

ニュースと株出来高により 日々更新できる株式銘柄知識

最新の話題が企業に及ぼす影響 を把握して、株式市場予測に活用

インターネットで配信される経済ニュースから、話題（トピック）とその影響が波及する銘柄との関係を知識として、辞書化する技術を開発しています。トピックと、その影響を受ける銘柄の関連及び影響の強さである影響度を合わせて辞書化し、トピックに関するニュース件数と各銘柄の出来高を参照して影響度を評価し更新します。この“トピック辞書”により、最新の話題が企業に及ぼす影響を容易に把握できるようになります。

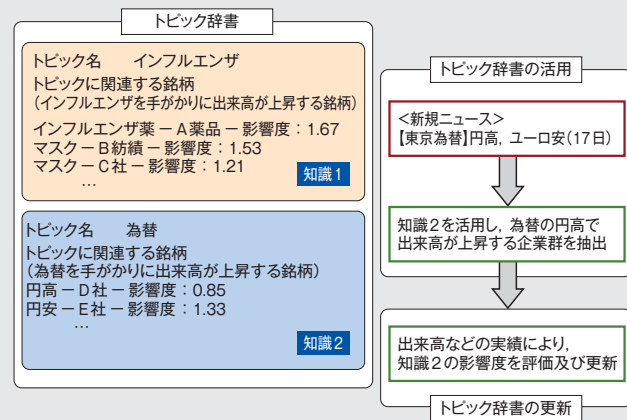


図1. ニュースと株出来高の活用 — 日々のニュース配信件数と出来高で影響度を更新することで、ニュース配信内容や社会状況などの変化に柔軟に対応できます。

表1. 東芝に対して影響度が大きい上位15のトピックとその影響度

年/月/日	2011/3/1		2011/3/17		2011/4/28		2011/7/29	
社会状況	東芝の専用眼鏡なし3Dテレビ発売約3か月後		東日本大震災直後		東日本大震災約1か月半後		地上アナログ放送終了直後	
項目	トピック	影響度	トピック	影響度	トピック	影響度	トピック	影響度
評価結果	3D	3.878	テレビ	5.000	テレビ	4.641	テレビ	3.866
	テレビ	3.260	iPhone	4.237	3Dテレビ	3.256	半導体	3.326
	3Dテレビ	3.206	3D	3.623	液晶テレビ	2.898	3Dテレビ	3.256
	半導体	2.512	電力	3.474	3D	2.853	液晶テレビ	3.098
	液晶テレビ	2.471	原子力	3.474	半導体	2.712	iPhone	2.242
	LED	1.977	3Dテレビ	3.256	鉄道	2.390	LED	2.147
	スマートフォン	1.876	鉄道	3.213	iPhone	2.222	メモリ	2.069
	メモリ	1.823	原子力発電	2.974	LED	2.162	HDD	1.939
	LSI	1.760	液晶テレビ	2.521	原子力	1.868	ノートPC	1.898
	システムLSI	1.660	半導体	2.512	メモリ	1.823	LSI	1.810
	iPhone ^(注1)	1.580	LED	2.202	LSI	1.810	蓄電池	1.677
	原子力発電	1.462	メモリ	1.823	システムLSI	1.660	システムLSI	1.660
	発電	1.332	LSI	1.760	HDD	1.520	NAND	1.543
	ノートPC	1.300	システムLSI	1.660	ノートPC	1.450	中小型液晶	1.518
	原子力	1.282	スマートフォン	1.566	NAND	1.342	液晶パネル	1.518

LED:発光ダイオード PC:パソコン HDD:ハードディスクドライブ

トピック辞書の評価を行いました。評価は、“東芝”に関して経済ニュースに記載された知識（トピック230種で構成したトピック辞書）に対して、前述の補正ニュース件数と出来高による評価と影響度更新で行いました。翌日の東芝の出来高の予測は、東芝に関連するトピックの補正ニュース件数の合計を手がかりとします。

影響度更新を続けた結果、東芝に関しては、表1に示すようなトピックと影響度が得られ、次のような社会状況の変化への対応が見られました。

- 2011年3月1日（東芝の専用眼鏡なし3D（立体視）テレビ発売約3か月後） “3D”や“3Dテレビ”などの影響度が大きい、すなわち東芝の出来高に関連の強いトピックと評価されています。
- 同年3月17日（東日本大震災直後） 震災で影響を受けたライフラインのうち、東芝に関連する

(注1) iPhoneは、Apple Inc.の商標。

“電力”などの影響度が大きな値になっています。

- 同年4月28日（大震災約1か月半後） “電力”がニュースで取り上げられても東芝の出来高に影響がなくなりつつあり、影響度も小さくなっています。
- 同年7月29日（地上アナログ放送終了直後） “テレビ”の影響度がやや小さくなりますが、電力不足対応で注目されるようになった“蓄電池”が大きな影響度を得るようになっています。

このように、トピック辞書は、世間で起きたでき事の影響の把握と知識化に有効であると言えます。

トピック辞書の利用効果について、ニュース配信件数と出来高の関係も次の三者を比較し評価しました。

- 明示的に“東芝”と書かれたニュースだけの配信件数
- 影響度を導入せずに230種のトピックと東芝に関するニュース配

信件数の合計件数
(3) 影響度を乗じた補正件数と東芝に関するニュース配信件数の合計件数
(3)で出来高との相関係数が改善し、トピック辞書の効果が確認できました。

今後の展開

銘柄に関するトピックのニュース件数と出来高を参照して、トピックと銘柄の関連の強さである影響度を更新することで、知識を社会状況に対応した内容に更新できました。

影響度を導入した精度の高いトピック辞書を作成し、更新していけば、最新の話題が企業に及ぼす影響を把握することが容易になります。今後、マイクロブログの発信傾向と企業への影響の関係を把握するなど他分野への適用や、必要な要素機能の開発を進めます。

牧野 恭子

東芝ソリューション（株）
IT研究開発センター 研究開発部主務

コンピュータによる株取引

2010年の（株）東京証券取引所 arrowhead 稼働により、コンピュータシステムが株価や出来高などに応じて自動的にタイミングや数量を決めて株式売買注文をするアルゴリズム取引が活発になりました。株取引に有効な情報をコンピュータが活用できる形で整備することがますます重要になります。

ニュースが企業に及ぼす影響

世間のでき事、例えばインフルエンザ流行の場合、“インフルエンザ関連銘柄”として知られる薬品会社やマスク製造会社などの株価や出来高に大きな変動が起きます。知識のある投資家は、インフルエンザ関係ニュースを入手すると、ニュースと関連性の強い銘柄を想起して取引を行います。株取引ではこのように、社会ニュースなどから関連銘柄を想起できる知識が有効です。関連銘柄の知識は、各社の事業内容や日々配信される経済

ニュース、特に“インフルエンザ流行開始が報じられたことからA薬品やB紡績などインフルエンザ関連銘柄が買われた。”のように株式市場の状況を伝えるニュースから得ることができます。

株式銘柄知識の更新

東芝ソリューション（株）が開発している“トピック辞書”は、経済ニュースに記載される銘柄をグルーピングする知識を体系化したものです。既存技術ではシソーラス（ことばを意味上の類似関係などによって分類した辞書）やオントロジー（概念を相互に関連付け知識として体系化したもの）で知識体系を表現します。企業情報などから知識体系を自動獲得する研究もあります。

トピック辞書では、知識を更に有効にするために、従来の知識体系に、トピック（グループ名）と銘柄間の影響の強さを表す情報として“影響度”を加えています。

トピック辞書とその活用例を図1に示します。影響度が大きいほど、そのト

ピックのニュースが銘柄の出来高に与える影響が大きいことを示しています。影響度は、日々のニュース配信件数と銘柄の出来高などトピックと銘柄に関する数値情報から評価して更新します。こうすることで、トピックのニュース配信内容の変化や銘柄との関係の変化などの社会状況変化に柔軟に対応できます。

影響度の評価は、具体的には、トピックと銘柄の各組合せで、次の二つの指標の直前5営業日と当日を比べ、増減の方向や変化率が一致するか否かを判定しました。

- 補正ニュース件数（=（該当トピックのニュース配信件数）×（影響度））
 - 東芝の出来高
- 影響度更新は、補正ニュース件数と出来高で増減の方向と変化率が一致する方向に学習するように、ルールを設定しました。

トピック辞書の評価

インターネット配信ニュースを用い、