

安全な類似ドキュメント検索のための インデックスを使った高速化手法

Efficient Privacy-Preserving Similarity Search Schemes Using Indices

学籍番号：201821602

氏名：有友 大輔

Daisuke Aritomo

クラウドコンピューティングプラットフォームは従来のオンプレミス環境と比べ、非常に安価でスケーラブルな計算資源を提供している。また、このクラウド環境にはあらゆる種類のアプリケーションがデプロイされ、それらが処理するデータもデータも同時にクラウドに置かれている。しかしながら、クラウド環境の利用には従来存在しなかったプライバシーの問題がつきまとう。クラウド環境の活用は、結果的にクラウドプロバイダーにデータの読み取りを許すことに他ならないからだ。

情報検索（IR）アプリケーションにおいて、この問題を解決するための取り組みが進んでいる。ドキュメントの文書ベクトルとクエリのベクトルとのコサイン類似度を用いる類似文書検索については、準同型暗号を用いて安全に実装する方法が提案されている。この方法では準同型暗号を用いることで、秘密鍵へのアクセス権を持たないクラウド環境の計算資源におけるドキュメントのスコアリングを達成している。

しかしながら、準同型暗号を単純に適用するだけでは、検索のたびにクエリに対して全てのドキュメントの類似度を計算する必要が発生してしまう。そこで、本研究では、この問題を解決するためのインデキシングの手法を2つ提案する。

1つ目の手法は、ドキュメントに含まれる単語から、OR 検索を実現する木を構築する方法である。木の構造自体に意味がないため、セキュリティ面における優れた性質が期待できる。

2つ目の手法は、ドキュメントの文書ベクトルから VP-tree を作り、定数時間での計算を実現する方法である。VP-tree の動作を高速化するため、SimHash を利用した次元圧縮も導入する。さらに、この木の各ノードを共通鍵暗号を用いて暗号化することで、セキュリティ面での問題の解決も図る。

また、これら2つの手法について、セキュリティ面の評価を行い、また現実のデータセットを用いた実験を通してパフォーマンス面の評価を行う。

研究指導教員：森嶋 厚行

副研究指導教員：若林 啓