

Communication Technologies 2 (CT2)  
Machine Learning:  
Applications and Algorithms

# Alignment for Location Prediction

Fabian Frölich, Johann Götz, Olga Groh

Lecture in WS 2017 / 2018  
2018/03/22



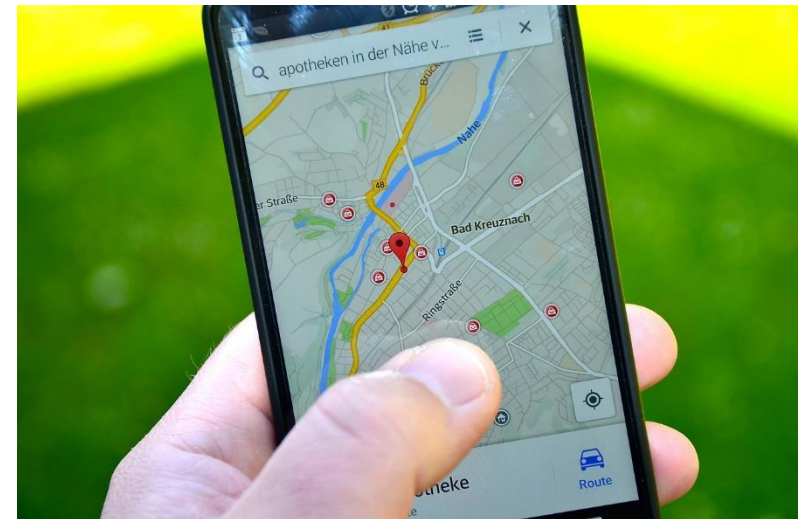
# Outline



- Introduction and Motivation
- Conception
- Implementation
- Evaluation
- Conclusion and Future Work

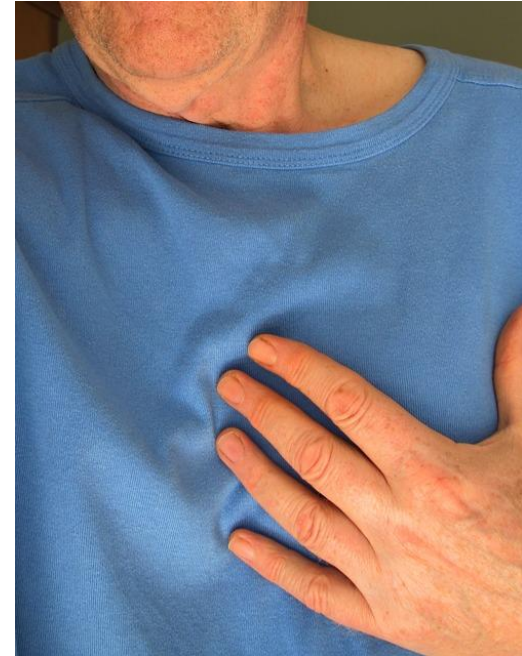
# Introduction

- **Goal:** Position prediction with alignment approach on Google Location History data
  - Working with GPS data
  - Gathered from modern mobile devices
  - Using Smith-Waterman algorithm



# Motivation

- Use cases:
  - Position prediction in case of emergency
  - Safety of elder people
  - Radiator turns on in time in smart-home environments



# Conception

## 1. Raw data:

Timestamp	Accuracy	Longitude	Latitude	Altitude	Location
1300040100	30	9.500.700	51.330.700	0	Home
1300028400	31	9.500.300	51.330.300	0	Sport
1300010400	31	9.500.500	51.330.500	0	Work

## 2. Sorted data:

Timestamp	Location
1300010400	Work
1300028400	Sport
1300040100	Home

## 3. Assigned characters:

Work → A

Sport → B

Home → C

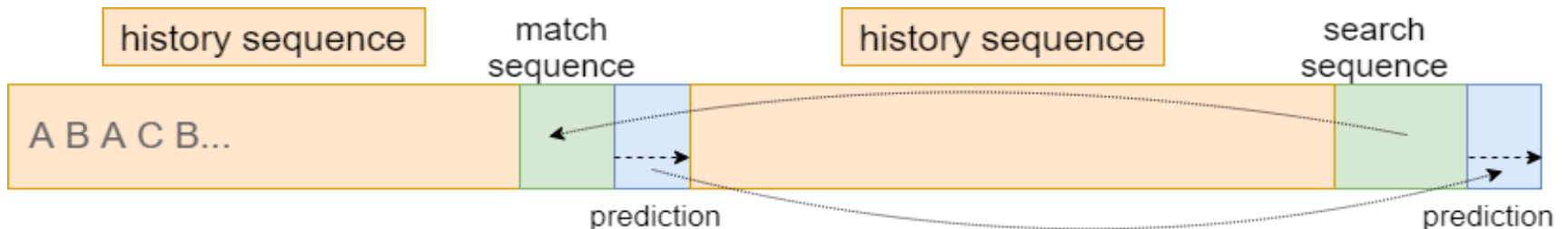
## 4. Result Sequence: ABC

- Compression:
  - Location data is gathered irregularly
  - Unchanged locations within 30 min interval are mapped to a single symbol
  - Achievement of regulated sequences

Data set	uncompressed	compressed	reduction
Sequence 1	1858	436	76,53%
Sequence 2	5499	1015	81,54%
Sequence 3	2222	581	73,85%

# Conception

- **History sequence:** The entire sequence
- **Search sequence:** Last x characters of history sequence
- **Match sequence:** Found search sequence
- **Prediction:** Continuation from match sequence



- Smith-Waterman algorithm used mainly in bioinformatics
  - Alignment algorithm invented by Temple F. Smith and Michael S. Waterman in 1981
  - Local Alignment algorithm
  - Compares evolution based differences in DNA and protein sequences [11]



# Conception – Alignment Algorithm

D(i,j)		A	C	B	A	C	C	A	B
	0	-1	-1	-1	-1	-1	-1	-1	-1
A	0	2	-1	-1	2	-1	-1	2	-1
C	-1	-1	2	-1	-1	2	2	-1	-1
A	-1	2	-1	-1	2	-1	-1	2	-1

D(i,j)		A	C	B	A	C	C	A	B
	0	-1	-1	-1	-1	-1	-1	-1	-1
A	-1	2	1	0	2	1	0	2	1
C	-1	1	4	3	2	4	3	2	1
A	-1	2	3	3	5	4	3	5	4

D(i,j)		A	C	B	A	C	C	A	B
	0	-1	-1	-1	-1	-1	-1	-1	-1
A	-1	2	-1	-1	2	-1	-1	2	-1
C	-1	-1	2	-1	-1	2	2	-1	-1
A	-1	2	-1	-1	2	-1	-1	2	-1

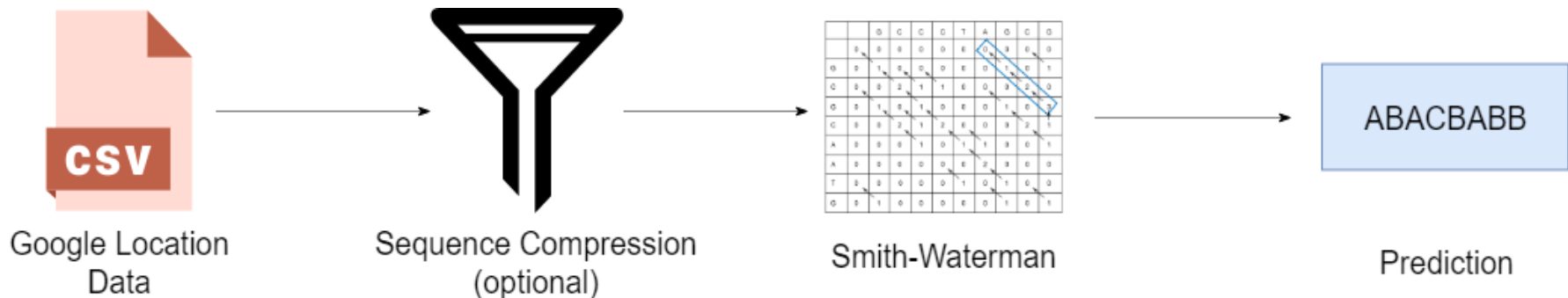
D(i,j)		A	C	B	A	C	C	A	B
	0	-1	-1	-1	-1	-1	-1	-1	-1
A	-1	2	1	0	2	1	0	2	1
C	-1	1	4	3	2	4	3	2	1
A	-1	2	3	3	5	4	3	5	4

**Result: A-CA**

**Reccurences:** Match = 2 Mismatch = -1 Gap = -1

# Implementation – Technical Facts

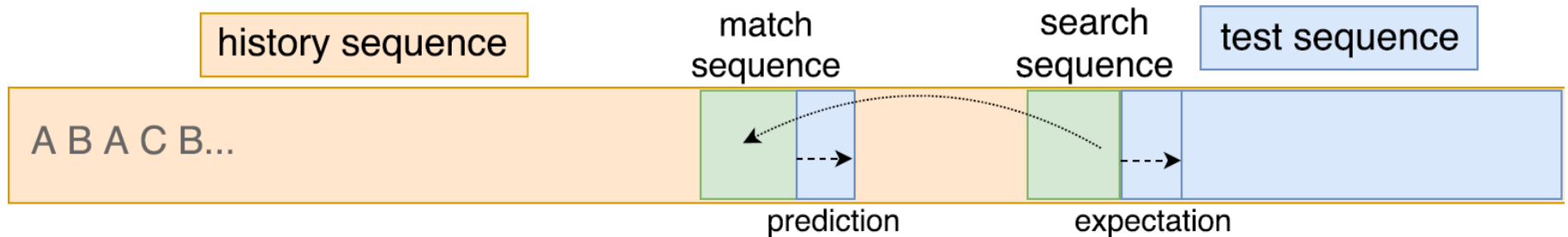
- JavaSE 1.8
- JUnit 5.0 for different parameter input



# Implementation – Output

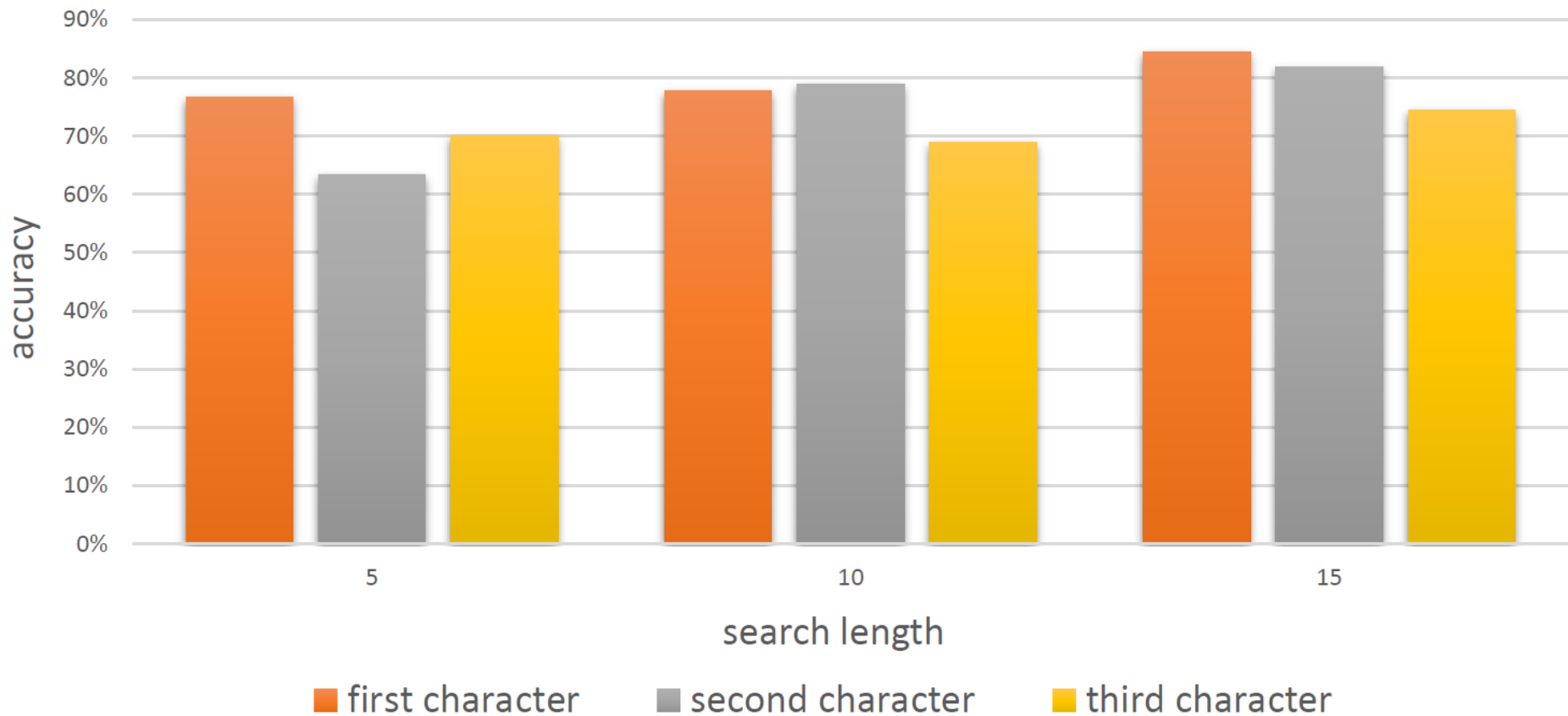
```
1 This is the following word mapping:
2
3 Friend1 --> E
4 Sport --> D
5 Work --> C
6 Shop1 --> F
7 Home --> A
8 unknown --> B
9
10 File path: google location data/Latitude_cv (2013_06_19 08_17_46 UTC).csv
11
12 =====SEQUENCE PREDICTION=====
13 Length of Data sequence: 436
14 Search sequence: ABCBC
15 In History sequence: ABCCBCBCCBCBCBCCBCBCBCBDBAAAAAAAAAAAAAAAAABABABACACAC ... AAAAAA
16
17 Traceback:
18 Found part in history sequence: ABCCBC
19 Match sequence: AB-CBC
20
21 Highest Score: 9 in row 5 and col 110
22
23 Predicted sequence of max length 3: BCB
24
25 Time needed for algorithm: 240ms
```

- Sequence testing
  - With original
  - With compressed
- Different search sequence lengths



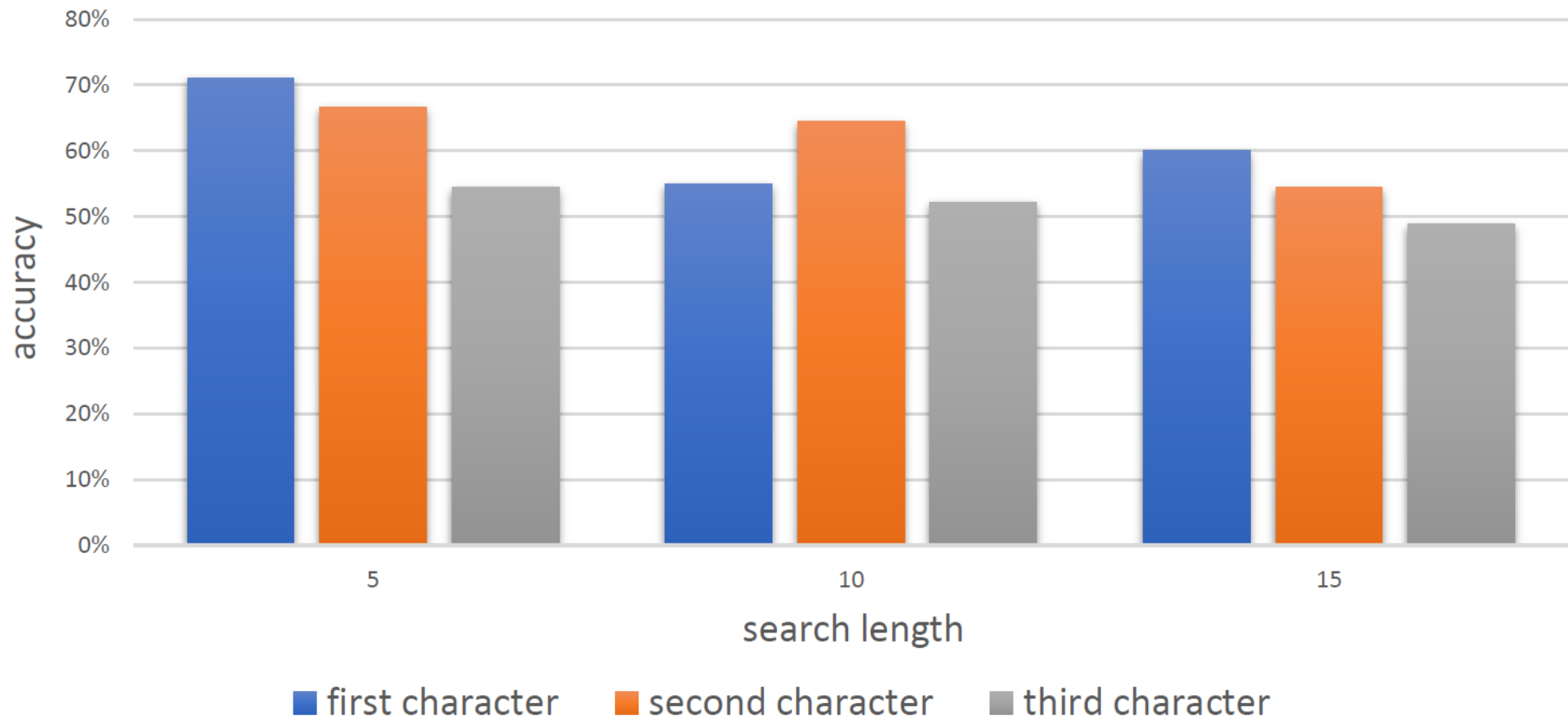
# Evaluation

- Original sequences



# Evaluation

- Compressed sequences



# Conclusion

- Smith-Waterman is suited for location prediction
- Uncompressed data accuracy is higher
  - Usable for high performance devices
- Compressed data redundancy is lower
  - Usable for low performance devices
- Future Work:
  - Real-time applications
  - Improve sequence prediction with same matrix scores
  - Take more data into account (longer sequences, coordinates)
  - Smart-Home: Radiator, Coffee machine, ...

# Thank you for your attention!





- [1] „Dynamic programming and sequence alignment“ <https://www.ibm.com/developerworks/library/j-seqalign/>, accessed: 2018-02-07.
- [2] „Google maps – timeline“ <https://www.google.com/maps/timeline>, accessed:2018-02-15.
- [3] I. Craig and M. Whitty, “Region formation for efficient offline locationprediction,” IEEE Pervasive Computing, vol. 16, no. 1, pp. 66–73, 2017.
- [4] T. Gueniche, P. Fournier-Viger, and V. S. Tseng, „Compact predictiontree: A lossless model for accurate sequence prediction“ in International Conference on Advanced Data Mining and Applications. Springer,2013, pp. 177–188.
- [5] D. S. Hirschberg, „A linear space algorithm for computing maximal common subsequences“ Communications of the ACM, vol. 18, no. 6, pp. 341–343, 1975.
- [6] A. Kirmse, T. Udeshi, P. Bellver, and J. Shuma, „Extracting patterns from location history“ in ACM SIGSPATIAL GIS 2011, <http://www.sigspatial.org/>, 2011, pp. 397–400.
- [7] S. B. Needleman and C. D. Wunsch, „A general method applicable to the search for similarities in the amino acid sequence of two proteins“ Journal of molecular biology, vol. 48, no. 3, pp. 443–453, 1970.
- [8] C. B. Rjeily, G. Badr, A. H. A. Hassani, and E. Andres, „Predicting heart failure class using a sequence prediction algorithm“ in 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), Oct 2017, pp. 1–4.
- [9] S. Sigg, S. Haseloff, and K. David, „An alignment approach for context prediction tasks in ubicomp environments“ IEEE Pervasive Computing, vol. 9, no. 4, pp. 90–97, 2010.
- [10] T. F. Smith and M. S. Waterman, „Identification of common molecular subsequences“ Journal of molecular biology, vol. 147, no. 1, pp. 195– 197, 1981.
- [11] M. S. Waterman, „Introduction to computational biology: maps, sequences and genomes“. CRC Press, 1995.
- [12] N. Casagrande, „Basic-Algorithms-of-Bioinformatics Applet“, <http://baba.sourceforge.net/>

- Rjeily et al. [8] use a sequence prediction approach to detect heart deficiencies of patients.
- Sigg et al. [9] describe an approach where context sequences can be predicted by using alignment. The utilization of alignment algorithms allows Sigg et al. to find the most similar partial sequences and with the aid of these to predict the next entry.
- Craig et al. [3] use a clustering approach for defining whereabouts of one person through clusters. This is done by dividing a location in preferably small regions and then combining similar adjacent regions to bigger clusters.