

時系列データによる 時制クラスの発見

法政大学 工学部 電気電子工学科

本吉 正博 渡辺浩平 三浦孝夫

産能大学経営情報学部

塩谷勇

目的

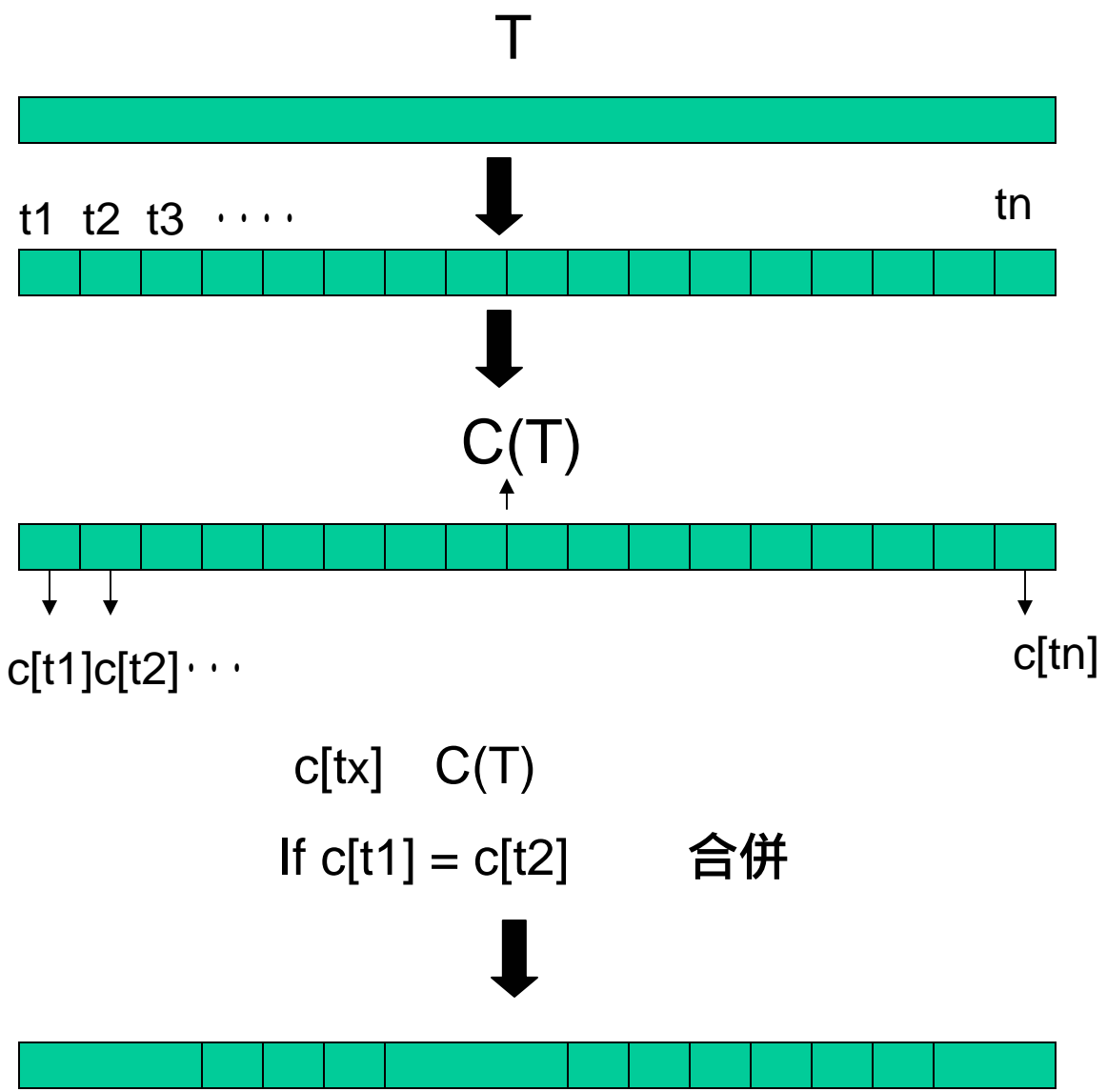
- 頻度に基づいた時制頻出クラスの抽出
- 時系列データのためのスキーマ発見
- 時系列データの
 - 離散化、集約
 - ノイズ除去

時制頻出クラス列生成の手順

1. ログ全体の時制頻出クラスを抽出
2. ログを最小単位のintervalに分割
3. 最小単位ごとの時制頻出クラスを抽出
4. 1と3の積集合を取る
5. 同一のinterval同士を合併
 - どのくらいいつながるか

時制頻出クラス列生成

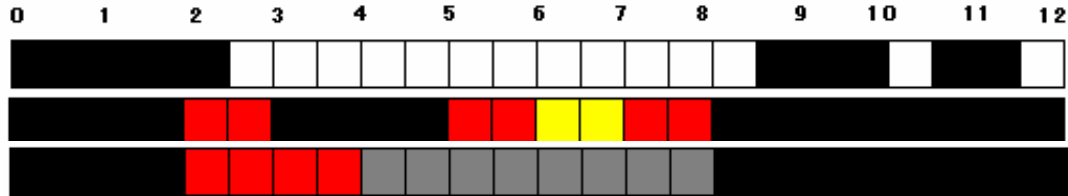
- 合併された(同じ状態が続いた)intervalにはなにかしら意味があるという仮定



WebLog

-
- [13:14:14] /job/ /job/mokuji.htm /job/info.htm /job/BK253.GIF
- [13:14:16] /job/2000/M20.htm
- [13:14:20] /job/ /job/ /job/mokuji.htm /job/info.htm /job/BK253.GIF ...
- [13:14:24] /job/suisen/index.htm
- [13:14:25] /job /job/ /job/mokuji.htm /job/info.htm /job/BK253.GIF ...
- [13:14:28] /job/suisen/index.htm
- [13:14:29] /job/suisen/april25.htm
- [13:14:30] /job/suisen/april25.htm
- [13:14:37] /job/ /job/ /job/mokuji.htm /job/info.htm /job/BK253.GIF ...
- [13:14:45] /job/ /job/mokuji.htm /job/info.htm /job/BK253.GIF ...
-

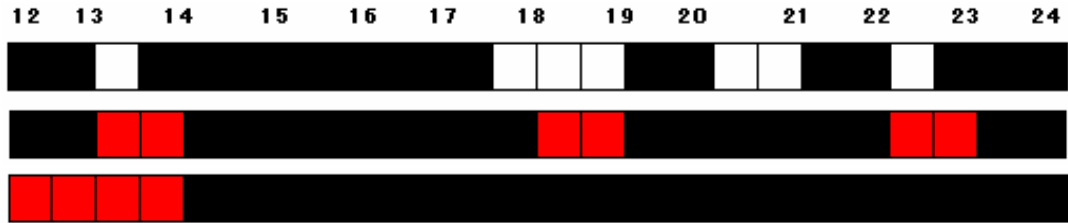
サポート10%の場合



30分

60分

120分

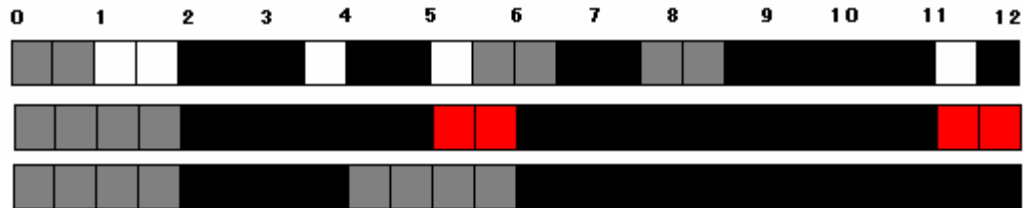


30分

60分

120分

サポート20%の場合



30分

60分

120分



30分

60分

120分

サポート30%の場合

0 1 2 3 4 5 6 7 8 9 10 11 12



30分
60分
120分

12 13 14 15 16 17 18 19 20 21 22 23 24



30分
60分
120分

考察

- 実際の学生の行動を反映する結果
 - 昼食、夕食、入浴時間を除く日中、夜が活発
 - 10%,20%と30%は大きく異なる結果
- 時制頻出クラス列を獲得
- サポートを低くすればより小さい頻出クラスが生成　ノイズを拾う