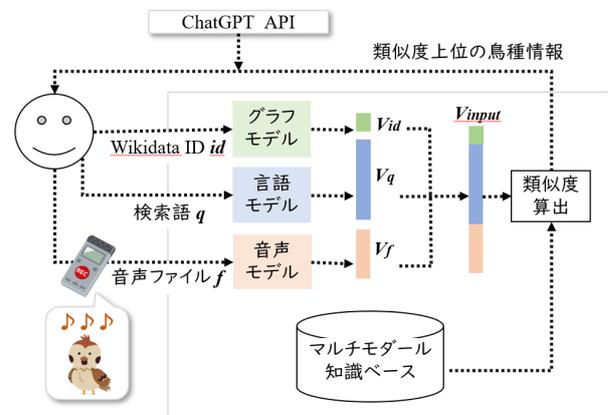


図1 マルチモーダル検索システムの概要

み統合グラフを構築し、各モダリティの類似度が閾値を超えた場合にエッジを張ることでモダリティが示す類似関係を可視化した。結果、特定のモダリティのみが検索可能な鳥種間の類似関係が多数存在することが判明し、複数モダリティを併せて鳥種の検索に用いることの重要性が示された。

### 謝辞

本研究は JSPS 科研費 No.20H00475, 19KK0260 の助成を受けた。

### 参考文献

- [1] 矢川雄一ほか. 鳥類図鑑 hyperbook における鳴き声検索. 全国大会講演論文集, データ処理, pp. 825–826, 1990.
- [2] Shah Nawaz, et al. Are these birds similar: Learning branched networks for fine-grained representations. In IVCNZ, pp. 1–5. IEEE, 2019.

\* teramoto@mil.doshisha.ac.jp

† kojima.ryosuke.8e@kyoto-u.ac.jp