

# 対象変化の再構成による設計支援

\*田中克明，堀浩一

東京大学 先端科学技術研究センター 知能工学研究室

〒 153-8904 東京都目黒区駒場 4-6-1

mycom2006@katsuaki-tanaka.net

**Abstract:** 複雑なモノは，多くの時間をかけてつくられた，結果である．この結果から，使用された知識を得ることは，むずかしい．しかし，設計の過程を時間を追って記録することにより，設計者が，なにを，どのように変化させたか，を検出することはできる．変化そのものは知識ではないが，変化を再現することにより，さも知識を用いたかのような結果を，つくり出すことができるはずである．以上の考えに基づき，設計過程で作成された文書から変化を抽出・再構成するシステムの検討を行った．

## 1 はじめに

複雑なモノは，多くの時間をかけて設計された，結果である．この「結果」と付随する多くの情報を管理するために，さまざまな情報管理，獲得，構造化技術が研究され [5]，情報の共有が試みられている．例えば，設計時に記述された文書のクラスタリングを行うことにより，どのような情報が設計とともに扱われたか，を把握することができる．しかしながら，なぜそのような情報を使うに至ったのか，つまり，そのために使用された知識を，設計の結果から得ることは，むずかしい．

一方，設計の過程を時間を追って記録することにより，設計者が，なにを，どのように変化させたか，を検出することはできる．変化そのものは知識ではないが，変化を再現することにより，さも知識を用いたかのような結果を，つくり出すことができるはずである．この際に，変化の組合せを，適用対象・文脈にあわせて再構成することにより，記録した変化からつくり出される「再現」の幅を広げ，よりそれらしい結果を，生み出すことが可能となる．

以上の考えに基づき，設計過程で作成・収集した文書を，分析して変化を抽出し，これらの変化を再構成するシステムの検討を行っている．

時間経過にともなう情報の変化を取り扱う研究としては，新聞記事などを対象として話題の大まかな遷移の可視化を行う ThemeRiver [6]，時系列に沿ったニュース記事から話題の抽出と分類を行う topic detection and tracking (TDT) [4]，同じくニュースなどから統計的な情報の通時的変化の抽出と可視化を行う研究 [1] などが行われている．これに対し，本研究では，専門的なシステム設計過程の議事録から，計算機が，特別な背景知識を用いずに，対象に関する情報構造とその変化の候補を取り出し，専門知識を持った人間が，それらを扱う仕組みを目指す．

本稿では，東京大学大学院工学系研究科航空宇宙工学専攻中須賀研究室において行われている，超小型衛星 CubeSat “XI-IV”<sup>1</sup> の設計過程を対象とし，設計過程において作成された議事録を情報源として用いる．議事録は，1999年11月30日から2002年12月3日までの397文書，文書の平均長は1350文字である．各議事録は，会議の行われた日付，会議の題名，記録者などのヘッダ部，および議事内容からなる．ヘッダ以外の部分には統一された書式はなく，記録者により，まちまちの形式で記述されている．

---

<sup>1</sup><http://www.space.t.u-tokyo.ac.jp/cubesat/>

## 2 主題構造変化の抽出

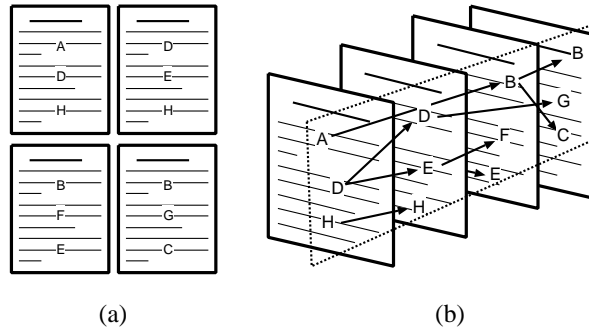


図 1: 主題抽出 (a) と主題構造変化抽出 (b)

各文書は1つ以上の主題を含んでいることが、ほとんどである。例えば、CubeSatの1つの設計議事録文書中には、全体の進捗報告の他、各設計担当ごとに課題があげられている。つまり、議事録から主題の抽出を行うためには、すなわち、文書を1つの単位とするのではなく、それ以下の細かい単位に断片化する必要がある。これらの断片を、クラスタリングにより分類すると、主題の候補を得られる(図1(a))。

クラスタリング対象とする文書断片を、全ての断片とするのではなく、文書の作成時刻を用いて、ある時点までの断片集合のクラスタリングを行うと、その時点までの主題候補を得ることができる。次に、設計がさらに進んだ時点までの断片集合を対象としてクラスタリングを行い、主題候補抽出し、以前の時点の主題候補との関連を計算すれば、時間経過にともなう主題構造変化の抽出を行うことができる(図1(b))[8]。

今回は、文書はあらかじめ定めた長さ(800バイトごと)に文書を断片化し、GETA[2]を用いてWard法[7]により、50クラスタへクラスタリングを行った。断片集合は、所属文書の作成時刻により50の集合を、各集合がそれ以前の断片を全て含むように定義した。クラスタリング後の隣接する文書断片集合に属するクラスタ間の関連度は、クラスタが共通して含む断片数を用い、 $sim(C_{n,i}, C_{m,j}) = \frac{|C_{n,i} \cap C_{m,j}|}{|C_{n,i}|}$ とした( $n$ は断片集合の出現順序、 $i$ はクラスタの番号)。

また、議事録中での言及がなくなった主題を、漸次、主題候補から外していくことを目的とし、新たな断片の追加が行われなかったクラスタに所属する文書断片について、その重みを、次のクラスタリング時に減少させる処理を行った。

## 3 主題構造変化の表示と再構成

### 3.1 俯瞰表示

類似度関数  $sim(C_{n,i}, C_{m,j})$  により、隣接した断片集合、すなわち時間上、直前直後の断片集合からつくられたクラスタ間の関係を可視化する。表示には Touch Graph LinkBrowser[9]を用いた(図2)。クラスタ  $C_{n,i}$ ,  $C_{n+1,j}$  間のリンク距離は、類似度と反比例させた。

各クラスタをグラフ上のひとつのノードとし、クラスタの特徴語と、文書集合の番号、およびクラスタに含まれる断片の数を、ノードのラベルとして表示させた。図2の矢印の向きが時間経過の方向を示す。あるノードを選択すると、ノードから、ユーザが指定したホップ数内の関連ノードが順次表示される。この動作を繰り返すことで、話題の変遷を俯瞰することができる。

### 3.2 再構成

各項目はもともと文書であるため、抽出された項目とその変化に対し、キーワードによる検索がかけられる。この検索により、関連した項目を集め、時間順に取り出すことで、変化の再構成が行える。

そこで、ユーザが指定した単語に基づいて、その単語に関連したクラスタを表示させる機能を持たせた。グラフにおいて、クラスタ間の距離以外に、単語が指定された回数以上出現するクラスタ  $C_{n,i}$  を、単語を示すノードから、クラスタが作られた断片集合の  $n$  に比例した距離を設定し、Touch Graph により表示を行う。これにより、計算機が背景知識なしに自動的に抽出した主題構造変化と、ユーザが持つ知識を組み合わせ、関連した情報の推移を選んで表示させることができる。

複数の単語を指定することにより、それらの語の間の関連を把握することも可能である。図3に、XI-VI 設計中に検討された3種類の無線機を表す語を指定し、関連する情報の推移を再構成した結果を示す。

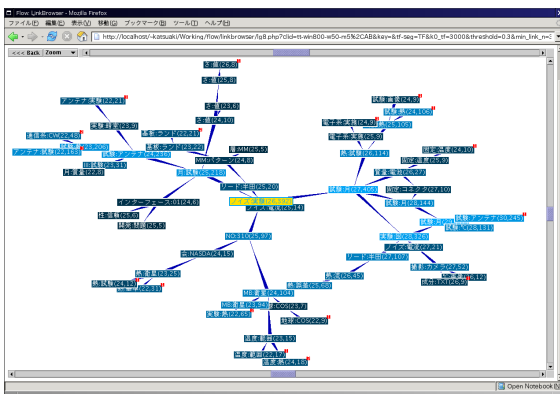


図 2: 俯瞰表示例

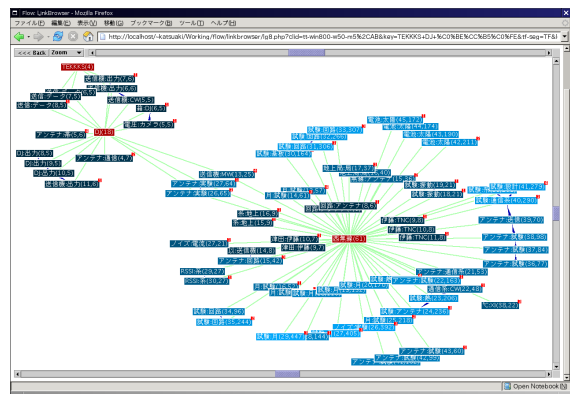


図 3: 再構成表示例

## 4 まとめ

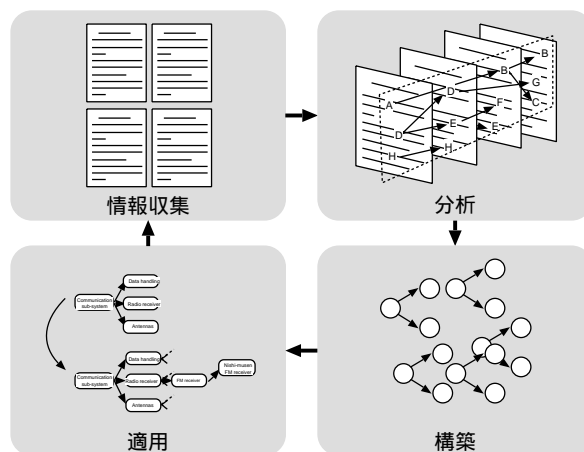


図 4: 知的処理のサイクル

設計作業、あるいは設計に限定されず、いわゆる知的処理は、情報の収集を行い、これを分析し、新たな情報を構築し、これを適用して何かを生み出す、というサイクルである(図4)。計算機が、あます

ことなく自律的に、このサイクルを成立させることは、まだ非常に難しい。

本研究では、計算機と人間が作業を分担することで、知的処理のサイクルを成立させることを目指す。そのために、超小型人工衛星の設計過程に作成された文書を対象として、特別な背景知識を持たない計算機が、文書に記述された内容の、時間経過にともなう変化を分析し、人間が持つ知識を、計算機による分析結果をぶつけることにより、変化の再構成を行い、行動に適用する、というシステムを作成した。

筆者らは、本稿で述べた、あらかじめ記述された文書群から、対象の変化を抽出・再構成するシステム以外にも、設計過程に沿って、動的に、さらに詳細な情報の変化を記録することで、計算機が背景知識に相当する情報を保持できるようにし、知的処理のサイクルにおける計算機の役目を拡大するシステムの検討も、行っている。

今後は、両システムをあわせ、本稿で述べた情報分析・構築手法の精緻化、計算機が収集する情報の拡大、精密化を行うことにより、知的サイクルにおける計算機が分担可能な範囲の拡大を、目指す予定である。

## 参考文献

- [1] 加藤恒昭, 松下光範, 平尾努: 動向情報の要約と可視化に関するワークショップの提案, 情報処理学会自然言語処理研究会, 2004-NL-164 (15), pp.89-94 (2004)
- [2] 高野明彦, 丹羽芳樹, 西岡真吾, 岩山真, 今一修, 久光徹: 汎用連想計算エンジン GETA, <http://geta.ex.nii.ac.jp/> (2002)
- [3] 田中克明, 赤石美奈, 高須淳宏, 堀浩一: 設計議事録からの主題構造変化の抽出と再構成, 人工知能学会 第 59 回人工知能基本問題研究会資料, pp. 1-6 (2005)
- [4] Allan, J.: *Topic Detection and Tracking: Event-based Information Organization*, Kluwer Academic Publishers (2002)
- [5] Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, Addison Wesley (1999)
- [6] Havre, S., Hetzler, B., and Nowell, L.: *ThemeRiver: Visualizing Theme Changes over Time*, IEEE Symposium on Information Visualization (2000)
- [7] Sharma, S.: *Applied Multivariate Techniques*, John Wiley & Sons (1996)
- [8] Tanaka, K. and Takasu, A.: *Topic Change Extraction from Problem Solving Records*, Proc. of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics, Vol. 1, pp. 377-382 (2004)
- [9] TouchGraph, <http://www.touchgraph.com/>