

審査の結果の要旨

氏 名 牛久 祥孝

本論文は「Automatic Sentence Generation for Images via Key-phrase Estimation using Large-Scale Captioned Images」と題し、画像とその説明文のみからなる大規模データセットを用いて一般的な画像の説明文を自動生成する手法を提案するものであり、全7章からなる。入力画像を複数のラベルで表現する一般画像認識がこの10年で広く取り組まれてきたのに対し、それらの関係をも包含した文として入力画像を説明する技術が近年注目されている。だが従来の研究では、それぞれの画像の物体や動作、光景などを明示したラベルで事物間関係を表現し、新規画像の説明文を生成していた。多様な画像を説明するには大規模な画像データセットが必要不可欠だが、大量の画像に同様のラベルを付与するのは困難である。本論文は、画像とその説明文であればWebから大規模に収集可能である点に着目し、そのようなデータセットを学習して新規画像の説明文を生成する手法を提案している。実験を通じて大規模データセットの有用性を確かめながら、より人手のかかるデータセットを用いる従来研究よりも高精度に説明文を生成できることを示している。

第1章「Introduction」では、既存の画像や動画の説明文を生成する手法について、「主体、動作、光景」などの属性を表すラベルを手作業で与える必要があり、多様で大量なデータへの拡張が困難である点を指摘し、画像と説明文のみからなるデータセットでも画像説明文を生成できる手法の開発を本論文の目的としている。

第2章「Methods to Describe Multimedia with Natural Language」では、従来の画像認識から近年の画像説明文生成手法までを俯瞰した上で、「画像の内容はいくつかのキーフレーズで表現でき、これらを単純な文法モデルで正しく繋げば説明文を生成できる」という仮説を立てている。この仮説に基づき、画像と説明文のみからなるデータセットを学習した上で新規画像のキーフレーズを推定する問題を新たに提起している。

第3章「Investigation of Online Learning Methods for Multiclass Classification」では、まず大規模画像分類問題に注目している。データを一つずつ読み込むオンライン学習は大規模データと相性が良いが、近年のオンライン学習手法はコンピュータビジョンの枠組みで評価・比較されていない。主要な学習手法で大規模画像の分類性能を実験的に調べ、画像認識におけるオンライン学習の重要な知見を得ている。

第4章「Multi-Keyphrase Problem and Sentence Generation」では、複数のキーフレーズを文法モデルで繋ぎ、説明文を生成する手法について提案している。実際に画像と説明文のみからなるデータセットを用い、画像の説明文が生成可能であることを実験的に確認している。

第5章「CoSMoS: Common Subspace for Model and Similarity」では、部分空間を利用した効率的かつ高精度なキーフレーズ推定手法を提案している。キーフレーズ学習は従来の一般画像認識よりもラベルごとの学習データ量が少なくなるため、安定した高精度なキーフレーズ推定が困難になる。同じキーフレーズを持つ画像同士が近づくような部分空間と、部分空間内での各事物への分類器を同時学習する手法を提案し、高精度に画像の事物を認識できることを実験的に示している。

第6章「Evaluation of Sentential Description for Images」では、画像と説明文からなる大規模データセットに提案手法を適用している。規模の異なる3種類のデータセットを用い、100万枚の画像を含む大規模なデータセットでも提案手法が適用可能なこと、従来の画像説明文生成の手法よりも高精度に説明文が生成できることを確かめている。さらに、データ量が増えるほど説明文の精度が改善される結果も示されている。

第7章「Conclusion and Future Work」では、以上を総括した上で、画像と文章のみからなるデータセットで説明文を生成する、キーフレーズ推定アプローチの提案と、大量なキーフレーズを効率的に学習するための、部分空間を用いた新たな学習手法が本論文の学術的寄与であると結論付けている。

以上、これを要するに、本論文は、画像への説明文生成で最も重要な問題であった、画像とその説明文のみと言う大規模に収集可能なデータからでも画像説明文を高精度に生成する手法を提案している。画像の事物やそれらの関係性を端的に説明するいくつかのキーフレーズを推定し、これらを文法的に正しくつなげるというキーフレーズ推定アプローチと部分空間を利用した高精度なキーフレーズ推定手法を提案し、その有効性を示したものである。これにより、従来は事実上不可能であった一般的な画像に対する説明文生成への端緒が開かれたと言える。

以上の理由から、本論文は知能機械情報学上貢献するところ大である。よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。