

画像検索における画像生成技術の活用に関する研究

稲垣 遼

近年では、検索アルゴリズムの発展などによって、検索エンジンにクエリを投げかけると多くの場合自分が求めている画像を得ることができる。しかし、学校でのレポートやプレゼンテーションの準備などで特定の画像が必要になる際、求めている画像の特徴を短くまとめたテキストで表すことが難しく、クエリが長くなり、求めている画像が得られないことがある。ここで、検索効率が減少してしまうような長いクエリは冗長なクエリと呼ばれる。本研究では、冗長なクエリが与えられた時求めている画像が得られにくいという問題を解決する手法の一つとして、敵対的生成ネットワークに着目し、特にテキストから画像を生成することが可能な **Attentional Generative Adversarial Networks (AttnGAN)** というモデルに着目した。AttnGAN では、従来のモデルに比べて、クエリの各単語の意味を考慮した画像の生成が可能になったため、冗長なクエリをテキストのまま画像検索するのではなく、AttnGAN を用いて画像を生成し、その画像を用いて類似画像検索を行った方が良い結果が出るのではないかと考えた。したがって、本研究では、冗長なクエリから AttnGAN を用いて画像を生成し、その画像を基に類似画像検索をするというフレームワークを提案した。

本研究は、画像検索における画像生成技術の活用を目的とし、多くある画像の一つとして鳥という画像群を対象とした。鳥に関する冗長なクエリが与えられた時の、テキストによる画像検索結果と、AttnGAN を用いて生成した画像による画像検索結果の上位 5 枚ずつを比較することで、提案手法の有効性を検証した。

鳥に関する冗長なクエリは Clickture-Lite という画像検索のログデータから 10 個取得し、AttnGAN は Caltech-UCSD Birds-200-2011 という鳥の画像とその鳥の特徴の説明から成るデータセットを用いて学習させた。評価実験では、一つのクエリに対してテキストの検索結果と提案手法の検索結果、各 5 枚ずつ、計 100 枚の画像に対して適合度を評価した。

実験結果を NDCG, MAP を用いて評価したところ、多くの場合、ベースラインであるテキストによる画像検索の方がスコアが高くなったものの、一部のクエリにおいては提案手法の方がスコアが高くなり、提案手法が有効な場面があることが示された。

10 個のクエリの NDCG のスコアを平均した値において、ベースラインでは、評価に用いる画像数の増加に伴う NDCG の平均値の増加は見られなかったが、提案手法では、評価に用いる画像数の増加に伴う NDCG の平均値の増加が見られた。したがって、今後の課題として、実験で評価に用いる画像を上位 5 件ではなく上位 10 件、20 件と増やし、再度評価実験を行う必要がある。また、本研究で使用したデータセットやモデルでは鳥の画像しか生成できず、画像の精度にも限界があるので、規模の大きい画像生成モデルを用いた時の提案したフレームワークの性能を検証する必要がある。

(指導教員 于海濤)