大規模画像検索ログデータを活用した画像キャプションの自動生成

村本　尚生

画像キャプション生成とは，画像の内容を理解し，説明する文を生成することである．画像キャプション生成は人間にとっては簡単な問題である．しかし，コンピュータが画像のキャプションを生成するためには，画像の内容を理解し，言語で説明する必要があるため，研究者が長年にわたって取り組んできた課題である．コンピュータが画像のキャプションを生成する意義は多くある．例えば，画像にキャプションが付いているとキーワードを使って画像を探すことは簡単になり，閲覧による検索が必要なくなるなど，画像にキャプションがついていると，我々は様々な恩恵を受ける．

画像のキャプションを生成する際に，画像キャプション生成モデルは，画像の特徴量をキャプションへ"翻訳"している．画像の特徴量として，先行研究では，多くの場合，畳み込みニューラルネットワーク(CNN：Convolutional Neural Network)が抽出した，特徴ベクトルを使用している．しかし，画像には，複数種の特徴量があり，画像キャプション生成の際に，どの特徴に焦点を当てるべきか判断することは，大きな課題となっている．この課題を解決するために，CNN が抽出した特徴ベクトルに加えて，画像検索ログデータのクエリからユーザが焦点を当てた画像の特徴を表している単語を，ヒントワードとして抽出し，画像の特徴として与えるモデルを提案した．

画像検索ログデータを活用した画像キャプション生成手法の提案，画像検索ログデータの画像キャプション生成への有用性を検証するために，2 つのステップから成る実験をした．実験のステップ 1 では，キャプション生成モデルにヒントワードを与える際に，ヒントワードの数，選び方などをどのように設定するとパフォーマンスが向上するか検証した．適切なヒントワードを与えると，適切なキャプションを生成するモデルを訓練した．訓練データとして，8000 枚の画像とそれぞれの画像に 5 文のキャプションが与えられているデータセット，Flickr8k を用いた．実験のステップ 2 では，画像検索のログデータ(Clickture-Lite)の画像について，Clickture-Lite のクエリからヒントワードをステップ 1 で優れた結果を出した方法で抽出し，与えた提案モデルと Marc Tanti らによって発表された論文にある，"merge-model"に基づいたベースラインモデル，それぞれによりキャプションを生成し，比較実験を行った．2 つの実験の結果，ヒントワードとして，名詞をランダムに選び，ヒントワードの数が 2 つのモデルが最も優れており，ベースラインモデルと比べ，画像キャプション生成に役立つと分かった．

以上より，実験で最も優れた結果を出した手法を，画像検索ログデータを活用した画像キャプション生成手法として提案した．また，実験結果から画像検索のログデータが画像キャプション生成に活用できることが示された．

（指導教員　于　海涛）