符号木同定とそれに関連した同期系列の研究

情報理工学系研究科 数理情報学専攻 今泉 隆文
指導教員 合原 一幸 教授

2006 年 2 月 6 日

1 概要

符号木同定とそれに関連した同期系列の話題を取り上げ研究を行う. 符号木の同期系列の有無を判定する手法 [1] では, 符号木の各状態に対応する頂点を作り, 頂点から符号木の符号語により遷移する頂点とを矢印で結んだ, 頂点と矢印からなるグラフを作る. そして, グラフ上の全頂点から根に対応する頂点にパスが存在すれば同期系列を持つと判定する. 本研究では, グラフ上の一部の頂点から符号木の根に対応する頂点にパスがあるかどうかを調べればよいことを示す.

次に, 共通同期系列を用いた符号木同定 [5, 6] において, 情報源文字種類数が 3 のときに符号語系列長が有限である場合に誤って符号木同定してしまう確率の評価を行う. 次に, 隠れマルコフ連鎖を用いた符号木同定 [2] において符号語系列の途中から盗聴されると仮定すると, 符号木を隠れマルコフ連鎖に変換するとき, 同期系列を持つ符号木から作られる隠れマルコフ連鎖は一意に定まるが同期系列を持たない符号木の場合は複数の隠れマルコフ連鎖を考慮しなければならないことを示す. さらに, 同期系列を持たない符号木に対して, 考慮しなければならない隠れマルコフ連鎖の数を評価する.

Splay 符号は, 情報源系列を読み込み符号語系列に変換しながら適応的に符号木を更新して符号化する動的な符号であり, 暗号性の高い符号であると考えられている. 本田 [6] により Jones の Splay 符号 [4] に対して符号木同定される可能性が指摘されているが, 本研究では, 情報源文字順序を保存する Splay 符号に対して同期系列が存在し, 盗聴者に符号木同定される可能性があることを示す.

2 符号木と同期系列

情報源系列を符号化又は復号化するのに用いる符号語集合を木で表したものを符号木と呼ぶ. 図 2 の符号木は, 情報源文字 (a, b, c, d) に対して符号語 $(0, 100, 101, 11)$ を割り当てる符号木である.

符号木の内部節点を符号木の状態と定義する. ここで, 葉は根と同じ状態とする. また, 系列 w により状態 X から状態 Y に遷移するとは, 状態 X から系列 w に対応する枝を葉の向きにたどっていくと状態 Y に着くことをいう. すると, 同期系列は, ある系列 w が符号木全ての状態を同時に根の状態に遷移させる系列と定義できる. 系列 00 は図 2 の符号木の状態 A, B, C を全て同時に状態 A に遷移させるので同期系列である.

3 同期系列の有無を判定する手法の改善

まず, 符号木の 0^m 状態集合と 1^m 状態集合とを定義する.

定義 3.1　ある符号木において, 十分大きい正整数 m に対して, 根の状態から系列 0^m により遷移する全ての状態を要素とする集合を 0^m 状態集合と呼ぶ. 同様に, 系列 1^m により遷移する全ての状態を要素とする集合を 1^m 状態集合と呼ぶ.

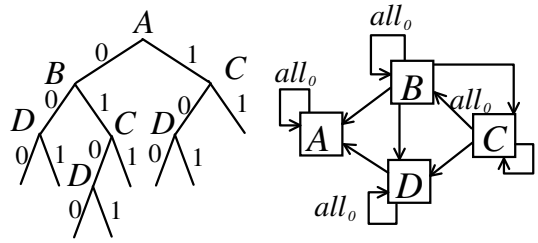

図 1　符号木とそれに対応するグラフ

1 を含まない符号語を all_0 と表す. グラフ上の全頂点から all_0 がラベル付けされた矢印を十分たどると 0^m 状態集合に対応した頂点にたどり着く. よって, グラフ上の全頂点は 0^m 状態集合に対応する頂点のどれかに矢印をたどって着ける. よって, 0^m 状態集合に対応した全頂点から根の状態に対応した頂点に矢印をたどってたどり着くことがいえれば, グラフ上の全頂点から根の状態に対応した頂点に矢印をたどってたどり着くことがいえる. 同様のことが 1^m 状態集合に対してもいえる.

以上から, 符号木の 0^m 状態集合又は 1^m 状態集合に対応するグラフ上の全頂点から符号木の根の状態に対応する頂点へ矢印をたどって着けることが言えれば, 符号木は同期系列を持つと言える.

4 符号語系列が途中から盗聴された場合の隠れマルコフ連鎖を用いた符号木同定

情報源は定常無記憶, 情報源文字の種類数は n とする. また, この仮定は盗聴者も知ってるものとする. この条件のもとで, 盗聴者が無限の長さの符号語系列全体を盗聴し情報源系列を符号化するのに用いた符号木が S か T かを同定したい場合, 符号木を隠れマルコフ連鎖に変換し, 盗聴された符号語系列の部分列の頻度と符号木から変換された隠れマルコフ連鎖の出力系列の部分列の頻度とを比較することにより符号木同定を行う [2].

本研究では, 符号語系列が途中から盗聴されると仮定した. すると, 符号語系列の初めの隠れ状態をどう定めるか曖昧さが生じ

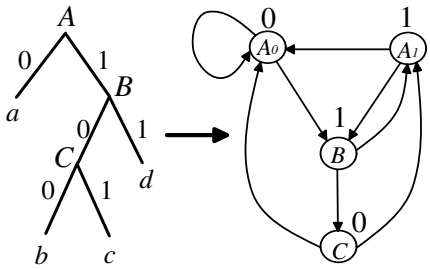

図 2　符号木とそれに対応する隠れマルコフ連鎖

る. 符号木に同期系列が存在する場合は, 同期系列は符号木の全ての状態を根の状態に遷移させるので, 同期系列以降の符号語系列の隠れ状態は, 符号語系列の初めの隠れ状態をどう定めようとも一つの状態に定まるので, 隠れマルコフ連鎖は一意に定まる. 符号木に同期系列が存在しない場合は複数の隠れマルコフ連鎖を考えなければならない. なお, 同期系列を持たない符号木の遷移確率行列を求めるときに必要な符号語系列の初めの隠れ状態は最短同期系列を求めるアルゴリズム [5, 6] を用いることにより求めることが出来る.

5　同期系列を持たない符号木の隠れマルコフ連鎖の候補数の評価

Honda-Yamamoto[5, 6] により同期系列を持たない符号木の十分条件が示されている. 本研究ではこれらの符号木に対して隠れマルコフ連鎖の候補数がどれだけになるか見積もった. そのうちの一つを紹介する.

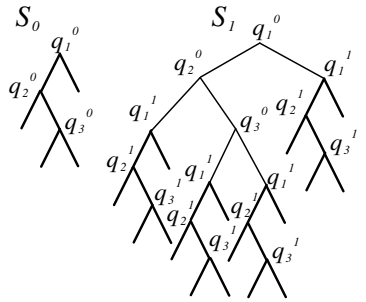

図 3　同期系列を持たない符号木の例

任意の符号木 S_0 に対して, 符号木 S_n が, 符号木 S_{n-1} の全ての葉に部分木として符号木 S_0 を接続して再帰的に構成されているものとする. このとき, 任意の n に対して, 符号木 S_n の隠れマルコフ連鎖の候補数は, 符号木 S_0 が同期系列を持つ場合, $n+1$ となる. 符号木 S_0 が同期系列を持たず, 隠れマルコフ連鎖の候補数が d 個の場合, $d(n+1)$ となる.

6　情報源文字順序を保持する Splay 符号の符号木同定

Splay 符号の同期系列を, 葉の数が等しい任意の符号木の任意の状態から, ある一つの符号木のある一つの状態に遷移させる系列と定義する. 情報源文字の順序を保存する Splay 符号 [3] では, 符号木において情報源文字に対応した葉から根までのパスに Semisplaying ステップを繰り返し適応することにより符号木を更新する.

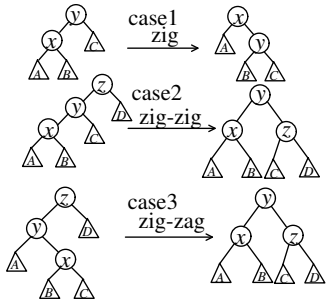

図 4　SemiSplaying ステップ

情報源系列文字種類数が 3 の場合を考えよう. このとき, 符号木の種類は図 5 の二種類であり, 状態数は 4 である. 状態遷移図を描くと図 5 のようになり同期系列が存在する. ゆえに符号木が盗聴者に同定される可能性がある. 同様にして, 情報源文字種類数が 4,5 のときも同期系列が存在し盗聴者に符号木同定される可能性があることを示した.

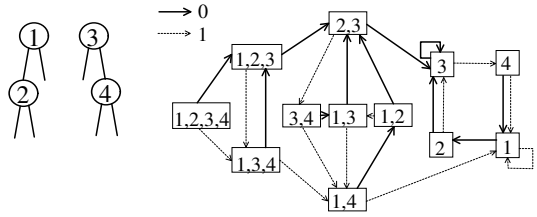

図 5　情報源文字種類数が 3 の場合の Splay 符号木とその状態遷移図

7　まとめ

同期系列の有無を判定する手法において, 0^m 状態集合が有用に働くことを示した. また, 隠れマルコフ連鎖を用いた符号木同定を, 符号語系列の途中から盗聴されるという問題設定で議論し同期系列が有用に働くことを示した. また, 同期系列を持たない符号木に対して隠れマルコフ連鎖の候補数を評価した. さらに, 情報源系列文字順序を保存する Splay 符号に対して, 符号木が特定される可能性があることを示した.

参考文献

[1] R. M. Capocelli, L. Gargano, and U. Vaccaro: IEEE Trans. Inform Theory, 34, no. 4, pp. 817–825, 1988.

[2] D. W. Gillman, M. Mohtashemi, and R. L. Rivest: IEEE Trans. Inform Theory, 42, no. 3, pp. 972–976, 1996.

[3] D. Grinberg, S. Rajagopalan, and R. Venkatesan: The 6th Annual ACM-SIAM Symposium on Discrete Algorithm, pp. 522–530, 1995.

[4] D. W. Jones: Communications of the ACM, 31, no. 8, pp. 996–1007, 1988.

[5] T. Honda and H. Yamamoto: AEW4, pp. 31–34, 2004.

[6] 本田司: 東京大学修士論文, 2005.