A Spherical Hidden Markov Model for Semantics-Rich Human Mobility Modeling

Wanzheng Zhu (WZ)*, Chao Zhang*, Shuochao Yao, Xiaobin Gao, Jiawei Han

University of Illinois, at Urbana-Champaign

wz6@illinois.edu

Human Mobility Modeling – Problem Description

Input

Human Trace Data: *Twitter*

Time	Fri Aug 01 00:00:30 CDT 2014
Location	33.982552, -118.309053
Text	absolutely_loved guardian_galaxyI absolutely loved Guardians of the Galaxy😁

Output

Mobility Modeling

Pattern Discovery

- Understanding Human Movement
- Next Location Prediction
- Traffic Scheduling
- Urban Planning
- Activity Recommendation

What & ***When*** are people's activities at different regions?

How & ***Why*** do people move from one region to another?

Human Mobility Modeling – Related Works

Pattern-based Methods

Discovering specific mobility patterns that occur frequently

- Frequent Sequential Patterns
- Co-location Patterns
- Periodic Patterns

Model-based Methods

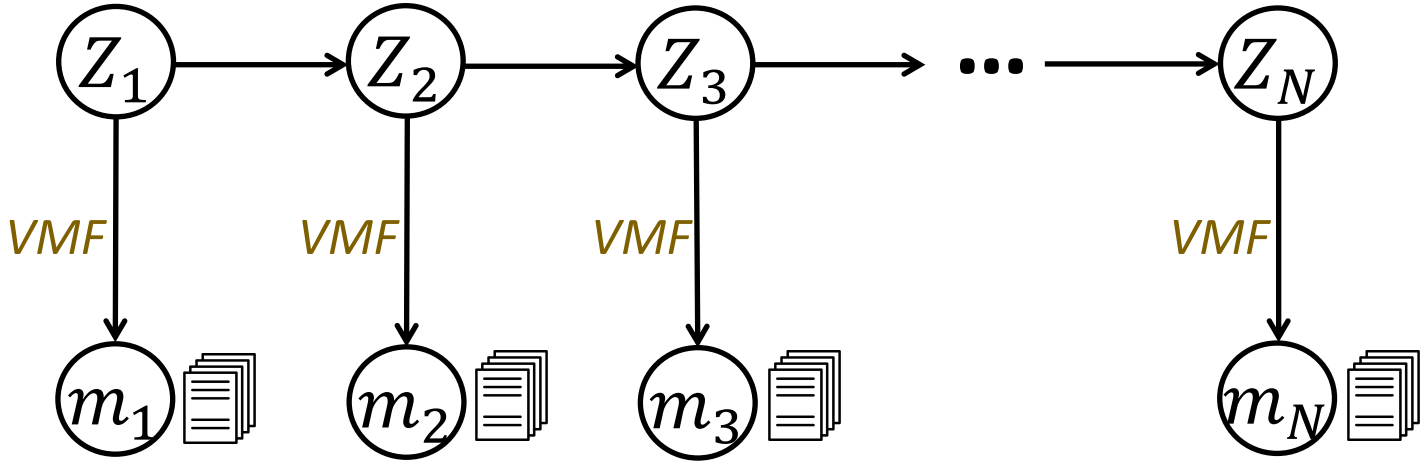
Statistical models – fitting data into the model.

- Hidden Markov Models
- Continuous-time Random-walk Models
- Periodicity Underlying Human Movement

Most previous works do NOT consider TEXT information!

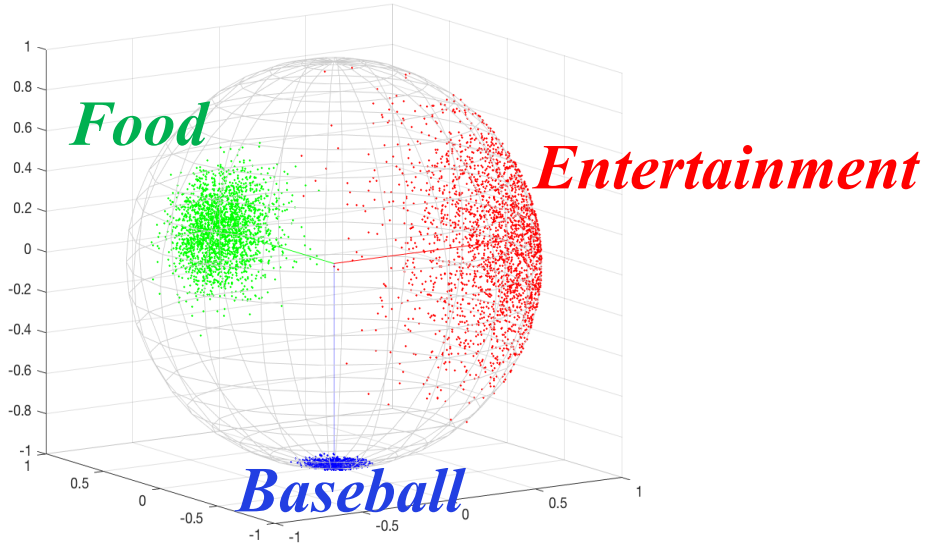
Previous *State-of-the-art* work: GMove (Zhang *et al.* 2016a)

Model Description (HMM + VMF)



Hidden States

Word Embeddings



The vMF Distribution

Intuition:

- Word embeddings are on the surface of a **unit hypersphere**
- ***The Von Mises-Fisher (vMF) distribution*** is well suited for *unit-length* directional data with different *concentration* levels.

Interesting States

State 1: **Basketball**	State 2: **Baseball**	State 3: **General Sports**	State 4: **Restaurant/ Food**	State 5: **Beach**	State 6: **Universal Studios**	State 7: **Airport**
lakers	homeruns	cfbplayoff	osteriamamma	monica	universal	international
grizzlies	dodgers	oilers	island's	beachselfie	waterworld	angeles
celtics	giants	2-May	hotwings	zoltar	hhn2014	los
nba	162	jets	jicama	beach	hollywoodhorrornights	airport
bryant	sfvspit	preseason	flore	fishbarmb	universalcitywalk	lax

Dataset: ~30K *Tweets* in LA, from Aug/01/2014 to Nov/30/2014

Interesting Patterns:

- ***A***: LA Airport/Flights → Restaurants in LA
- ***B***: Beach Activities → Leisure Activities
- ***C***: Food → Concert/ Show

The vMF Distribution – Parameter Inference

How to update the HMM parameters?

M-step of the EM algorithm:

$$\theta := argmax_\theta \sum_i \sum_{z_i} Q_i(z_i) \log \boldsymbol{p(m_i, z_i|\theta)}$$

- Very complicated math expression

$$\boldsymbol{f_p(x; \mu, \kappa)} = \frac{\boldsymbol{\kappa^{p/2-1}}}{\boldsymbol{(2\pi)^{p/2} I_{p/2-1}(\kappa)}} \boldsymbol{exp(\kappa \mu^T x)}$$

Modified Bessel Function: $\boldsymbol{I_{p/2-1}} = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+p/2)} \left(\frac{x}{2}\right)^{2m+p/2-1}$

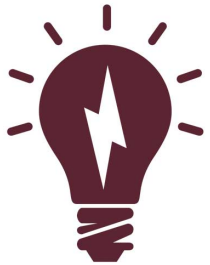
Gamma function

Theoretical Contribution of the vMF distribution

There is ***no closed form solution*** for updating $\boldsymbol{\kappa}$ in the M-step.

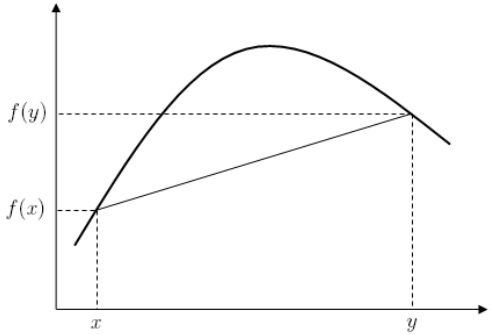
Existence: There exists ***one unique*** $\boldsymbol{\kappa}$ to maximize the M-step.

- By ***the intermediate value theorem***



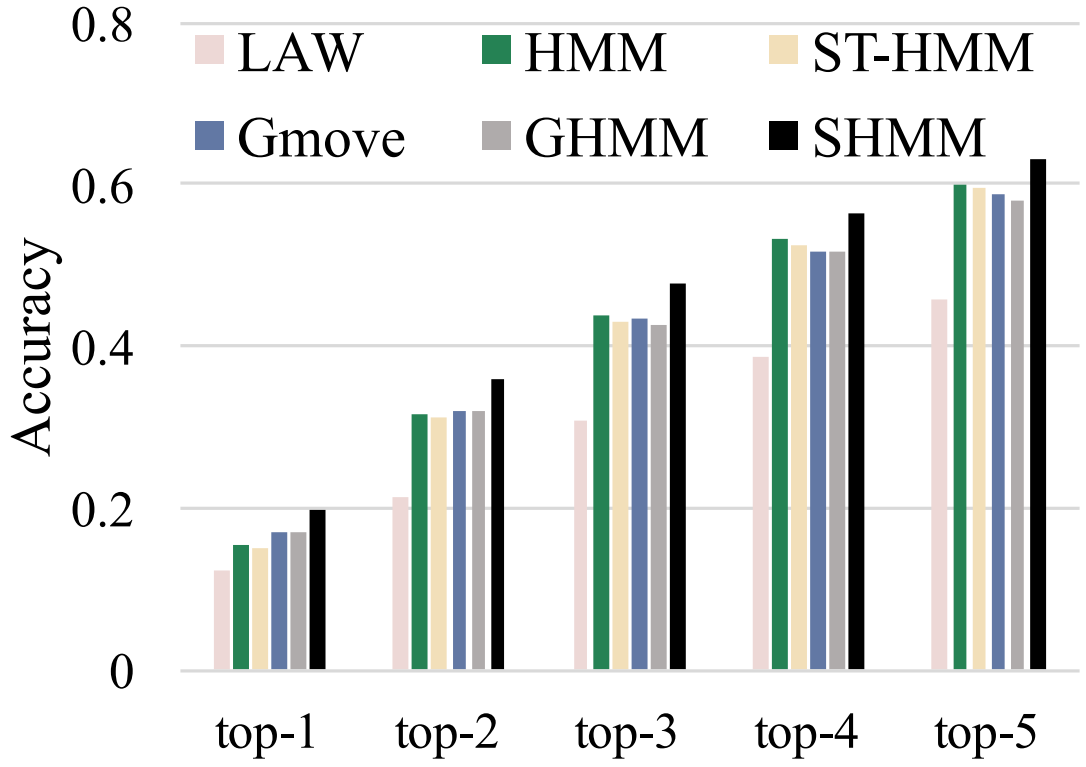
Feasibility: ***Newton's method*** is guaranteed to converge to the optimal κ.

- By its ***concavity*** property
- Quadratic convergence rate

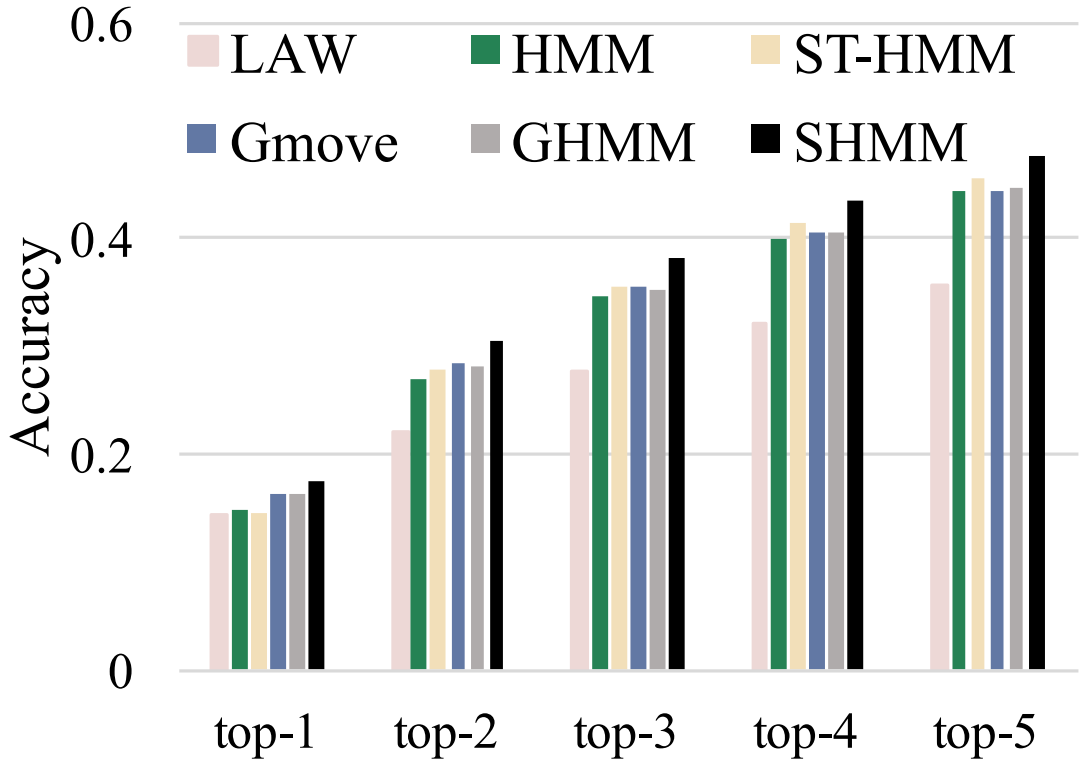


Simulation: Synthetic data

Results – Next Location Prediction



Dataset: ~30k *Tweets* in LA, from Aug/01/2014 to Nov/30/2014



Dataset: ~42k Tweets in NYC, from Aug/01/2014 to Nov/30/2014

Accuracy Improvement: ~**3.2**%

Contributions and Conclusion

Model

HMM + VMF

Mobility Modeling

Downstream Applications

Semantics

Interpretable

Theory

Existence

Feasibility

HMM + VMF

EM + VMF

Thank you!

A Spherical Hidden Markov Model for Semantics-Rich Human Mobility Modeling

Wanzheng Zhu (WZ)
wz6@illinois.edu

Code & Data & Paper & Presentation:
https://github.com/WanzhengZhu/SHMM