

Weighting Methods for Information Retrieval Models and Video Retrieval Experiments

その他のタイトル	情報検索モデルにおける重み付け法と映像検索データを使用した実証実験
学位授与年月日	2017-09-15
URL	http://doi.org/10.15083/00077681

審査の結果の要旨

氏 名 村 田 眞 哉

本論文は、「Weighting Methods for Information Retrieval Models and Video Retrieval Experiments (情報検索モデルにおける重み付け法と映像検索データを使用した実証実験)」と題し、5章よりなる英文で書かれている。本論文は文書検索分野で発展を遂げている情報検索モデルと、そのインスタンスサーチと呼ばれる特定物体検索(映像検索)タスクへの適用に関するものである。著者は既存の情報検索モデルに新しいアプローチを導入し、特徴量の重み付け法として3つの新方式を提案している。そして各提案法の有効性をインスタンスサーチの実験により実証している。また、提案モデルと既存モデルの関係性を明確化し、本論文の結果が情報検索分野や映像検索分野にもたらす影響を論じている。

第1章は、「Introduction (序論)」であり、既存の情報検索モデルの内、特に代表的な手法が体系的に分類され、整理されている。また、本論文の貢献が体系内に位置付けられ、既存の検索モデルとの関連が論じられている。そしてインスタンスサーチのデータセットを使用した実験により、提案法の有効性を検証することが示されている。

第2章は、「BM25 with Exponential IDF for Instance Search (Exponential IDFを用いたBM25によるインスタンスサーチ)」と題し、これまで適用されてきた簡単な検索モデルではなく、文書検索分野の代表的な確率的検索モデル(BM25)を映像検索に応用することを提案している。その際、BM25をそのままの型で適用するとインスタンスサーチの精度は不十分であるため、BM25-IDFと呼ばれる特徴量の重み付け部分を変更することを提案している。その結果、従来法と比べて高い検索精度が得られることを実験的に示している。

第3章は、「Bayesian EIDF and ROI for Enhancing Instance Search Accuracy (Bayesian EIDFとROIを使用したインスタンス検索精度の向上)」と題し、第2章で提案されたExponential IDFの設計理論を展開している。ある特徴量が映像の中で主題となる確率に事前情報を付与し、画像クエリに対する適合情報が与えられた後の事後分布をベイズの定理に基づき評価することで、Exponential IDFと類似した重みが導出されることが示されている。また、本理論では、BM25-IDFは事前情報として無情報事前分布を採用することに対応し、有情報事前分布を使用する提案法は通常のBM25の拡張

になる。本章の結果から、映像検索タスクの精度向上においてはベータ分布に基づく情報事前分布の使用が効果的であることがわかり、この点が文書検索と大きく異なる。

第4章は、「IR Model using Generalized Pareto Distribution and Its Application to Instance Search (Generalized Pareto Distributionに基づく情報検索モデルとそのインスタンスサーチへの適用)」と題し、極値統計学の知見を検索モデルに適用することを提案している。この様なアイデアは初めてではなく、実際、既存研究では percentile-based model と呼ばれる検索モデルに極値統計学の知見が適用され、高い検索精度が示されている。反対に、著者らは information model と呼ばれるシンプルな検索モデルに極値統計学の知見を適用し、検索対象のデータの平均超過関数 (mean excess function) からパラメータが推定される、パラメータチューニングが不要な検索モデルの構築を実現している。また、インスタンスサーチの実験結果から十分な検索精度も確認し、ある閾値以上のデータのみでも高い検索精度が保持されることを示している。従って、その様な閾値超過データのみを使用による、処理コストの大幅な低減についても触れられている。

第5章は、「Conclusions (結論)」と題し、本論文の貢献のまとめ、各提案法の分類と位置づけ、および未解決の問題が示されている。未解決の問題としては、axiomatic approach と呼ばれる経験的な公理群による各提案モデルの解析や、その他の特徴量に対する Generalized Pareto distribution の仮定の正当性検証が挙げられている。また、今後の研究方針として、Generalized Pareto distribution の離散形版に対応する Zipfian distribution に基づく検索モデルの構築の可能性が論じられている。

以上これを要するに、本論文では、情報検索モデルにおける特徴量の重み付け法として3つの新手法が展開され、インスタンスサーチによる映像検索実験によりそれらの有効性が実証されている。3つの手法の内最初の2つは従来法と比べて高い検索精度を示し、最後の1つはデータからパラメータが推定可能なデータ駆動型の検索モデルであり、今後、実問題への応用が期待される。さらに、情報検索モデルに基づく映像検索技術の基礎を構築したと言え、電子情報学上貢献するところが少なくない。

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。