

## 審査の結果の要旨

氏 名 顧 彦 慧

本論文は「**Research on Efficient Similar Sentence Extraction**（効率的な類似文抽出の研究）」と題し、英文8章から構成されている。ウェブ検索、推薦システム、質問回答システム等の環境において数語のキーワード問合せに対する効率のよい類似文抽出手法の確立を目的とし、候補データの逆引き索引を用いてデータの探索コストを減少させることで、精度を保ったまま効率よく短文を抽出する手法を提案している。本手法を用い、地理情報を考慮した推薦システムにおいてユーザ要求に適合した結果を効率よく提示するべく、類似文の抽出手法と地理情報の探索手法を統合し、データ探索量を減少する索引を構築し、効率良く探索結果を提示可能な手法を提案し、ウェブ上の実データを用いて提案手法の有効性を示している。

第1章は、「**Introduction**（序章）」であり、本論文の背景および目的について概観し、本論文の構成を述べている。

第2章は、「**Problem Definition and Related Work**（問題定義と関連研究）」と題し、効率のよい類似文抽出における課題を明らかにすると共に、類似文抽出技法に関し、文字列、語順、知識ベース、コーパスベースの手法についてまとめている。

第3章は、「**Efficient Similar Sentence Extraction**（効率のよい類似文抽出）」と題し、数語のキーワード問合せにおける効率のよい類似文抽出手法の確立を目的とし、文字列の類似手法およびコーパスを用いた類似手法を組み合わせ、各手法に候補データへの逆引き索引を導入し、データの探索量を減少させることで、精度を保ったまま上位k件の類似文を抽出する手法を提案している。

第4章は、「**Performance Evaluation of Efficient Similar Sentence Extraction**（効率のよい類似文抽出手法の性能評価）」と題し、文字列ベースの類似度手法およびコーパスベースの類似度手法の組合せにおける重み付けについて最適化を試みるとともに、提案手法の効率性について、データサイズ、抽出文数、および語数を変化させ、その有効性を確認するとともに、従来の手法と比較して同等の類似度が保てることを確認した。

第5章は、「**Extension on Efficient Similar Sentence Extraction**（効率のよい類似文抽出手法の拡張）」と題し、WordNet, Wikipediaをコーパスとして利用する手法を提案し、ベンチマークデータセットを用いて評価を行い、類似度を改善できる余地があることを示した。

第6章は、「**Application of Efficient Similar Sentence Extraction by Incorporating**

**Spatial Information**（空間情報を利用した効率のよい類似文抽出のアプリケーション）」と題し、地理情報を考慮した推薦システムにおいて、アプリケーション実行時にユーザ要求により適合した結果を効率よく提示するべく、類似文の抽出手法と地理情報の探索手法を統合し、データ探索量を減少する索引を導入した。また、位置情報サービスである foursquare のデータを用い、提案手法を評価し、有効性を確認した。

第7章「**Discussion**（議論）」では、提案手法で用いる語単位の類似検索手法として異なる手法の利用可能性、語と独立した類似度検索手法の可能性、効率のよい空間情報類似検索手法に関し、提案手法との組合せ等について議論している。

第8章「**Conclusion**（結論）」では、本論文の成果と今後の課題について総括している。

以上これを要するに、本論文は、数語のキーワードから精度を保ちつつ増大するデータ量に対応可能な類似文抽出手法の確立を目的として、候補データへの逆引き索引に基づき上位  $k$  件の類似文を効率よく抽出する手法を提案し、様々なデータセットを用いた実験により提案手法の有効性を明らかにしており、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。