

区間長制約付き正規表現を用いた時間軸を持つデータ編集環境

2Q-1

北村 孝明 中山 健 前川 守

電気通信大学大学院 情報システム学研究所

1 はじめに

最近、動画や音声のように、時間軸を持つデータが大量に蓄積できるようになってきている。これらを有効に利用するには、記録されているものの出現順序や時間間隔に関する条件を様々に組み合わせて検索し、それを基に編集できる機能が必要である。それらを解決する枠組を考えられている中に、区間長制約付き正規表現 [1] がある。

区間長制約付き正規表現は、通常の記号に加えて時間長を持つ記号も導入し、これら2種類から成る記号列を対象として検索パターンを行うものである。動画や心電図など時間軸をもつデータに付与する属性をこのような記号列で表現しておけば、必要な部分を検索するために使える。しかし、区間長制約付き正規表現は、従来の正規表現に比べて複雑な表現になることが多く、インタラクティブに検索パターンを入力、修正したり、結果を確認したりできる環境が実際上不可欠である。

時間軸を持つデータにその内容を表す記号列を予め付与しておき、意図した部分の検索や編集を記号列操作として行うシステムを作成した。本システムは記号列操作の部分だけを扱い、原データから記号列を付与するまでの部分は含まない。

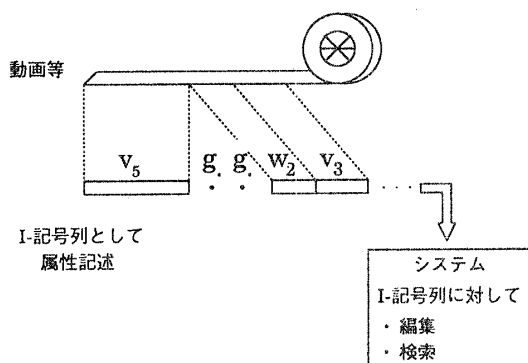


図 1: 時間軸を持つデータに付与される記号列

2 区間長制約付き正規表現

動画や心電図など時間軸のあるデータを検索する枠組み

検索対象となるデータは、予め何らかの方法で

- ある時間長をもつ区間の属性を表す記号

(例) $p_{10}, c_{3.5}$

- ある瞬間の属性を表す記号

(例) $e., m.$

という2種類の記号の列として表現されていると仮定する。

その記号列に対する検索方法として、従来の正規表現を時間長に関する制約も記述できるように拡張してある。

2.1 具体例: サッカー

a, b の2チームでサッカーを行う。コード α は、ボールをキープしているチームと時間を示している。

$$\alpha = a_3b_{4.5}g \cdot a_{1.8}b_{1.3}g \cdot a_4b_{8.4}a_2$$

・4分以内に2ゴール

$$U_+ \equiv (a_+|b_+|g.)$$

$$U_+^* \frac{g \cdot (a_+|b_+)^* g}{(0, 4)} U_+^*$$

3 インタラクティブな検索環境

意図した検索と区間長制約付き正規表現間のギャップや入力間違い等によって、必ずしも意図した検索結果が得られる訳ではない。よって、検索結果から戻って修正するようなインタラクティブな検索環境 (図 2) が必要である。

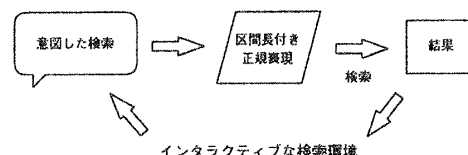


図 2: インタラクティブな検索環境

A continuous media editing environment using interval constrained regular expression.
Takaaki Kitamura, Ken Nakayama, and Mamoru Maekawa
Graduate School of Information Systems, University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan

また、従来の正規表現のように、検索パターンを記号列としてキーボードから入力するといった素朴なインターフェースでは非常に使い難い。

更に、以下のような機能が重要である。

- 検索パターンを入力補助、検索パターンの形式の正しさのチェック
- インクリメンタルな検索
- 検索結果のサマリーを表示して、意図した検索が行なわれているかどうかを確認する方法の提供
- 過去に入力した検索パターンの履歴の記録と修正

4 編集体系

時間軸を持つデータを編集するのに、検索と整合性を持たせた編集体系でなければならない。

- データの切り出し / 連結
- 付与されている属性の変更
- マッチングが重複部分の意図した抽出
- 複数の検索結果の組合せ

5 システム

5.1 概要

インタラクティブな検索、編集の環境構築にあたって、対象(人、状態)や制約条件を柔軟に変更でき、検索結果を確認できる環境は不可欠である。よって、記号列操作で意図した部分の検索や編集をするシステムを構築した。

このシステムにより、複数の検索結果を組み合わせることができる。順を追って、検索を組み立てることにより、複雑なパターンの組み立て(図 3)を支援する。

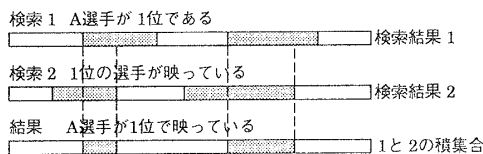


図 3: 検索結果の組合せ

本システムは Java によって実装し、記号列操作の部分だけを扱い、原データから記号列を付与するまでの部分は含まない。

時間軸を持つデータに、予め付与していた記号列に対して、時間長制約を扱うのに以下の理由よりシンプレックス法を用いた。

- 検索パターンと制約条件は、すべて線形である。
- 区間長制約付き正規表現は、すべての係数は 1 である。

5.2 システムの機能

本システムで予め付与していた記号列に対して、意図した検索パターンをどのような操作で支援するかを示す。

- 記号を組合せて検索パターンを作成する際に、Symbol ウィンドウにある記号からマウスで選ぶことができ、キーボードから入力する必要がない。
- Retrieve ウィンドウに表示されている検索パターンに対し、パターンを追加したごとに検索できるインクリメンタルサーチと検索パターンをすべて入力してから検索する機能の 2 通りを用意した。
- よく使う検索パターンをマクロとして登録することができ、入力の手間を軽減することができる。

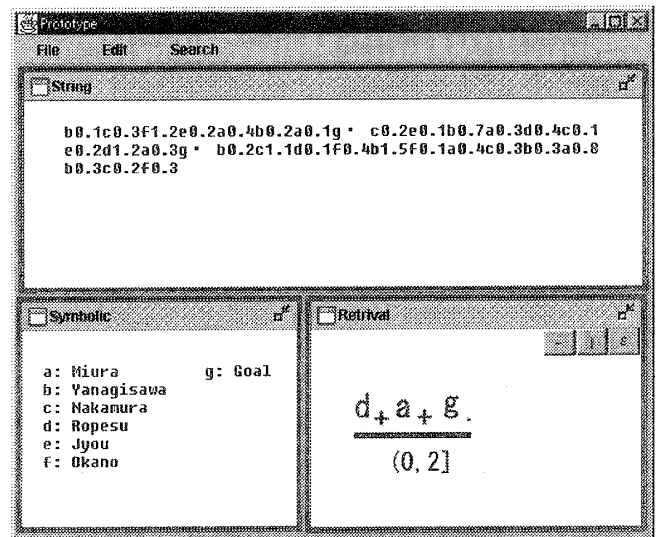


図 4: システムの実行画面

6 おわりに

現在、記号列操作のシステムの開発をしたが、検索パターンが意図通りかどうかをわかりやすくできる視覚的な環境の構築が必要である。また、区間長制約付き正規表現は検索の指定方法を与えているだけなので、動画等の切断や連結の操作も統合した編集対系を作らなければならない。

参考文献

- [1] Ken Nakayama, Kazunori Yamaguchi, and Satoru Kawai. I-regular expression: Regular expression with continuous interval constraints. In *Proceedings of the sixth international conference on Information and Knowledge management*, pp. 40-50, 1997.