

An Update on Language Model Training for Track Finding

Jakapat Kannika

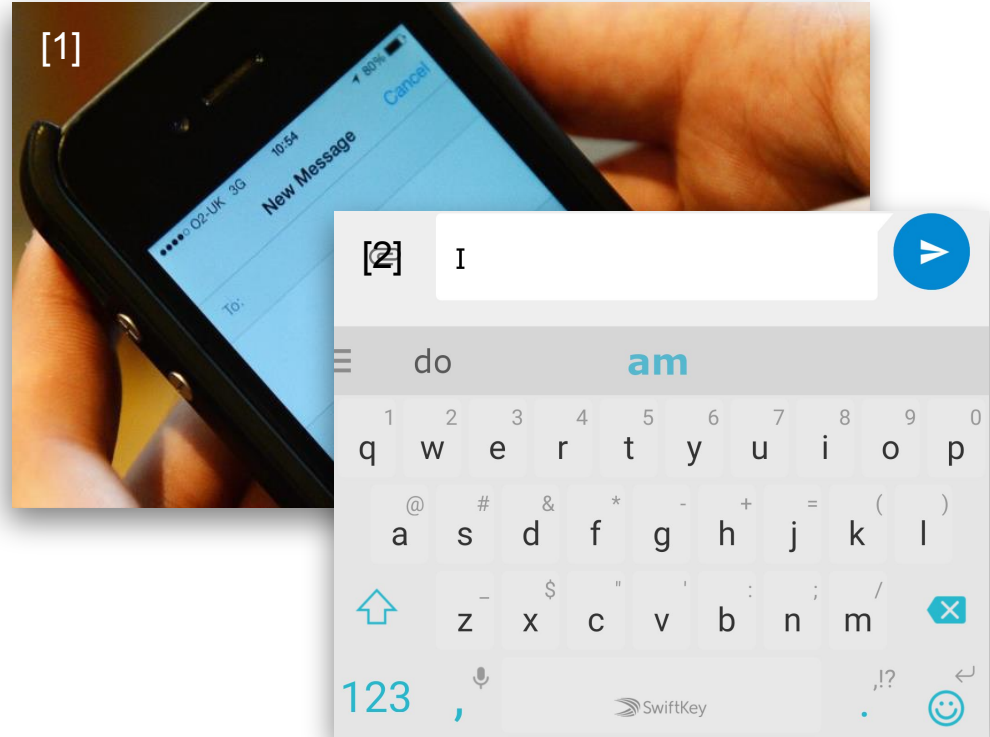
Language Model

Next word prediction

“I am Sam”

“Sam I am”

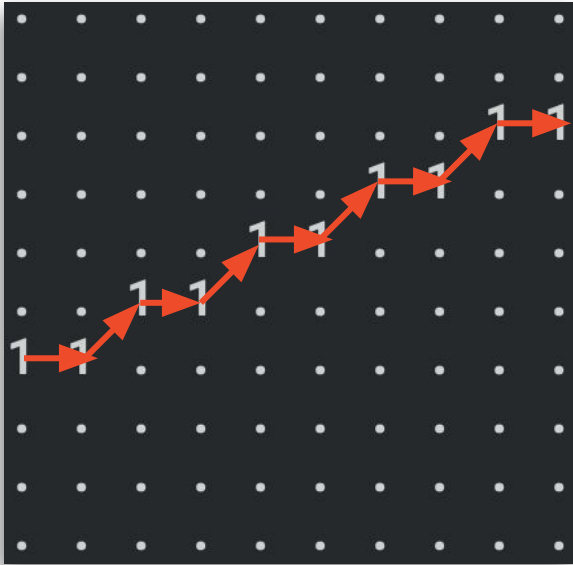
“I do not like green eggs and ham”



Sources:

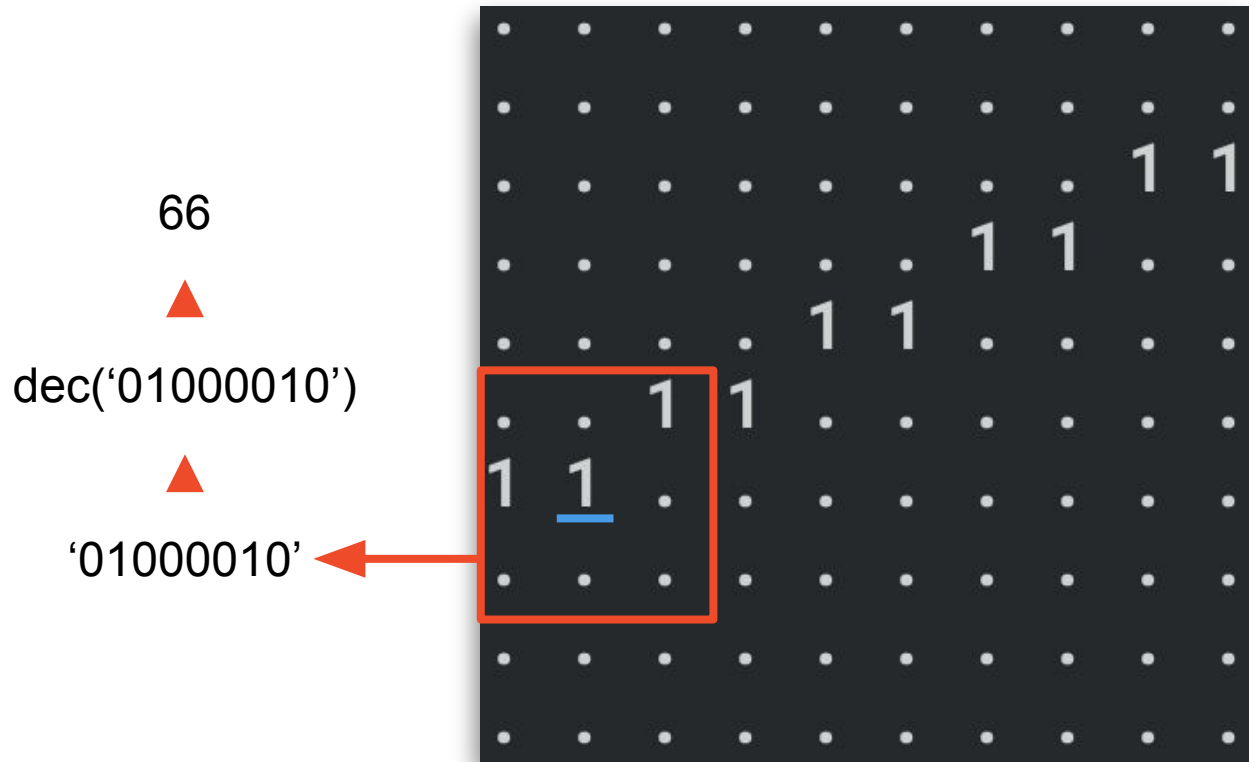
[1] <http://www.androidpolice.com>

[2] <http://qocall.com>



Moving direction tokens:

90 45 90 45 90 45 90 45 90



- 66 is a *neighbor pattern id*.
- There are 255 patterns (excluding one that has zero neighbor point).

Target language models

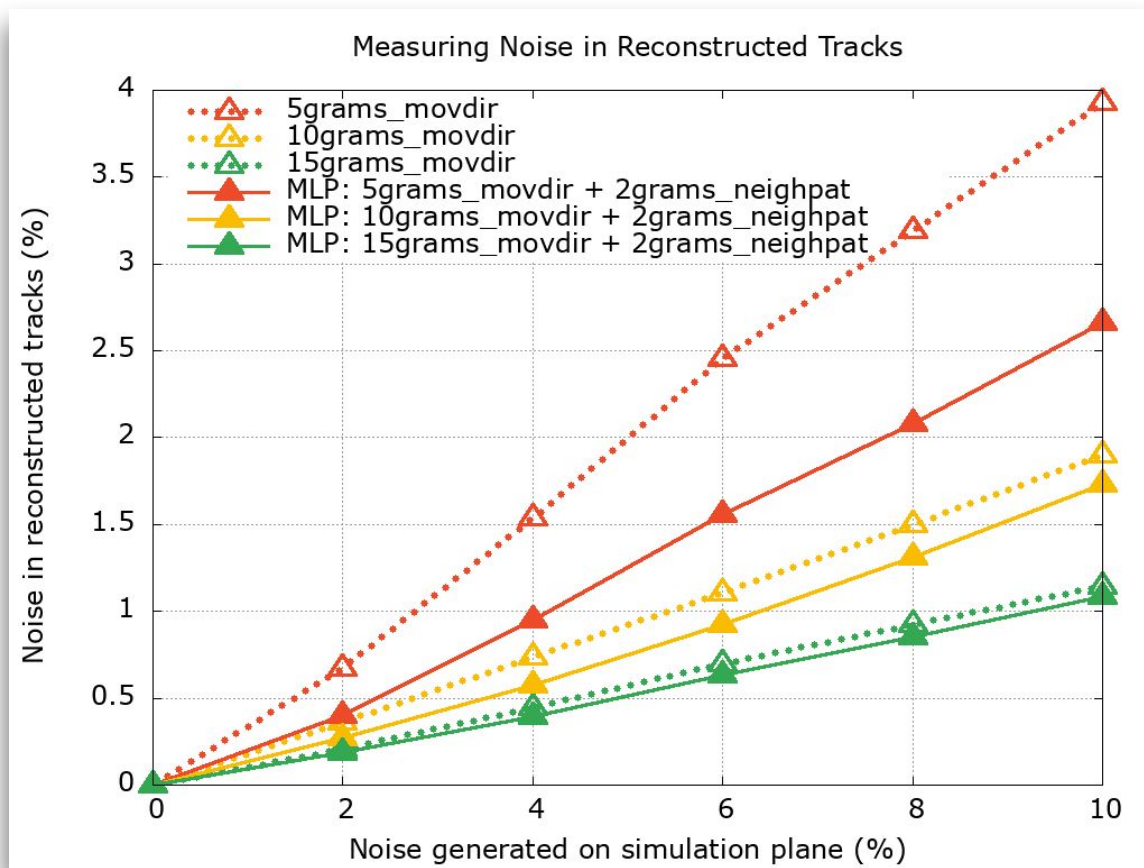
Moving directions:

- 5-gram model,
- 10-gram model,
- 15-gram model.

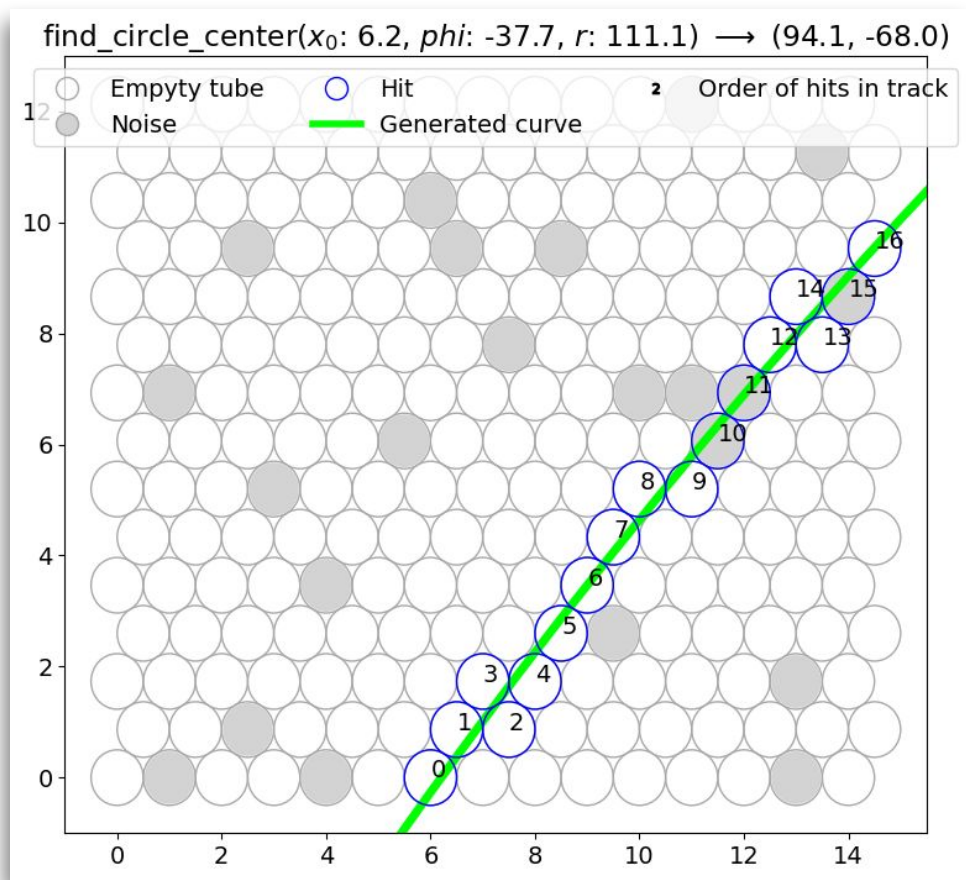
Neighbor pattern:

- 2-skip-2-gram model.

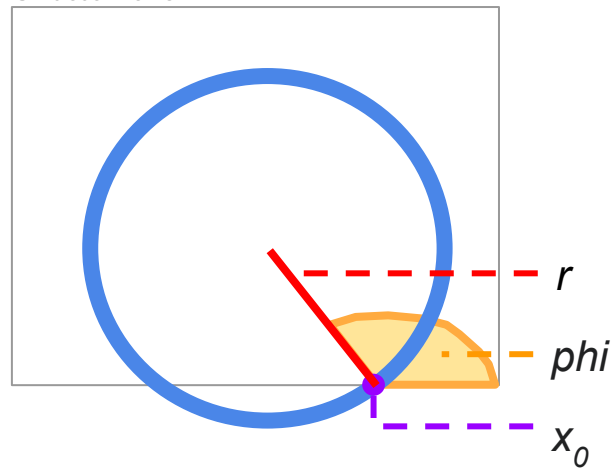
Pattern	Probability
150:30:150:30:150	0.730
150:30:150:30:90	0.001
150:30:150:30:</s>	0.267
150:30:150:90:150	0.030
150:30:150:90:90	0.382
...	...

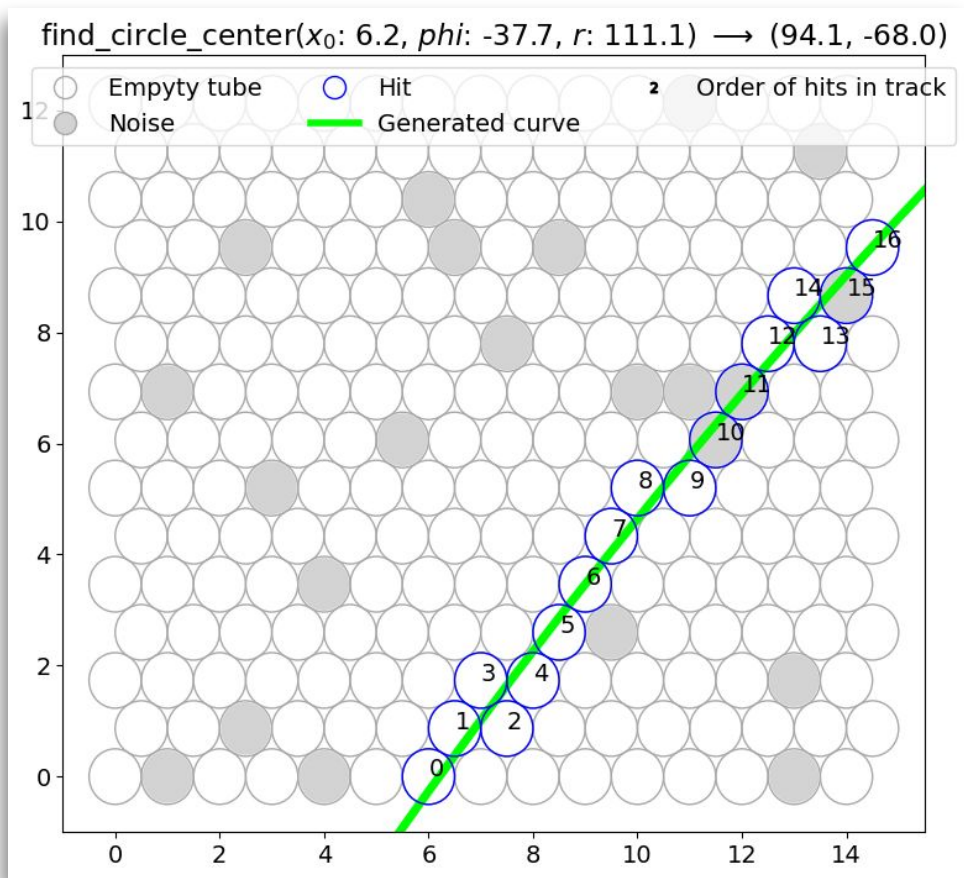


Generating Toy Data with STT-like Structure



Simulation frame





Positions: [

[6.0, 0.0], [6.5, 0.9], [7.5, 0.9], [7.0, 1.7],
 [8.0, 1.7], [8.5, 2.6], [9.0, 3.5], [9.5, 4.3],
 [10.0, 5.2], [11.0, 5.2], [11.5, 6.1], [12.0,
 6.9], [12.5, 7.8], [13.5, 7.8], [13.0, 8.7],
 [14.0, 8.7], [14.5, 9.5]]

Tracking features

Moving directions: [

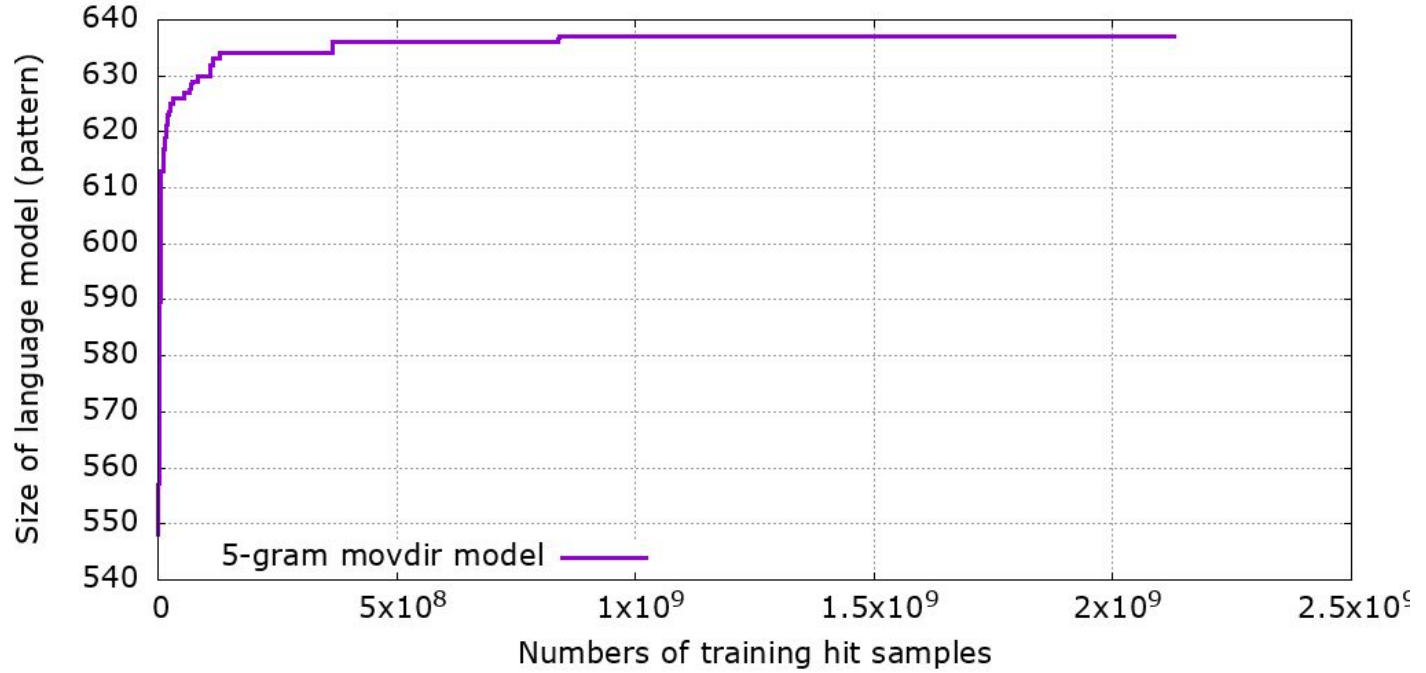
60, 0, 120, 0, 60, 60, 60, 60, 0, 60, 60,
 60, 0, 120, 0, 60]

Neighbor patterns: [

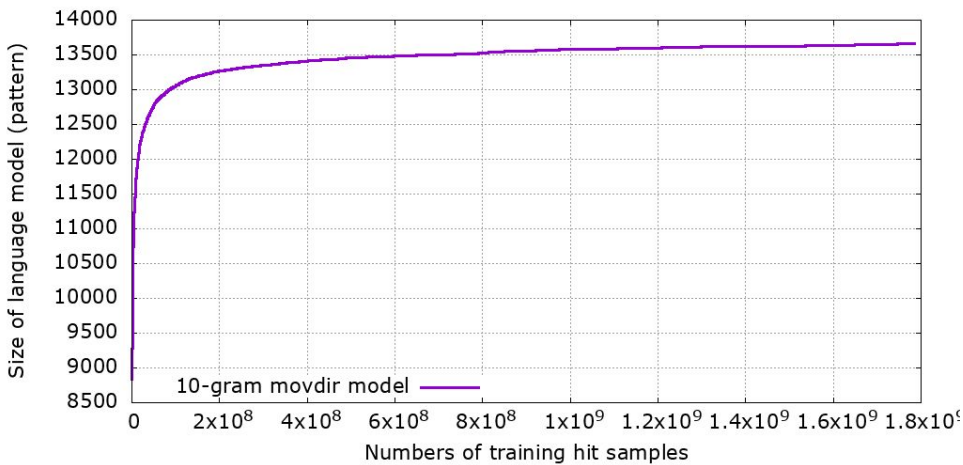
1, 41, 7, 56, 13, 41, 25, 9, 40, 5, 11,
 13, 41, 7, 56, 13, 8]

Language Model Training

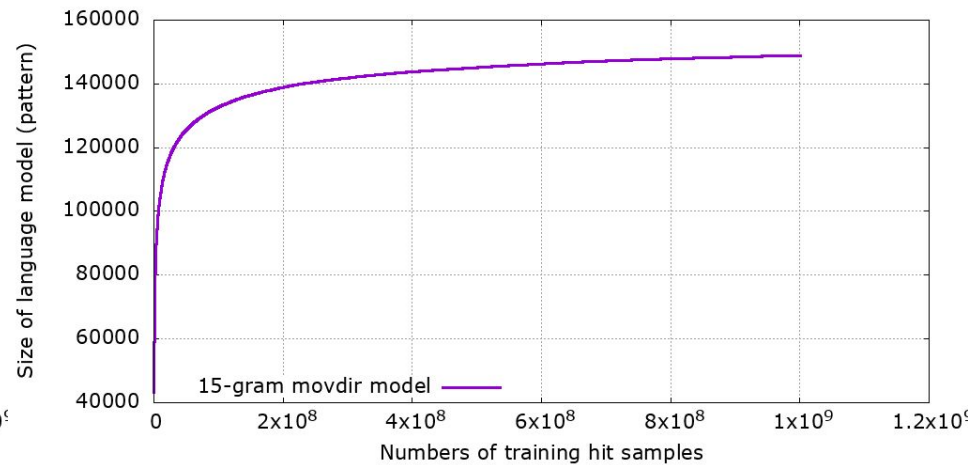
5-gram model



10-gram model



15-gram model

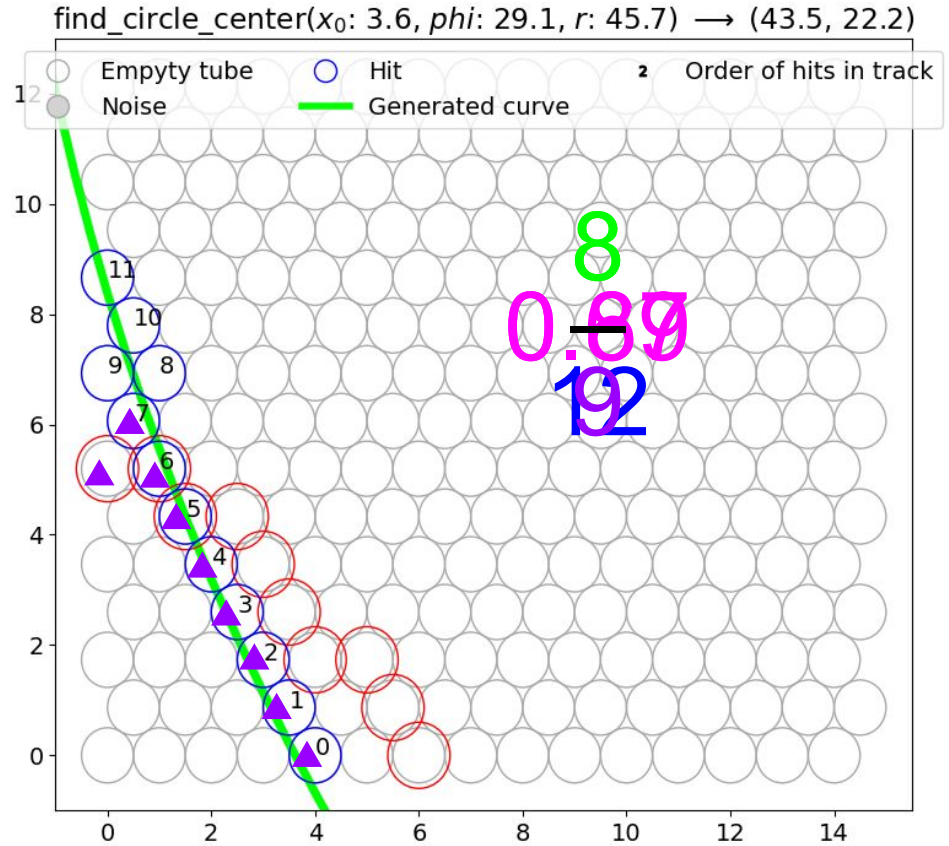


End-to-end Evaluation

End-to-end evaluation

Metrics:

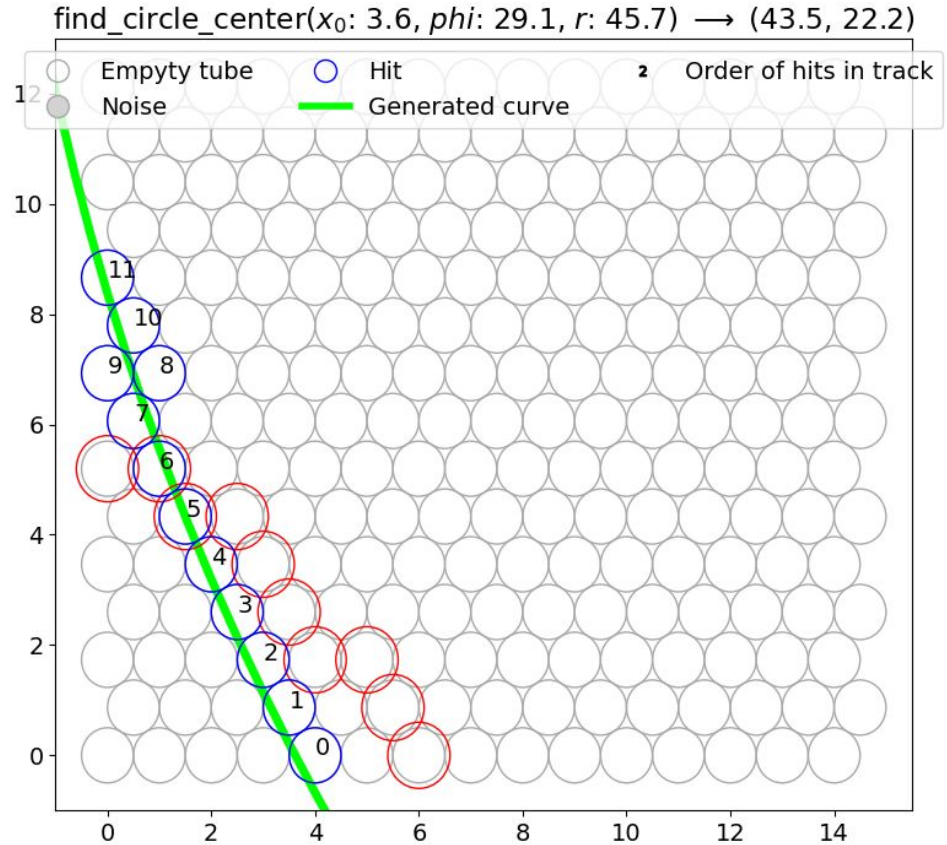
- Track completeness,
- Track purity.



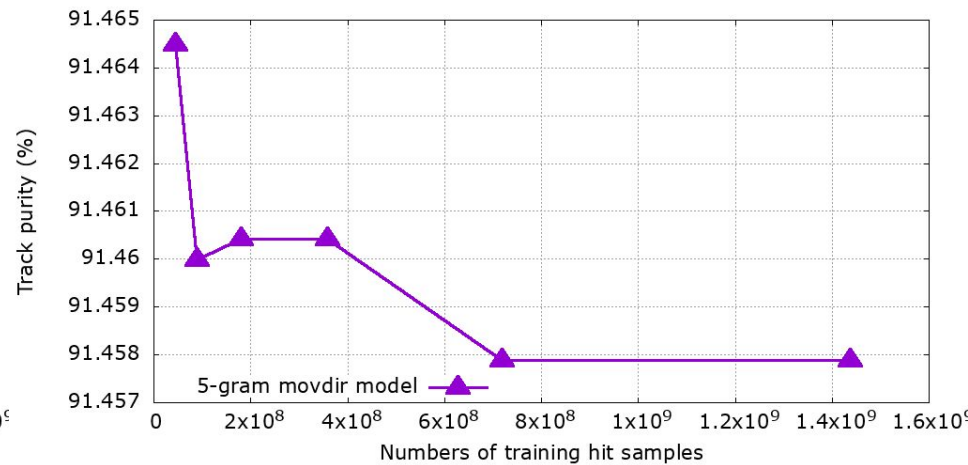
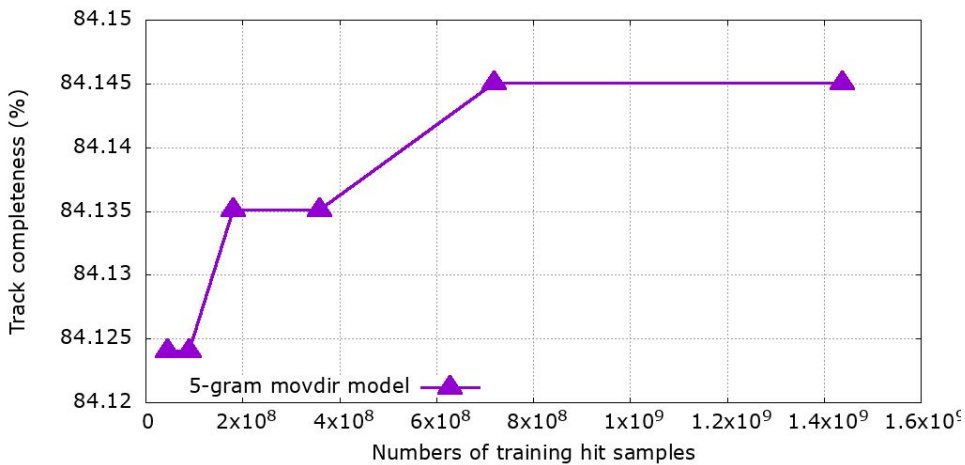
End-to-end evaluation

Test samples are generated with:

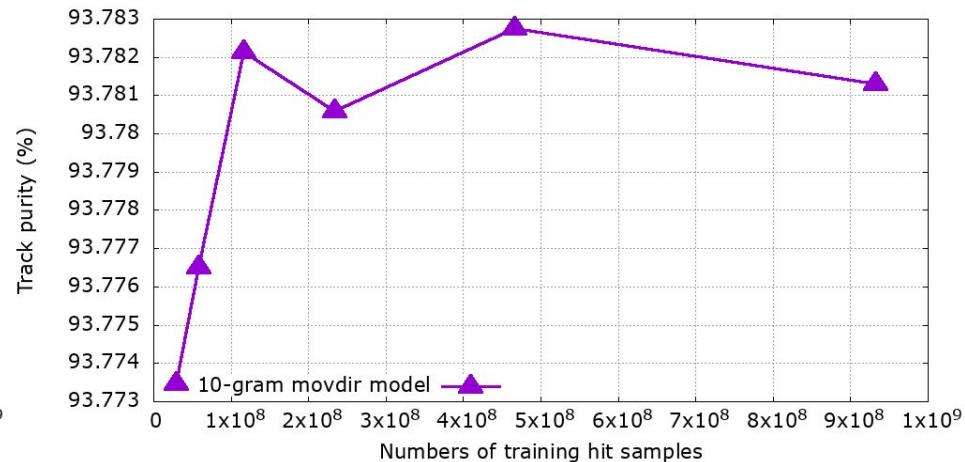
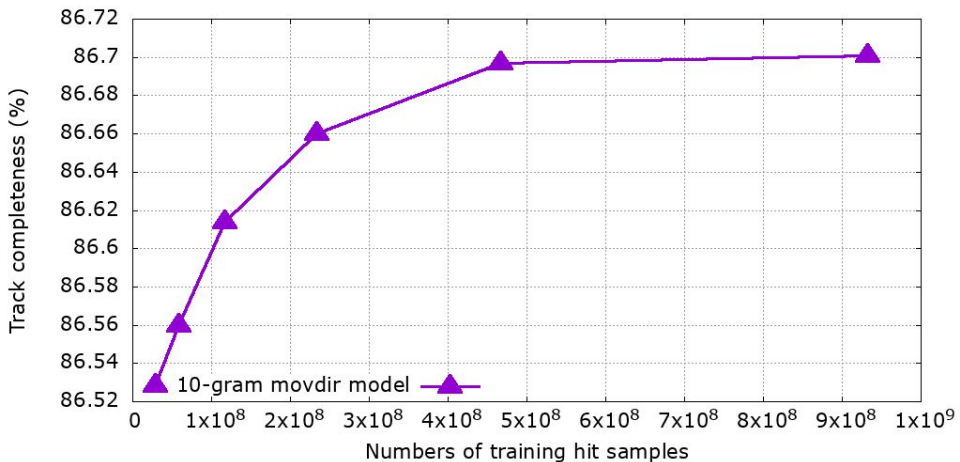
- **Number of tubes:**
25 x 25 tubes.
- **Number of samples:**
100,000 samples.
- **Number of tracks per sample:**
2 tracks per sample.
- **Conditions:**
Crossing tracks,
15 seed points.



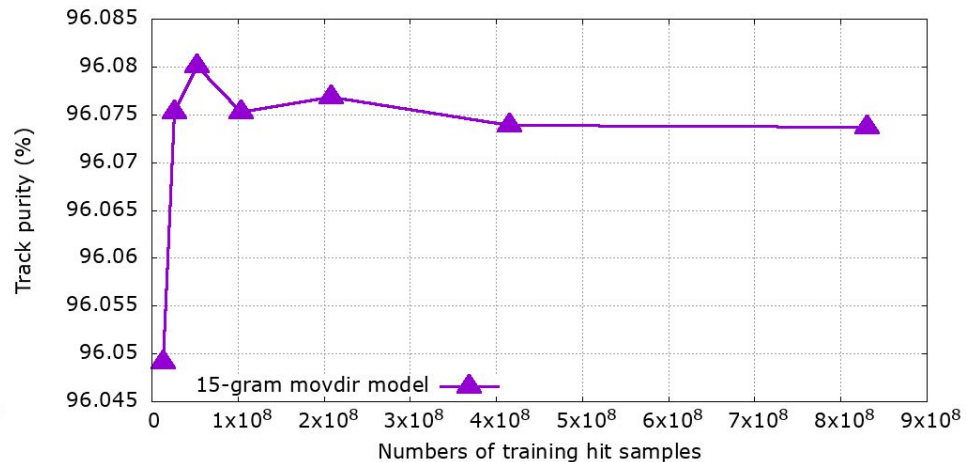
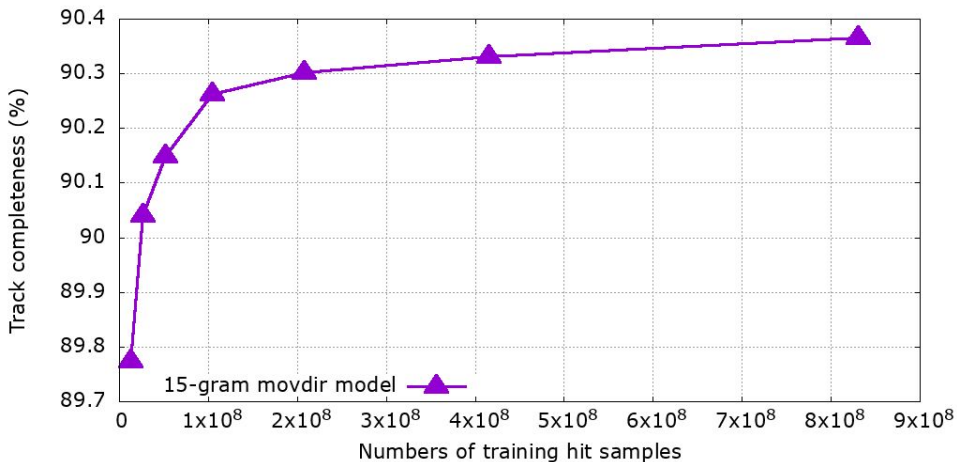
5-gram model: completeness vs purity



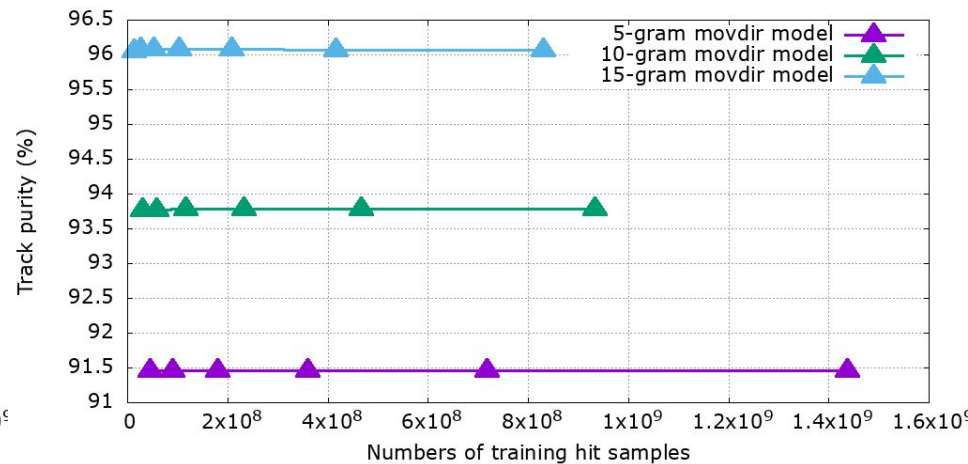
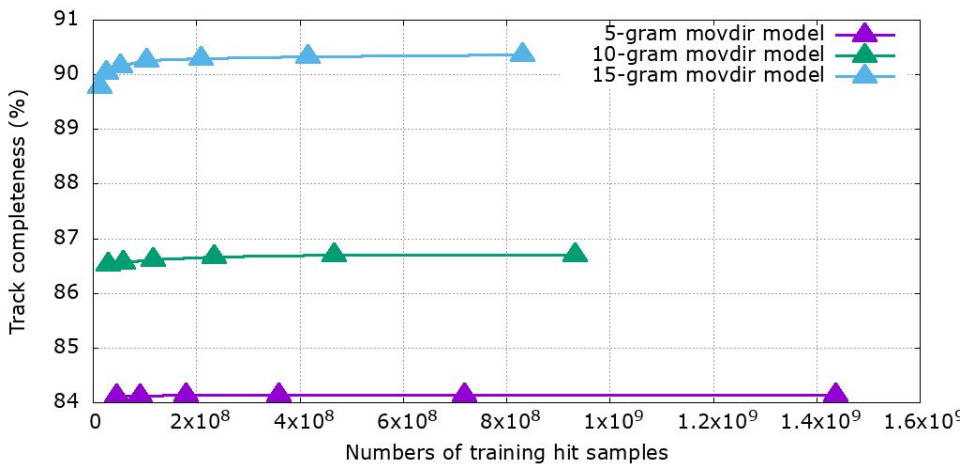
10-gram model: completeness vs purity



15-gram model: completeness vs purity



All models: completeness vs purity



Summary and outlook

Summary

- Using curved tracks in training dramatically increase the training time and size of the language models,
- End-to-end evaluations can be helpful for decide a reasonable terminating time for the model training especially in the case of large-sized language models.

Outlook

- We are currently working on designing a sequential pattern recognition model for tracking isochrone radii.

