

Web 上のテキストを利用した金融市場予測の高精度化

研究代表者

大原 剛三

【研究の目的】

株価や為替などの金融市場の変動予測は、金融工学、統計学、情報工学など様々な分野において多くの試みがなされてきたが、十分な予測精度を達成するには至っていない。6割程度の正答率であれば高い方であり、個別の条件下で高い正答率を達成していても、別の条件下では同様の性能を再現できないのが現状である。そのような中、近年、数値的なテクニカル指標に加えて、Web 上から取得できるテキスト情報を活用したファンダメンタル分析、市場ムードの推定など、いわゆる Web マイニング、テキストマイニングと呼ばれる技術の適用事例が多く報告されるようになってきている。これらの研究では、市場ムードの変化が金融市場の変動の手がかりとなり得ること、双方の変化には一定のタイムラグが生じることなど、興味深い結果が報告されているものの、その成果は一般化されるまでには至っておらず、対象データ（銘柄、期間等の違い）に依存している。このような背景の下、本研究では、Web 上の複数の情報源（Twitter、株式市場関連掲示板、ニュースサイト）から収集したテキスト情報からより正確に市場ムードを推定し、それにより市場変動予測をより正確に予測するシステムの実現を目指す。そのために、1) 市場変動の予測値と実際の観測値の乖離度を前処理系にフィードバックする市場変動予測モデル、2) 評価表現とそれに対する評価極性を記録する評価極性辞書をフィードバック情報に基づき動的に更新することで市場ムードをより正確に推定する評価極性分析システム、および3) 収集テキストの情報源の比重をフィードバック情報に基づき適応的に変化させるアクティブ情報収集システムの実現を、それぞれ機械学習（大原・豊田）、自然言語処理（原田・谷津）、Web 情報処理（Dürst, 荘司）を専門とする研究グループが協力することにより機動的に進める。最終的にこれらの技術を統合することにより、長期的な利用を通して市場変動予測の精度を漸進的に改善するシステムを実現する。

【本研究の特色】

従来の Web 上のテキストを利用した金融市場予測では、市場ムードの予測に利用する情報源は基本的に単一であり、かつ情報収集方法は固定的であった。加えて、収集したテキストからの市場ムードの予測には、汎用的な評価極性辞書を用いていた。これに対し、本研究で実現を目指す枠組みは、アクティブ情報収集システム、評価極性分析システム、および市場変動予測モデルを有機的に結合し、市場変動予測モデルから与えられるフィードバック情報に基づき極性評価辞書の内容、およびテキストを収集する情報源に対する重みを適応的に更新するものであり、そのような連続的な利用過程において各システムが連動して機

能するフレームワーク自体に独自性がある。加えて、評価表現に対する極性値の動的更新、情報収集対象の能動的選択という個々の要素もこれまでにない独自のアプローチである。

【期待できる成果】

近年、金融分野では情報技術、とりわけ機械学習をはじめとした人工知能技術の導入が進んでおり、Fintech という新たな領域となっている。そのような領域において、ここで実現する技術の導入は、より低コスト、かつ信頼性の高いサービスの提供を可能とする。また、本研究では金融市場の変動予測を対象としているが、ここで確立する要素技術、およびフレームワーク全体は、同様にインターネット上のテキスト情報を情報源として各種の予測・推定を行うソーシャルセンシング一般に転用可能なものである。予測モデルからのフィードバック情報の自動生成、それに基づくアクティブ情報収集、評価極性辞書の動的更新のいずれもこれまでにない技術であり、ソーシャルセンシング分野に与える影響は大きい。

【研究計画】

研究初年度である 2018 年においては、本研究を構成する 3 つの主要課題について以下の目標の実現を目指す。

(1) フィードバック情報を自動生成する市場変動予測モデルの学習

収集したテキスト情報から推定される市場ムードを入力にもつ市場変動予測モデルを構築し、その予測値と観測値の乖離度から市場ムードの推定に利用する評価極性辞書、および情報収集対象を適応的に修正するためのフィードバック情報を自動生成する方法を確立する。

(2) 評価表現辞書の動的更新

過去の各種金融市場の変動率と各情報源から収集したテキスト情報から得られる評価極性値の変動の関係性に基づき、汎用的な評価極性辞書から各情報源に適した評価極性辞書（動的更新における基準辞書）を構築する。

(3) Web からのアクティブ情報収集

これまでの予備的分析の結果に基づき、Twitter、株式市場関連掲示板、経済関連ニュース記事を主な情報源とし、正確、かつ定常的に情報を収集できる基盤技術を確立する。特に、Twitter および掲示板記事におけるノイズ的な記述を高精度に除外する技術を確立する。

2019 年度以降は、初年度に検討を進めた各主要課題に対する成果の統合を進める。具体的には、市場変動予測モデルが出力するフィードバック情報を評価極性辞書更新システム、およびアクティブ情報収集システムが受け取り、収集した情報から市場ムードをより正確に推定し、市場変動の予測精度向上に必要な情報をより多く収集する技術を確立する。また、必要に応じて、評価極性辞書更新システムもアクティブ情報収集システムに対してフィードバック情報を与えることを検討する。さらに、長期的な運用テストを通して手法の洗練化を図るとともに、その有用性を明らかにする。