氏　　名	森　郁海
学　位　名	博士（システム情報科学）
学位記番号	第６０号
学位授与年月日	令和４年３月２２日
学位論文題目	エッジ AI における訓練データ検索システムに関する研究
論文審査委員	主査　稲村　浩
	副査　白石　陽
	副査　新美　礼彦
	副査　藤野　雄一
	副査　中村　嘉隆（京都橘大学　准教授）

論　文　要　旨

エッジ AI（Artificial Intelligence）は，セキュリティの問題や応答時間へのリアルタイム要求に応えるなどの目的で，データの生成元に近いエッジ上で機械学習や深層学習などの知的処理を実行するものである．

エッジ AI が適したユースケースは，工場の検品作業や人物行動分析，自動運転などのような，画像を入力として即時的な出力が要求されるものである．さらに，これらのユースケースでは，高度なセキュリティ対策やプライバシー保護が求められるため，クラウド上に訓練用の画像データを集約することが難しい．したがって，エッジ上で収集した画像データのみを訓練に利用することになるが，しばしば訓練データの不足が問題となる．

このような訓練データ不足を解消するために，高度なセキュリティ対策やプライバシー保護を実現しながら，AI の学習に有効な訓練データを検索するシステムが必要である．この訓練データ検索システムでは，まず，訓練データの暗号化したキーワード群と，その訓練データを所有するエッジの所在地を示す，暗号化した URI（Uniform Resource Identifier）からなるカタログ情報をクラウド上に集約しておく．次に，訓練データを検索するエッジは，クラウド上のカタログ情報の暗号化されたキーワードに対して検索を行い，訓練データの候補となるデータの URI を得る．そして，訓練データを検索するエッジは，URI に示されたエッジにアクセスし，データの利用許諾を取得した後，データをダウンロードする．最後に，訓練データを検索するエッジは，ダウンロードした画像データが訓練データとして有効かどうかを判定する．

以上のような訓練データ検索システムを実現するためには，クラウドにおいてスケーラ

ビリティとセキュリティを維持しながら暗号化されたキーワードどうしを検索する技術と，エッジにおいて訓練データとして有効な画像かどうかを判定する技術の開発が必要である．本論文では，クラウドとエッジは，担保すべきセキュリティ対策とプライバシー保護のレベルが異なり，処理も分離可能であることから，これらの技術開発を独立した 2 つの問題ととらえ，各々の問題を解決することで訓練データ検索システムが構築可能であることを示す．

1 つ目の問題である，暗号化されたキーワードどうしの検索を高速に実行する技術については，高度なセキュリティ対策とプライバシー保護を実現するために，キーワードに加え検索処理自体もクラウドから秘匿する検索可能暗号を用いる．そのうえで，スケーラビリティを確保するために，検索速度とセキュリティを考慮した検索可能暗号の高速化パラメータを自動的に決定する手法を提案する．従来は，高速化パラメータの設定値をシステム管理者が手動で決定していたが，提案手法は，検索に使用するキーワードの最小エントロピーと k-匿名性を用いて，検索性能とセキュリティがバランスするパラメータ値を自動的に求める．数十万キーワード規模での評価の結果，提案手法によって検索時間を最大 97.2%削減しつつ，データベースが 2,598-匿名性を持ち，検索速度を向上させながら高いセキュリティレベルを達成していることを確認した．

2 つ目の問題である，訓練データとして有効な画像かどうかを判定する技術については，エッジの限られた計算資源で実行できるよう，特徴点マッチングをベースとした類似画像検索を用いる．そのうえで，訓練データとして有効な画特を判定できるような，特徴点マッチングの類似度指標を提案する．従来のユークリッド距離などの単純な類似度指標は，照明変動や画像に占める被写体の割合の違いなどの環境ノイズを含む画像を検索しにくい．そこで，提案方式は，このような環境ノイズに対して不変性を持たせるために，画像のヒストグラムの形状に着目して類似度を計算する．具体的には，ヒストグラムの類似度計算において，ヒストグラムの形状が平行移動したり，伸縮したり，相似形である場合でも類似度が高くなるように，類似度評価区間を極値で分割し，区間ごとに DTW（Dynamic Time Warping）距離を求め，各距離を結合することで類似度を得る．画像認識でのユースケースを想定した評価において，提案手法は，ヒストグラムの形状に平行移動や伸縮，相似形が存在する画像どうしの類似度を高く算出できることを確認した．

1 つ目の提案により，十分なスケーラビリティを持ち，高いセキュリティを維持しつつ，検索速度が向上する秘匿検索技術を確立した．2 つ目の提案により，環境ノイズに対して不変性を持つ訓練画像データを検索できるようになり，エッジが収集した,ドメインが似ている画像データを AI の学習に相互利用できる技術を確立した．これらの結果から，エッジ AI における訓練データ検索システムの実現課題である 2 つの独立した問題が解決され，システムが構築可能であることが示された．

審査結果の要旨

クラウドとエッジが連携するAI向けの訓練データ検索システムの実現を目的として，システムの検索速度や検索精度のボトルネックを解消する以下の２つの新しい手法を提案し，その有効性を論じている．それぞれ（１）暗号化されたキーワードの検索における性能とセキュリティを考慮した高速化パラメータの決定手法　（２）データ転移に好ましい訓練画像データが検索可能な特徴点マッチングの類似度指標である．

１章では，発展を続けているエッジAIにおいてその推論機構の構成には転移学習が有効であることを受け，要求に合致する訓練データを検索によって見付けるためには，異なるセキュリティ要件のもとでの高速な検索手法が求められることを述べている．２章では，検索可能暗号と特徴的マッチングの類似度指標についての関連研究を述べている．

３章では，検索可能暗号における性能とセキュリティを考慮した高速化パラメータの決定手法の詳細を述べている．この章では，検索性能とセキュリティがトレードオフとなる検索可能暗号に対して，検索性能とセキュリティがバランスする高速化パラメータを，検索に使用するキーワードの最小エントロピーとk-匿名性を用いて求める方法を提案している．

４章では，データ転移に好ましい訓練データが検索可能な特徴点マッチングの類似度指標の詳細を述べ，データの類似度を，画像のヒストグラムの形状に着目し計算する方法を提案している．訓練画像の類似度計算において，画像のヒストグラムの形状が平行移動したり，伸縮したり，相似形である場合でも類似度が高くなるように，類似度評価区間を極値で分割し，区間ごとにDynamic Time Warping(DTW)距離を求め，各距離を結合することで類似度を計算する方法を提案している．

５章では，結論と本研究の成果を利用した新たな研究領域について述べている．

本研究は深層学習に代表されるアルゴリズムの発展に伴い提供可能になった知的な処理を，人間の社会活動に適用するためにはIoT(Internet to Things)での推論実行が必要であり，そのためには訓練学習データの円滑な流通が鍵であることを指摘し，その解決に有効な手法を提案している．この成果はシステム情報科学の発展に寄与するものであり博士論文として十分であると判断する．