

# グラフ情報と協調するテキスト要約 Text Summarization Synchronizing with 2-D Chart Information

渡邊 千明<sup>†</sup>  
Chiaki Watanabe

小林 一郎<sup>‡</sup>  
Ichiro Kobayashi

## 1. 研究背景と目的

インターネットが普及するにつれ、インターネット上の膨大な情報を利用できる人、そうでない人の格差であるデジタルデバイドという社会現象が起きている。この要因の一つとして考えられるのが、インターネットから得られる情報の内容や表示が難解であり、情報を提供する側において、ユーザが欲しい情報を欲しい形で提供するなどの工夫がなされていないためである。そのため本研究では、このような現状を踏まえ、情報の内容や表示を誰にでも理解されやすいよう動的に変化させる、知的コンテンツの開発を試みる。その具体例として、テキストとグラフという異なるメディア同士を協調させることにより、大まかな情報を必要とするユーザ、詳細な情報を必要とするユーザなど、それぞれのユーザに適した情報を提示させることを目的とする。

## 2. 対象コンテンツ

本研究では、日経平均株価の動向を示すテキスト情報とグラフ情報を対象とする。テキストデータとして、国立情報学研究所の主催で実施されている評価型ワークショップ (NTCIR) のひとつである「動向情報の要約と可視化に関するワークショップ」[3] で提供されている MuST コーパスを利用している。MuST コーパスとは、1998 年から 1999 年の 2 年間分の毎日新聞を基に、ガソリン価格やパソコン出荷状況など 20 トピックについて時系列になっている記事を収集し、各トピックにつき 3 つ前後の統計量を選んで、これらの統計量の可視化に必要な要素に対して、XML 文書の形式で、人手でタグを付与したものである。タグの例として、指定された統計量や出来事に言及している部分に付与されている unit タグ、「10日」「今月」「昨年」等の時刻の表現に付与されている date タグ、統計量の値そのものではないが、その値の差や比に付与されている rel タグなどがある。

## 3. 提案手法

本研究の準備として、日経平均株価の数値情報と、MuST コーパスによって得られるその日の株価の動向情報を対応させ、グラフとテキストを関連付ける。MuST コーパスと日経平均株価の数値情報の対応の仕方を図 1 に示す。

### 3.1 処理概要

表示されたグラフから、その一部を取り出したり、目盛りの間隔を変更したりすることで、変更されたグラフの詳細度、範囲に対応して、ニュース本文が要約され、グラフの表示とテキストが協調した情報提示を行っている。

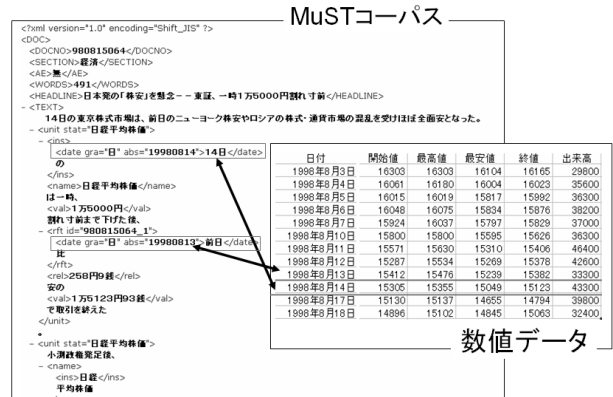


図 1: MuST コーパスと日経平均株価の数値情報

る。生成される要約文は、元となる MuST コーパスの本文の 1 文ごとに情報としての重要度を決めておき、変更されたグラフの状態に対応した重要度の高い文を抽出することで生成される。システムの概要を図 2 に示す。

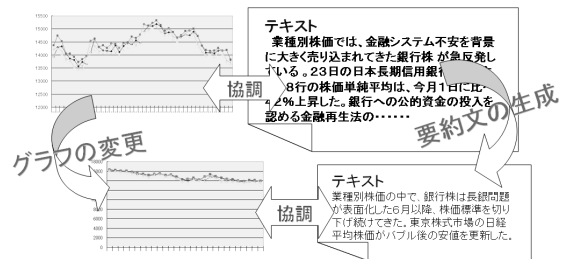


図 2: システムの概要

### 3.2 システム構成図

数値データから必要な部分を取り出し、グラフを表示させる。MuST コーパスの方も同様に、グラフの詳細度に対応して記事をサンプリングし、重要度の高い文を抽出させ要約文を表示する。これにより表示されるグラフとテキストの協調を実現している。

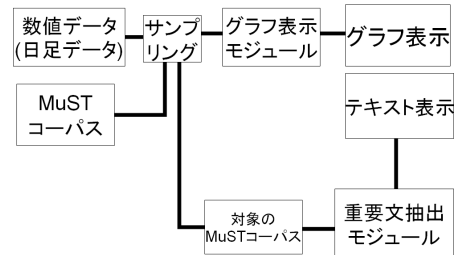


図 3: システム構成図

<sup>†</sup>お茶の水女子大学理学部情報科学科小林研究室, Dept. of Information Sciences, Faculty of Science, Ochanomizu University

<sup>‡</sup>お茶の水女子大学理学部情報科学科, Dept. of Information Sciences, Faculty of Science, Ochanomizu University

### 3.3 グラフの目盛り間隔を変更

グラフが変更され、2日おき、4日おきのように目盛りの間隔が広がった場合、2日ごと、4日ごとのように、重要文を抽出してテキストをまとめる。さらに、それぞれから抽出されたテキストから、さらに重要度の高い文を抽出する。

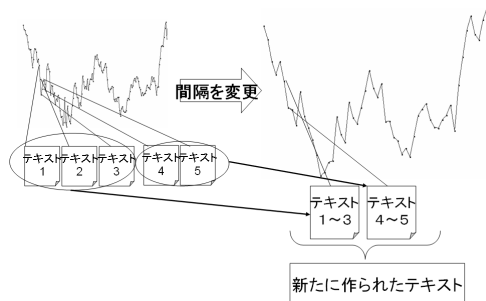


図 4: グラフの目盛り間隔の変更

この処理により、抽出されるテキストは減り、動向を示す情報の詳細度が低くなる。また、間隔が狭くなれば抽出されるテキストが増え、詳細度が高くなる。システムの実行例を図 5 に示す。

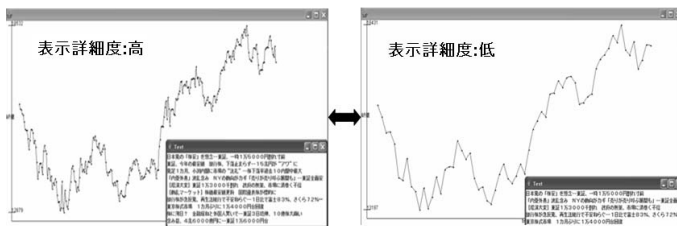


図 5: 実行例 (グラフの目盛り間隔の変更)

### 3.4 範囲を選択

グラフの一部分が選択された場合、選択された範囲にある部分のテキストから重要度の高い文を抽出する。このとき、抽出する文の長さを決めておく。

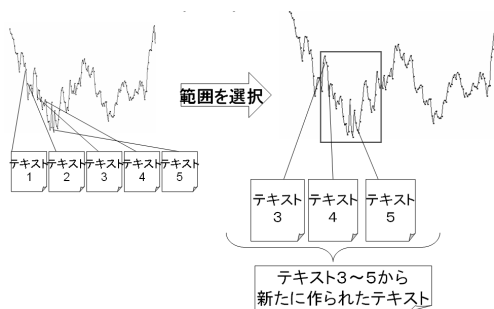


図 6: 特定箇所の情報抽出

この処理により、テキストも選択した部分を焦点とした内容となる。範囲が狭くなればなるほど、変更する前には抽出されなかった重要度の低い文も抽出されるよう

になり、その範囲のみをより詳しく説明したテキストとなる。システムの実行例を図 7 に示す。

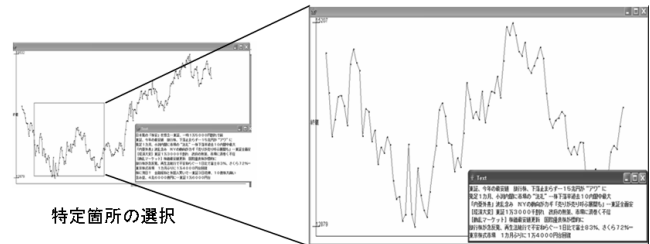


図 7: 実行例 (特定箇所の情報抽出)

### 3.5 要約対象となる文の重要度の決定方法

各文の重要度を決定する方法として、MuST コーパスに用意されているタグを参考にすることが挙げられる。unit タグは、具体的にグラフの挙動(数値情報が得られるところ)が記載されている文に付与されているので、その文は重要度が高いとする。unit タグが付与されている文の中でも、「前日比」というように、短い期間の挙動について言及しているものから、「3年前と比べて」というように、長い期間の言及している文もある。日付が付与されている date タグを参考にし、同じ unit タグが付与されている文の中でも、より長い期間の挙動について言及している文を重要度がより高いとする。また、見出しに付与されている HEADLINE タグを参考にし、見出しで取り上げられている話題に言及している文を重要とする。

## 4. まとめ

本研究では、異なるモダリティが協調することにより情報を効果的に提示する技術開発の一環として、グラフとテキストの異なる 2 つのモダリティ情報を用いた、グラフの表示状態に対応したテキストの表示を行った。これにより、ユーザにグラフを変更させ、ユーザの情報閲覧の焦点を判断し、その要望に対応した提示方法を提案した。今後の課題として、MuST コーパスで用意されているタグが付与されていない文の中で、重要度が高いとする文には、新たにタグを追加し重要度を判断する基準とする等、グラフとテキストの情報がより協調する仕組みを工夫し、提示方法を自由に变化させることができるコンテンツの開発を進める。

### 参考文献

- [1] サイバーアシストコンソーシアムシンポジウム 2005「人間中心の情報社会」の実現へ向けて セマンティックコンピューティングの応用展開, Oct 26 2005.
- [2] 長尾確: アノテーションに基づくデジタルコンテンツの高度利用, IPSJ Magazine Vol.42 No.7 July 2001
- [3] 動向情報の要約と可視化に関するワークショップ, [http://www.kecl.ntt.co.jp/scl/workshop/must/must\\_index.html](http://www.kecl.ntt.co.jp/scl/workshop/must/must_index.html)
- [4] 松下光範, 加藤恒昭: 動向情報に基づく情報可視化の基礎検討, 第 19 回人工知能学会全国大会予稿集, 1E3-03, 2005.