審 査 の 結 果 の 要 旨

氏 名　　金 英子

本論文は「Extraction and Application of Social Networks from World Wide Web（Web からの社会ネットワークの抽出及び応用）」と題し，英文８章から成る．第３～６章を第Ⅰ部として Web からの社会ネットワーク抽出の研究成果について記し，第７章を第Ⅱ部として社会ネットワークの応用の研究成果を記している．

第１章は「Introduction（序論）」である．社会ネットワークはアクター（人物，会社や組織）をノードとして，アクター間に存在する繋がりや関係を結んでネットワークとしたものであり，その分析は社会科学分野で行われてきた．Web 上に大量の情報が公開されアクセス可能になったことで，Web 情報から社会ネットワークを抽出し，分析，利用することが Web マイニングの一種として新しい研究が興ってきていることを述べ，本研究の貢献事項について記している．

第２章「Background and Existing Studies（背景と既存研究）」では，まず背景となる関係研究として Web マイニング，Web からの情報抽出の状況を紹介している．次いで，社会ネットワークの抽出と分析の基礎となる考え方，技術について纏めている．

第３章「Problem Definition（問題設定）」では，Web からの社会ネットワーク抽出に関する関連研究について記し，既存手法の問題点について述べている．人間間関係抽出の典型的な既存手法では，２つのアクターを表すエンティティが１Web ページ内で共起する Web ページ数が多い程，関係が強いとの認識の下，Web 検索エンジンの AND 検索でこの共起回数を計測している．この第一の問題点は，例えば企業間関係の抽出では，２企業が良く知られた企業であるとこの２企業名の AND 検索を発すると，ニュース文等を含め様々な文脈で非常に多数の共起ページ数が得られることになってしまい，抽出したい種別の関係の強さとはかけ離れた計測値になってしまう．第二の問題点は，共起 Web ページ数により関係強度を計測する既存手法は各アクターが異なる文脈で Web に現れるような，不均質性を有する社会コミュニティでの関係については，良く働かないことである．そして本研究では，第一の問題点に対して目的とする種別の関係に絞って強度を計測し，社会ネットワークを抽出する手法，第二の問題に対しては関係存在有無判定の閾値を状況毎に適応的に調整して，社会ネットワークを抽出する手法を提案していることを述べている．

第４章「Social Network Extraction for Complex Relations（複雑な関係に対する社会ネットワーク抽出）」では，第３章の第一の問題に対処する手法として，Web から企業間の特定の関係を関係識別手法を導入して抽出する手法を示している．基本的な考え方は，２企業名に加えて特定の関係を表す関係語（複数）も加えて AND 検索し，そのヒット件数によりこの関係に焦点を当てた強度計測を可能にする．この検索に必要となる関係語は，最初に指定した関係語（以下に記す例では“提携”と“訴訟”）と Web で共起頻度が高い語を，Jaccard 係数に基づいて選定する方法を採っている．

以上の手法により，日本の電子情報企業 60 社を対象として提携関係（資本提携と業務提携あり），及び訴訟関係（係争段階と和解段階あり）について関係強度を求め，関係ネットワークを抽出している．関係語の使用法は，関係語無しの AND 検索，最も重要度が高い関係語も加えた AND 検索，２番目に重要度が高いと推定された関係語も加えた AND 検索の結果を統合することにより，適合率，再現率の点で良い結果が得られることを示している．

第５章「Social Network Extraction for Inhomogeneous Communities（不均質コミュニティに対する社会ネットワーク抽出）」では，第３章の第二の問題に対処する手法として，Web

上で永続的でない弱い社会関係を抽出するため，ネットワーク全体から見て弱い社会的関係であっても，あるアクターにとって相対的に強い関係となるアクターを，４つのパラメータを状況適応的に調整することにより抽出する手法を示している．ここでの４パラメータは，エッジを張るか否かを決めるオーバーラップ係数とマッチング係数の閾値２個と各ノードからのエッジの最小数に関する２パラメータである．本手法は 2005 年横浜トリエンナーレに参加した 133 名のアーティストの関係ネットワーク抽出に用いられ，ユーザをナビゲートする Web サイトとして実際に運用された．

第６章「General Model of Social Network Extraction（社会ネットワーク抽出の一般化モデル)」では，上記第４,５章の社会ネットワーク抽出の手法を纏めて，Web から様々なエンティティ間の社会ネットワークを抽出するための，汎用的枠組みを提示している．

第７章「Ranking Entities Based on the Social Network（社会ネットワークに基づくエンティティのランク付け)」では，社会ネットワークでの特徴が各ノードとなるエンティティの重要度ランキングの推定に利用できることを示している．特徴としては，ネットワーク中心性（次数中心性，近接中心性と媒介中心性を含む)，複数の関係についてのネットワークの組合せ，到達可能ノード数や隣接ノードが持つ接続数などのネットワーク構造などであり，どのような特徴量がランキングに影響するかを，ランキングが得られているデータへの相関が最も高くなるような学習により定めている．第一のデータとしては，電子産業分野の 312 社について株式時価総額, 社員の平均年間収入額, 優れた企業についてのランキングデータを用いている．この場合に用いる複数ネットワークとしては，共起頻度に基づくネットワーク，業務提携ネットワーク，資本提携ネットワーク，同一株式市場ネットワーク，株保有関係ネットワークなどである．第二のデータとしては，東大の 253 名の教員について，発表論文数のランキングデータを用いている．この場合に用いる複数ネットワークとしては，日英についての共起頻度に基づくネットワーク，同一所属のネットワーク，同一プロジェクトのネットワークなどである．第二のデータの場合だと，日本語での検索ヒット数と英語で抽出した共起回数ネットワークでの次数中心性が，発表論文数と高い相関を有するなどの知見が得られたことを述べている．

第８章「Conclusion and Future Work（結論と今後の研究)」では，本論文の成果を纏め，今後の研究について言及している．

以上を要するに，本論文は Web マイニングの一種としての Web からの社会ネットワーク抽出手法に関し，２つのアクターが各 1 Web ページ内で共起する回数が多い程，関係が強いとする既存手法ではうまくいかない場合の問題に対処する手法を提示している．第一は，例えば企業間関係ネットワーク抽出では，２企業が良く知られた企業であると AND 検索を行うと様々な文脈で多数の共起ページ数が得られて，抽出したい特定の関係の強さとはかけ離れた計測値になってしまう問題に対し，抽出したい種別の関係語も付加して AND 検索する手法である．第二は，各アクターが異なる文脈で Web に現れるような不均質なコミュニティに対しては，全体から見れば弱い社会的関係であっても，あるアクターにとっては相対的に強い関係となるアクターを，関係存在有無判定の閾値を状況毎に適応的に調整して取り上げ，社会ネットワークを抽出する手法である．以上の手法の有効性を実験を通して具体的に実証している．更に，このような Web から抽出した社会ネットワークのネットワーク中心性といったような各種特徴量と，各エンティティの重要度のランキングの相関について実験的に検討し，このようなネットワーク特徴量が各エンティティの重要度ランキングの推定に利用できることを示している．これらの研究成果により，本論文は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）の学位論文として合格と認められる．