Time-Warping Life Courses

Brendan Halpin
Department of Sociology
University of Limerick
brendan.halpin@ul.ie[1]

Sept 2 2008, RC33 Naples

sociology
UNIVERSITY OF LIMERICK

[1] Work in progress! See http://teaching.sociology.ul.ie/seqanal/naplestw.pdf

Appropriate algorithms for life-course data

- Sequence analysis is increasingly popular in sociology
  - For many sorts of data, including "life-course"
  - Nearly always uses optimal matching algorithm
- The OM algorithm works naturally with discrete-time/discrete-state sequences
- Life-course is usually continuous-time, discrete-state spell data
- Is OM valid for such data?
- This paper discusses "time warping" as an alternative algorithm

sociology
UNIVERSITY OF LIMERICK

Tokens and spells

- OM operates on strings that are sequences of tokens, e.g.,

$$\delta(\mathtt{ABCD}, \mathtt{ABBD}) = d(\mathtt{C}, \mathtt{B})$$

- But life-course data is best considered sequences of spells
  - Usual strategy: respresent spells as runs of tokens, as for many analyses
- But does OM do justice to this?
  - For instance, OM says **AAAB** is as distant from **AACB** as from **AABB** (given $\mathtt{sub}_{A,B} = \mathtt{sub}_{A,C}$)
  - Substantively, the first and third are very similar while the second introduces a completely new spell

Alternatives?

- Alternatives to OM exist
- Elzinga's X/t method counts common subsequences of spells, weighted by their duration (Elzinga, 2003, 2005)
- In previous work (Halpin, 2008) I proposed a variant on OM, "OMv"
  - Cost of elementary operations depends on the length of the spell
  - Comparable speed to OM
  - Results depend on amount of variability in sequences

sociology
UNIVERSITY OF LIMERICK

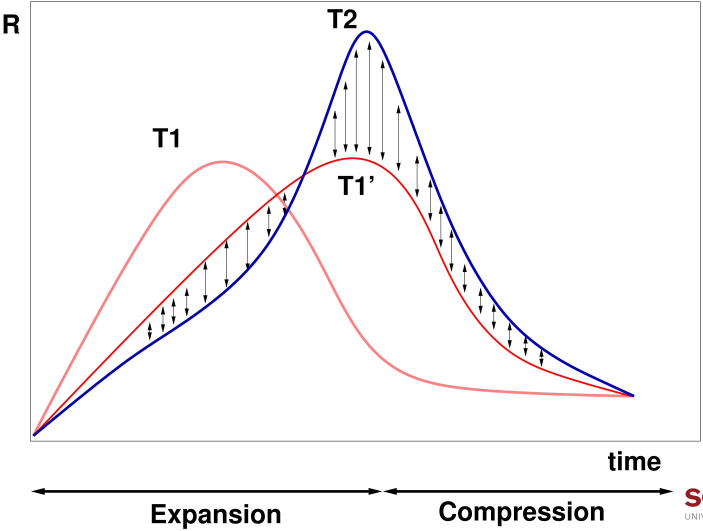
Warping time

- OMv “warps time” by weighting it differently in different spells
- Harks back to Abbott’s use of the term to suggest non-linear time scales (Abbott and Hrycak, 1990)
- In turn informed by Sankoff and Kruskal (1983), *Time Warps, String Edits and Macromolecules*

Time warping algorithms

- Formally, time warping is a family of algorithms that do "continuous time-series to time-series correction" while OM *et al* do "string to string correction" (Marteau, 2007)
- Focus on comparing pairs of continuous-time high-dimensional time-series in $\mathbf{R}^n$
- Operates by locally compressing or expanding the time scale of one trajectory to minimise the distance to the other
- Distance is usually Euclidean in $\mathbf{R}^n$ or other simple distance

1-dimensional time-warping



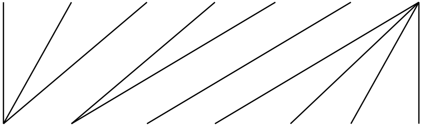
TW algorithms

- TW used widely: was used for speech recognition, signature verification, other machine learning tasks
- Typically used to match a high-dimensional time-series to a "dictionary" of standard elements
- Conceptually it is a continuous time approach but implementations must be discrete – sampling or periodic summaries:
  - e.g., sound at 44 kHz
  - rainfall daily
  - employment history monthly
- Kruskal and Liberman (1983) show that the continuous time logic can be faithfully implemented with discretised series

sociology
UNIVERSITY OF LIMERICK

Discrete time-warping

AAABBCC



ABCCCCC

TW for general SA?

- Sociological sequence analysis often uses cluster analysis to generate data-driven typologies
  - Also multi-dimensional scaling
  - Also comparing sequences to "typical" sequences
- For cluster and MDS, distance needs to be a "metric":

  1. $d(R, S) = 0 \implies R = S$
  2. $d(R, S) \geq 0$
  3. $d(R, S) = d(S, R)$
  4. $d(R, T) \leq d(R, S) + d(S, T)$ (triangle inequality)

- Conventional TW satisfies 1-3, not 4, thus usually limited to matching against a "dictionary"

sociology UNIVERSITY OF LIMERICK

TW with stiffness penalty

- Violation of the triangle inequality is due to TW usually having no cost to expansion or compression, only to the residual point-by-point distance
- Marteau (2007, 2008) proposes a TW algorithm that has a "stiffness" penalty
- Satisfies the triangle inequality
- Can be programmed very similarly to OM (recursive algorithm)
- Stiffness penalty like but not like *indel* cost – squeezing/stretching, not inserting/deleting
- Point-to-point distance just like substitution

sociology UNIVERSITY OF LIMERICK

Recursive algorithm

TW distance, $\delta(A^p, B^q) =$

$$min \begin{cases} \delta(A^{p-1}, B^q) & + d_{LP}(a_p, a_{p-1}) + \gamma d_{LP}(t_{a_p}, t_{a_{p-1}}) + \lambda \\ \delta(A^{p-1}, B^{q-1}) & + d_{LP}(a_p, b_q) \qquad + \gamma d_{LP}(t_{a_p}, t_{b_q}) \\ \delta(A^p, B^{q-1}) & + d_{LP}(b_q, b_{q-1}) + \gamma d_{LP}(t_{b_q}, t_{b_{q-1}}) + \lambda \end{cases}$$

(Marteau, 2007)

sociology
UNIVERSITY OF LIMERICK

Implementation

- Implemented as a Stata plugin
  - alongside similar implementations of OM and OMv
  - fast, comparable to OM plugin
  - but not platfrom independent
- **NB** – experimental implementation, tentative results!
- Will be made available once stable

Testing TW

- Test data consists of 5 years of monthly labour market history of females who have a birth at end of year 2
- BHPS, 675 cases, 223 excluding duplicates
- State space:

Full-time employed	■ (black)	0	1	2	3
Part-time employed	■ (brown)	1	0	1	2
Unemployed	■ (red)	2	1	0	1
Not in labour market	■ (yellow)	3	2	1	0

- *indel* cost is 2 for OM and OMv
- Range of γ "stiffness" parameters used (0.33 to 2.00)

sociology
UNIVERSITY OF LIMERICK

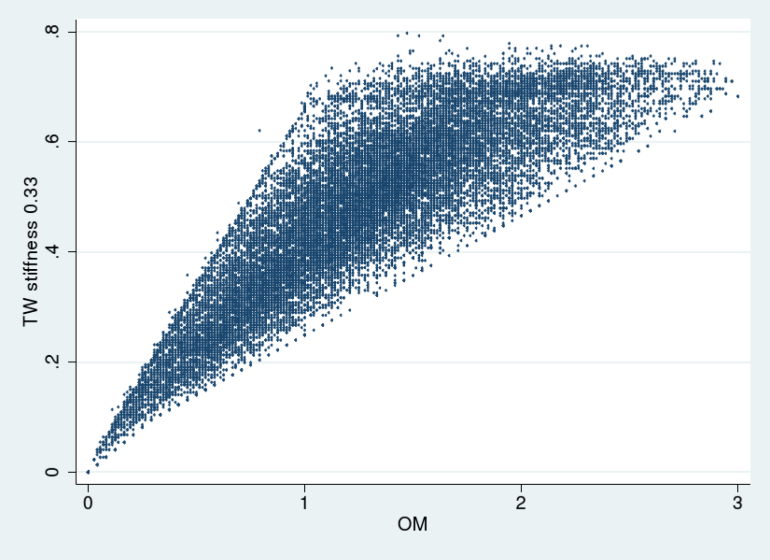
Correlations

	TW				OM	OMv	Ham.
	0.33	0.67	1.00	2.00			
TW, $\gamma = 0.33$	1.00	0.98	0.93	0.87	0.86	0.85	0.87
TW, $\gamma = 0.67$	0.98	1.00	0.98	0.93	0.92	0.88	0.93
TW, $\gamma = 1.00$	0.93	0.98	1.00	0.96	0.95	0.90	0.96
TW, $\gamma = 2.00$	0.87	0.93	0.96	1.00	0.98	0.92	1.00
OMA	0.86	0.92	0.95	0.98	1.00	0.89	0.92
OMv	0.85	0.88	0.90	0.92	0.89	1.00	0.92
Hamming	0.87	0.93	0.96	1.00	0.98	0.92	1.00

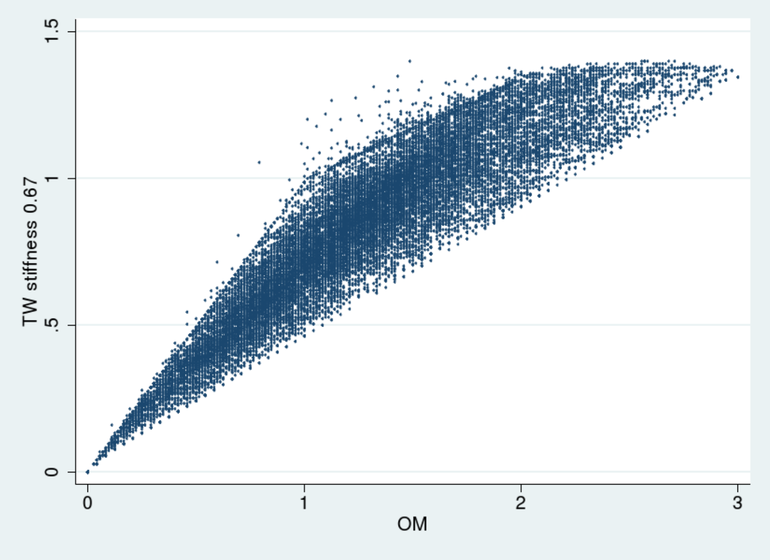
First impressions

- Generally very close in pattern to OM
- The higher γ is, the closer to OM...
- Up to a point: at $\gamma = 2$ and above, yields Hamming distance; i.e., "warping" completely suppressed
- Analogously, high *indel* costs suppress *indels* in OM...
- But at a rather higher threshold: warping and *indels* are not the same; *indel* costs and γ are not on the same scale relative to substitution costs

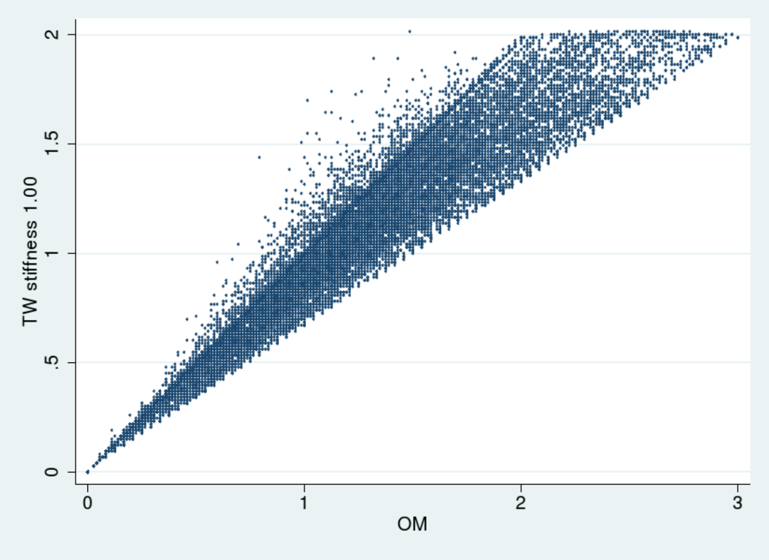
TW, stiffness 0.33 vs OM



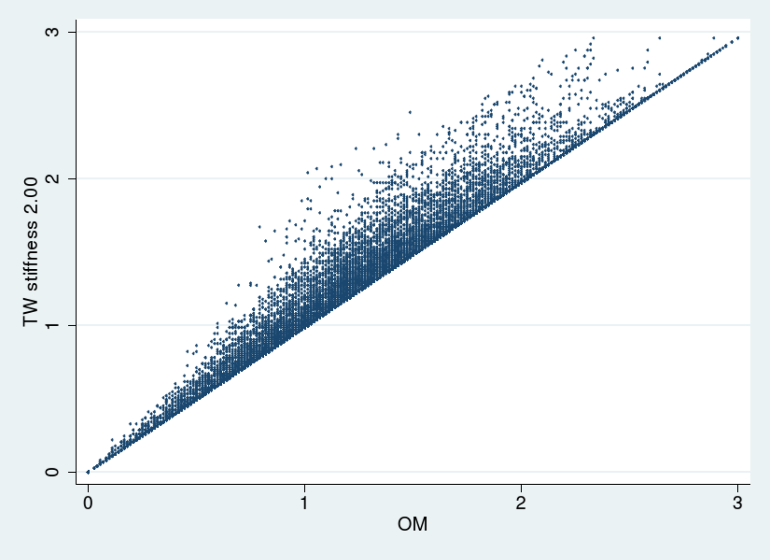
TW, stiffness 0.67 vs OM



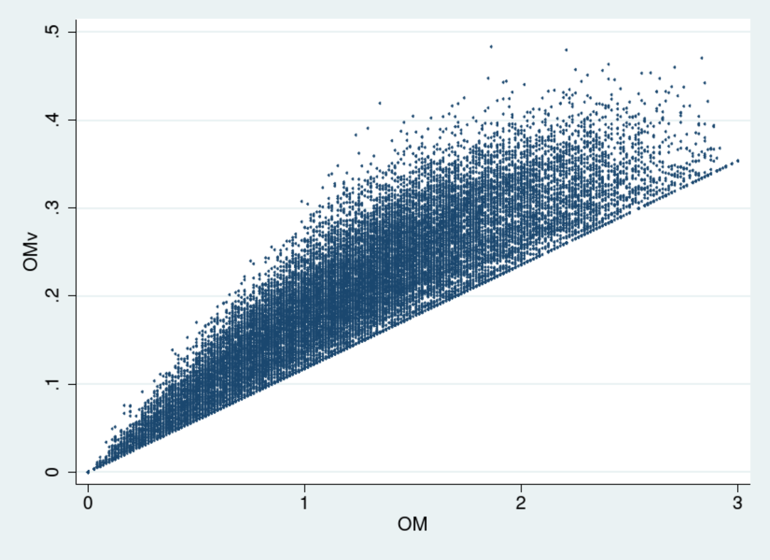
TW, stiffness 1 vs OM



TW, stiffness 2 vs OM (Hamming!)

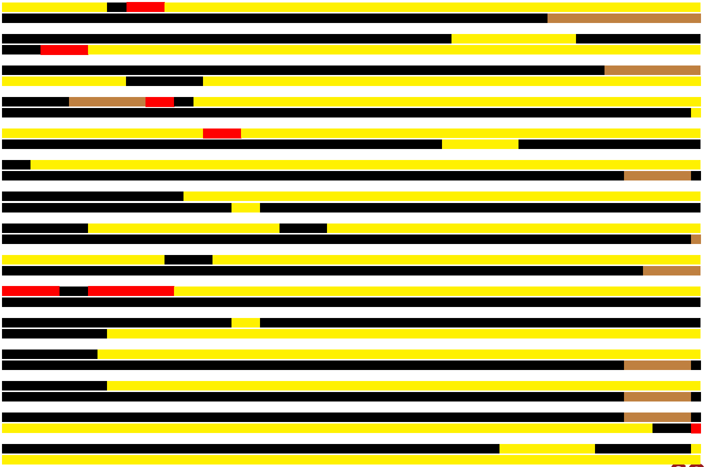


OMv vs OM



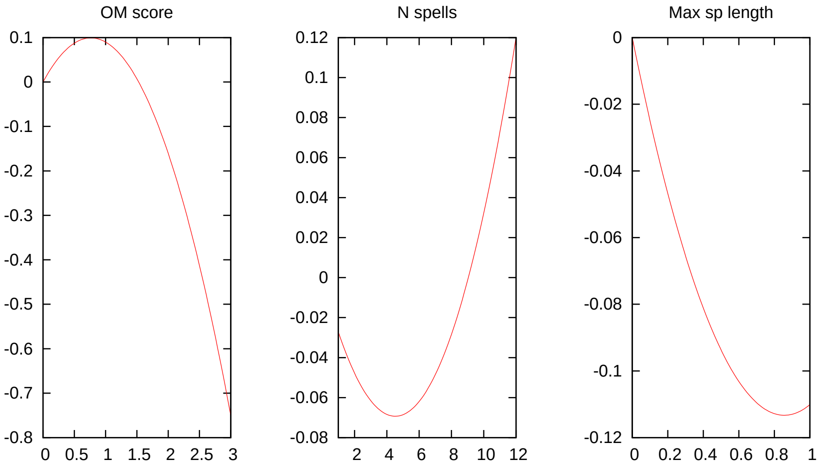
TW relatively bigger than OM

TW relatively smaller than OM



TW relatively close to OM

Regressing TW/OM difference



sociology
UNIVERSITY OF LIMERICK

Summary

- TW tends to be higher than OM distance, where there are moderate to high numbers of spells
- TW tends to be lower than OM when there are relatively few spells, in high-distance states
- Results are similar when the sequences have similar numbers of spells, usually well matching
- Also evidence that TW is greater where the number of spells is unequal

Conclusion

- TW is different, but not hugely – affinity between compression/expansion and *indels*
- A different "narrative" behind the algorithm, more suited to continuous time phenomena
- An alternative to the "edit operations" narrative
- Agreement is reassuring: OM is not very wrong!
- Intriguing in the differences – more choice in the SA toolbox!

Abbott, A. and Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology*, 96(1):144–85.

Elzinga, C. H. (2003). Sequence similarity: A non-aligning technique. *Sociological Methods and Research*, 32(1):3–29.

Elzinga, C. H. (2005). Combinatorial representations of token sequences. *Journal of Classification*, 22(1):87–118.

Halpin, B. (2008). Optimal Matching Analysis and life course data: The importance of duration. Working paper 2008-01, Dept of Sociology, University of Limerick.

Kruskal, J. B. and Liberman, M. (1983). The symmetric time-warping problem. In Sankoff and Kruskal (1983), pages 125–161.

Marteau, P.-F. (2007). Time Warp Edit Distance with Stiffness

Adjustment for Time Series Matching. *ArXiv Computer Science e-prints*.

Marteau, P.-F. (2008). Time Warp Edit Distance. *ArXiv e-prints*, 802.

Sankoff, D. and Kruskal, J. B., editors (1983). *Time Warps, String Edits and Macromolecules*. Addison-Wesley, Reading, MA.

sociology
UNIVERSITY OF LIMERICK