Many phenomena happen in predictable cycles:
- CPU clock, presidential elections, moon cycle

**Periodicity**: tendency for events to occur in a cyclic pattern
- Observations have a regular time interval.
- *Period* measures length of a cycle or interval.
- In time-series data we want to mine for trends:
  - Periodic patterns can be used to forecast future occurrences
  - Notice anomalies when periodic event is skipped
- Connection to mining frequent sequence patterns:
  - Partition input sequence by period length, starting from first occurrence.
- Many Applications
  - Improved recommendation based on habits
  - Biomedical – detecting telomere decay
**What is Periodic Pattern**

- **Periodic Pattern**: subsequence that repeats, following a specific or approximate frequency.

- **Notation**: pattern expressed as tuple of events from the sequence.
  - * event means “don’t care”

- **Sub Pattern**: is when one or more events in the cycle is replaced with *

- **Super Pattern**: a * event is replaced with an exact event.

- **Frequency Count**: Number of instances where the periodic pattern is found in the data.

- **Confidence**:

  \[
  \text{conf}(s) = \frac{\text{frequency count}(s)}{m}
  \]

  - \(m\) is the max frequency count = length(pattern)/length(sequence)
SUBTLETIES IN PERIODIC PATTERNS

- Periodic pattern does not always define the behavior taken
  - Scenario 1. There is one pattern: “put on clothes at 8am each day”.
    - Behavior is different depending on season (t-shirt or jacket)
    - Events are abstract (without defining concrete behavior)
  - Scenario 2. There are two patterns:
    - *Put on jacket at 8am each day*. Pattern applies for October-April
    - *Put on t-shirt at 8am each day*. Pattern applies for May-September
    - Events are concrete (behavior defined explicitly)
    - Some periodic patterns might not be active continuously
  - Observed behavior in either scenario is the same
  - Abstract events or concrete events depends on the contents of input
- A data sequence often contains multiple interleaved periodic patterns
  - Pattern 1 : *Wear blue jacket every Wednesday*
  - Pattern 2 : *Wear blue jacket every 3 days*
  - Typical solution is to selectively consider sequence elements for pattern matching
TYPES OF PERIODIC PATTERNS

**Full periodic pattern** : specifies behavior at every step in the period. Pattern tuple does not contain any ‘*’ events

Pattern: (A,B,C,D) → ABCD ABCD ABCD....

Ex: Every day contributes to year cycle. (day1, day2 ... day365)

**Partial periodic pattern** : specifies behavior at some points in time and not others. Pattern does contain unspecified events.

Pattern: (A,B,C,*) → ABCR ABCV ABCF....

Ex: Every Tuesday, CS512 students sit in lecture room from 9:30 to 10:45.
MINING FULL PERIODIC PATTERNS

- Fast Fourier Transform: Find pattern’s frequency using DFT
  - Fourier Transform: change sequence from time to frequency domain
  - Continuous, periodic functions can be written as linear combination of sines and cosines
  - DFT: Discrete Fourier Transform:
    - Takes samples from target function and finds a linear combination of sinusoid functions which match the sample point.

- Autocorrelation:
  - Find correlations between observation at time t and time t+k where k is some time shift
  - Use Pearson’s product moment correlation to find correlation between observed vector
    \[ v_1[i] = o_i \quad 1 \leq i \leq n - k \]
    and lagged vector
    \[ v_2[i] = o_{i+k} \]
• Autocorrelation function is \( \frac{1}{t-\tau} \int_{0}^{t-\tau} x(t)x(t+\tau)dt \) for time lag \( \tau \)
AUTOCORRELATION FUNCTION

- Autocorrelation function is \( \frac{1}{t - \tau} \int_0^{t-\tau} x(t)x(t + \tau)dt \) for time lag \( \tau \)
**AUTOCORRELATION FUNCTION**

- Autocorrelation function is
  \[
  \frac{1}{t - \tau} \int_{0}^{t-\tau} x(t)x(t + \tau)dt
  \]
  for time lag \( \tau \).

- Common technique is to perform FFT on autocorrelation curve.

- Result gives **prime periods**:
  - If \((a,b)\) is pattern with period \(t\), \((a,b,a,b)\) is a redundant pattern with period \(2t\).
  - Prime period is the smallest \(t\) which captures all redundant patterns.

**Phase aligned**

**Phase misaligned**

**Phase approximately aligned**
FREQUENT PARTIAL PATTERNS

• Exploiting Apriori Property
  • All sub-patterns of frequent pattern are also frequent
  • Candidate set shrinks slowly:
    
    | X | X | X | X | X | X |
    |---|---|---|---|---|---|
    | Y | Y | Y | Y | Y |   |

    \[ X = \text{Confidence of } (A, *) \]
    \[ Y = \text{Confidence of } (*, B) \]

    Confidence of \((A, B)\) is high if \(X, Y\) is high (pigeonhole)

• Candidate Max-Patterns [1]
  • Max-Pattern is compressed representation of frequent 1-patterns
  • All partial periodic patterns can be found by finding a set of frequent sub-patterns of Max-Pattern
  • Max-subpattern tree captures frequency of subpatterns, can find frequent partial periodic patterns.

\[ (A, *, *) \]
\[ (B, *, *) \]
\[ (*, C, *) \]
\[ \{A, B, C, *\} \]

Max-Pattern composed from three 1-patterns

• One scan to find 1-patterns and one scan to build the tree.
**VARYING PERIOD PATTERNS**

*Synchronous Pattern*: treat period as a stable value, not flexible.

- pattern occurs at a regular, predictable period interval
- *Ex*: Scheduling jobs on a computer to execute at precise times each day

*Asynchronous Pattern*: Period can misaligned from insertion of unexpected or noisy event.

- Pattern length must be flexible to account for unexpected events.
- Most students don’t show up exactly at 9:30, [9:20, 9:35]
- A student who shows up exactly at 9:30 might be late if they can’t find their pants.
FINDING THE PERIOD

- Lots of possible periods. Don’t want to try them all.

- **Distance-based pruning** [2]:
  - For periodic events, relative distance between occurrences is stable.
  - For each symbol occurrence, find distance to previous occurrences.
  - Filter out pairs of (symbol, distance) by threshold support value
  - Relationship to autocorrelation: distances with high support will act as good lag values for autocorrelation

- Remove Noise: **Dynamic Time Warping**
  - Noisy insertions cause a period to “stretch”
  - Solution: DP to find matching between two time series to minimize total distance.

---

Diagram illustrating the concept of dynamic time warping.
PERIODICITY FROM INCOMPLETE DATA

• Problem: incomplete data does not include all periodic occurrences

• Main idea: if good period value is selected, the occurrences of periodic events form clusters.[3]

• Measure of period quality:

\[ \gamma^+_\mathcal{X}(T) = \max_{I \in \mathcal{I}_T} \Delta^+_\mathcal{X}(I, T). \]

- \( T \) is candidate period = set of time ranges
- \( I \) is candidate pattern = some subset of \( T \)

• For each \( T \), pick candidate pattern to maximize probability measure:

\[ \Delta^+_\mathcal{X}(I, T) = \mu^+_\mathcal{X}(I, T) - \frac{|I|}{T}, \]

- Include time ranges where \( P(\text{event}) > P(\text{no event}) \)

• Return the \( T \) that maximizes the probability measure
PERIODIC BEHAVIOR FOR MOVING OBJECT

• Moving objects: cars and people (and more)!
  • Where will I be tomorrow?
• Moving objects have a path and a destination.
  • Many paths go to one destination, often may change
  • Destination is not easy to substitute
  • Reference point: place where object spends a lot of time
    • Basically means destination
• Objects visiting reference points can have periodic behavior
• Finding periodic behavior at each reference point:
  • Log timestamps for reference point visits
  • Find periodicity using autocorrelation and FFT
• Finding periodic patterns given period T:
  • Consider all reference points with period T
  • For sub-intervals in T and reference point o, calculate probability of object visiting o at each subinterval
Sources


[3] Li, Zhenhui, Jingjing Wang, and Jiawei Han. "Mining event periodicity from incomplete observations." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.