Centre for
Data Analytics

Insight

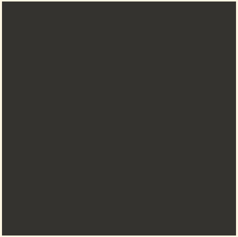
Towards the Recommendation of Personalised Activity Sequences in the Tourism Domain

Gunjan Kumar, Houssem Jerbi, and Michael P. O'Mahony
Insight Centre for Data Analytics
University College Dublin

RecTour 2017, Como

Aug 27, 2017

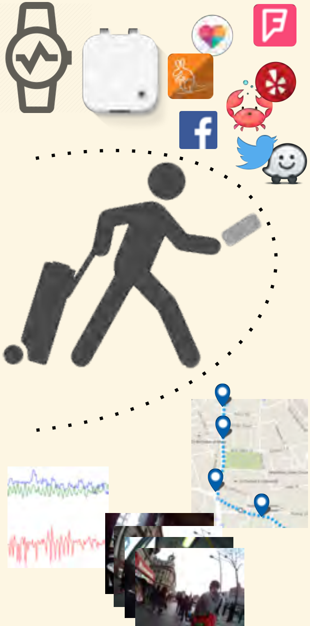
e-tourism



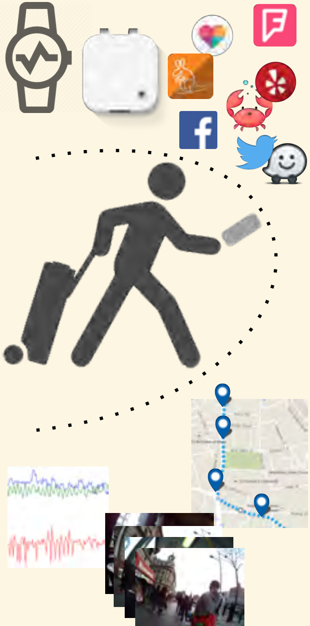
e-tourism

Booking.com
skyscanner
Tripomatic.com
Expedia
trivago
tripadvisor
KAYAK
yelp
Tripit from Concur

Sensors: Rich User Activity Data



Sensors: Rich User Activity Data



- **Sequential** nature of user activities
- Activities have associated **features/context**, e.g. location, time, weather, etc.

Rich User Activity Data

For Recommender Systems

Facilitates real time recommendations for a given user and context (e.g. time, location, weather, etc.)

Previous work:

A framework for sequence- and context-based recommendation of **next activity** (lifelogging/modes of transport) to perform. [Kumar et al., 2014, 2016]

Current Research Problem:

Recommending the next sequence of activities to users.

Rich User Activity Data

For Recommender Systems

Facilitates real time recommendations for a given user and context (e.g. time, location, weather, etc.)

Previous work:

A framework for sequence- and context-based recommendation of **next activity** (lifelogging/modes of transport) to perform. [Kumar et al., 2014, 2016]

Current Research Problem:

Recommending the next sequence of activities to users.
e.g. visiting a museum, having Italian food, and going to a theatre.

Related Work

Capturing Sequence

- Hierarchical-graph-based model:
  - [Li et al., 2008; Zheng et al., 2009; Yoon et al., 2010]
- All-k^{th}-order Markov models:
  - [Bohnenberger and Jameson, 2001; Deshpande and Karypis, 2004; Shani et al., 2005]

Capturing Context

- Tensor and matrix factorization models:
  - [Zheng et al., 2010, 2012; Wang et al., 2010; Symeonidis et al., 2011; ?; Adomavicius et al., 2011; Braunhofer et al., 2013]

Related Work

Capturing Both Sequence & Context

- To improve recommendations
  - [Adomavicius and Tuzhilin, 2005; Zheng et al., 2012]
- Content-based Activity Recommendation Framework
  - [Kumar et al., 2014, 2016]
- Stochastic Modelling
  - [Sun et al., 2016]

Recommending Sequences

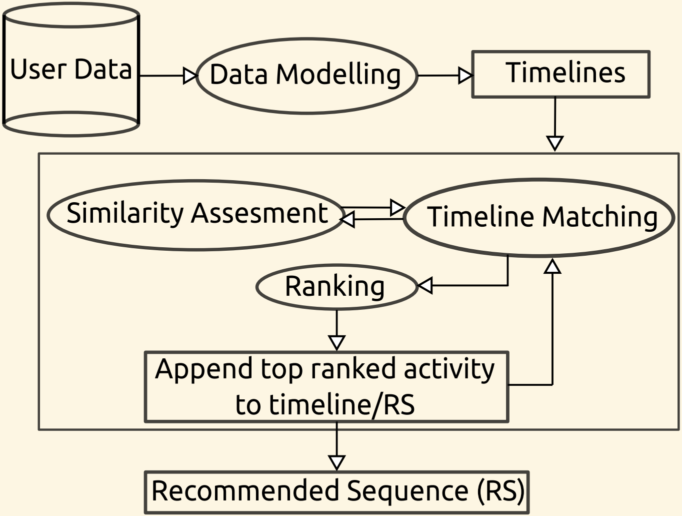
- Music playlists
  - [Baccigalupo and Plaza, 2006; Chen et al., 2012]
- POI/Itinerary
  - [Tai et al., 2008; Yoon et al., 2012]

Our Contribution

- A generic activity recommendation framework to recommend the next sequence of activities to users based on past activity patterns and context (extending [Kumar et al., 2014, 2016]).

- Application of the proposed approach in the tourism domain. Experiments using a location checkin dataset.

Framework Overview



Data Model

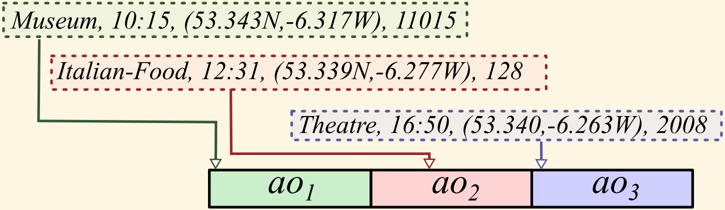
Activity Object

A single occurrence of an activity and consists of a set of features describing the activity or the context.

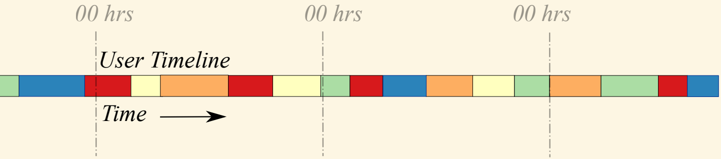
Activity Timeline

A chronological sequence of n activity objects performed by the user during a time interval δ:

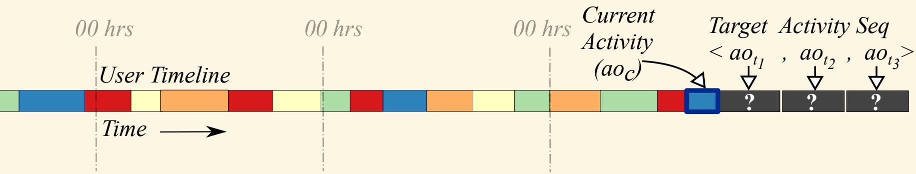
$$\mathcal{T} = < ao_1, ao_2, ..., ao_n >$$



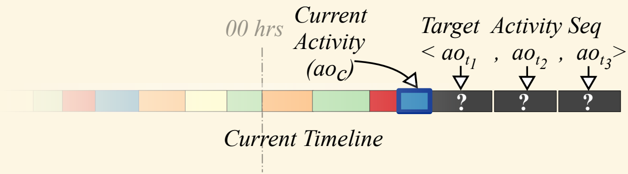
Recommendation Algorithm



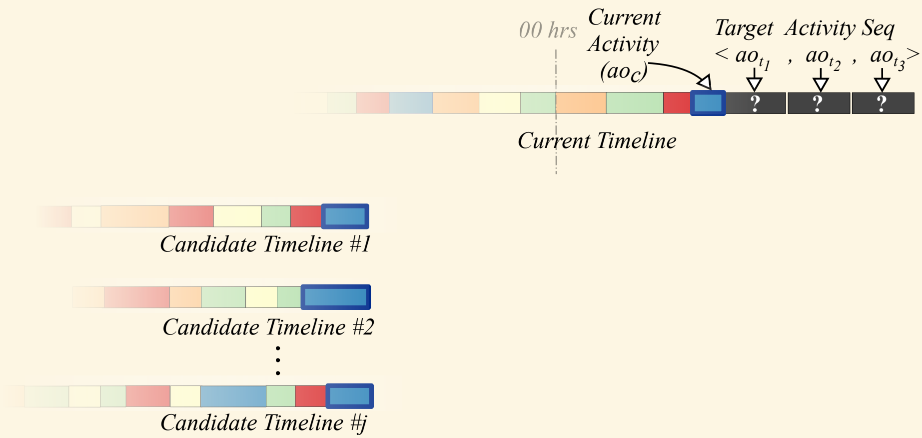
Recommendation Algorithm

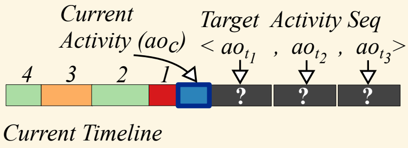


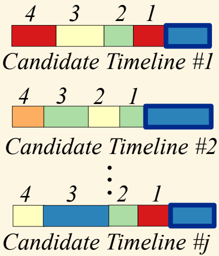
Recommendation Algorithm



Recommendation Algorithm







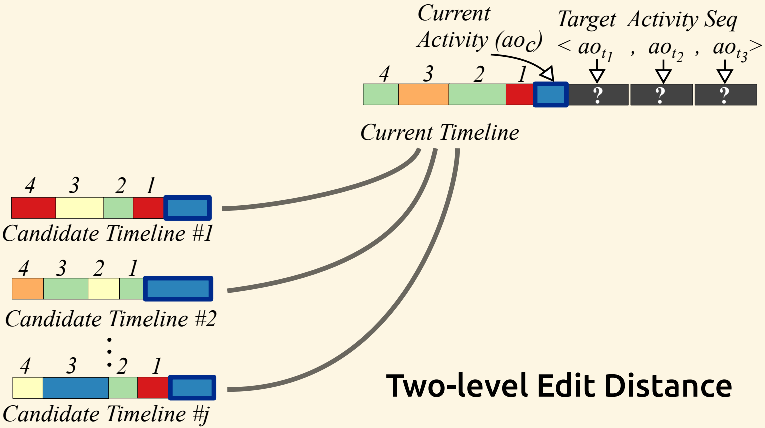
N-count matching ($N = 4$)

[Kumar et al., 2016]

Matching unit

Determines the length of the subsequences to be compared.

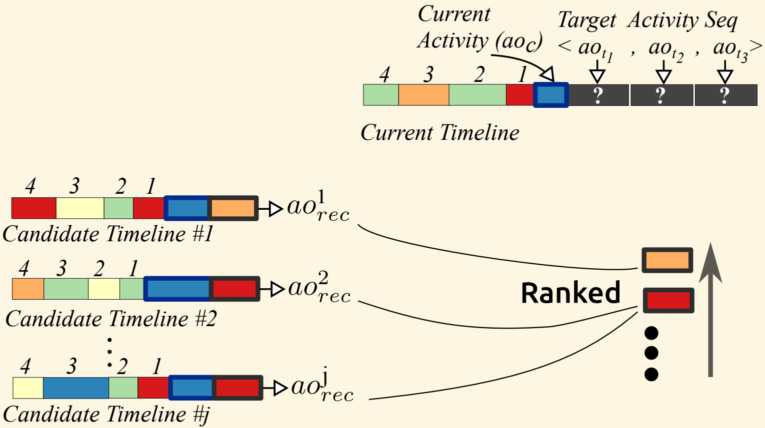
Similarity Assessment



Two-level Edit Distance

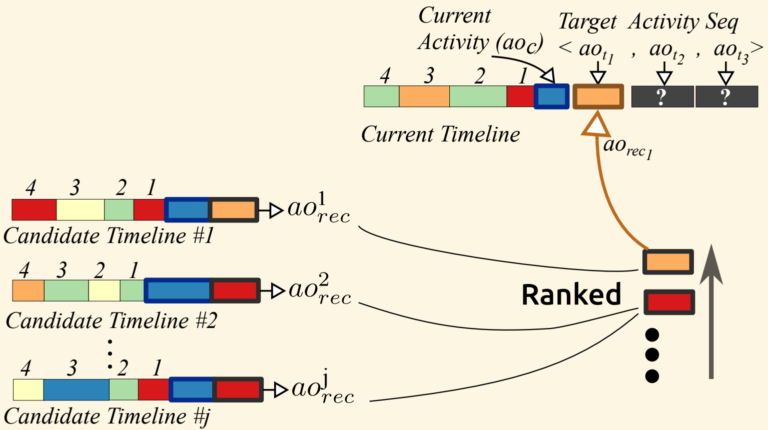
[Kumar et al., 2014]

Ranking

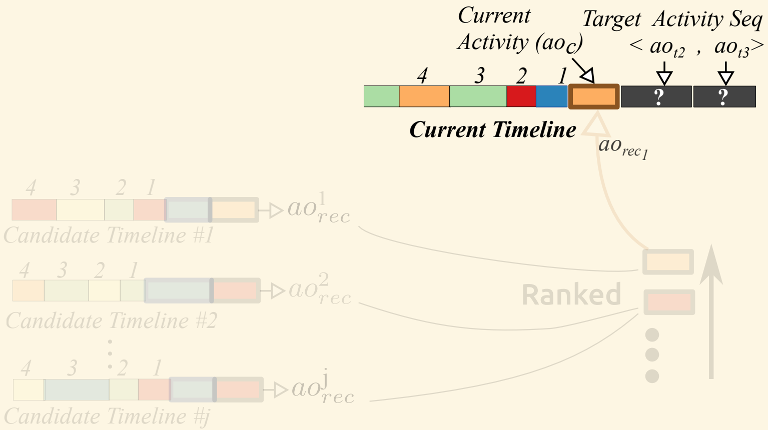


$$Score(ao^j_{rec}) = 1 - \frac{d(\mathcal{T}_j, \mathcal{T}_c) - \min_{\mathcal{T}_p \in \mathscr{T}} d(\mathcal{T}_p, \mathcal{T}_c)}{\max_{\mathcal{T}_p \in \mathscr{T}} d(\mathcal{T}_p, \mathcal{T}_c) - \min_{\mathcal{T}_p \in \mathscr{T}} d(\mathcal{T}_p, \mathcal{T}_c)}$$

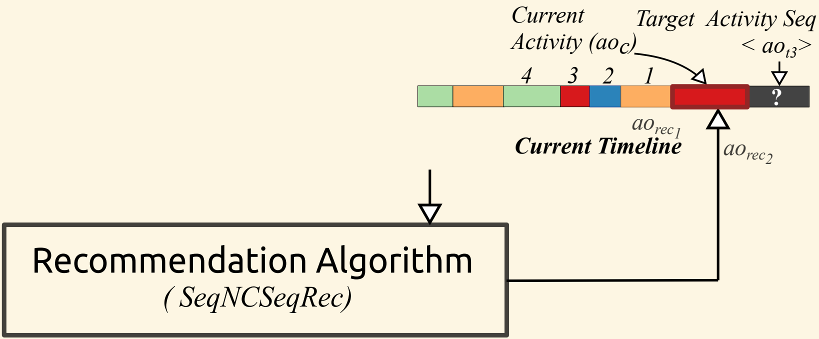
Recommending



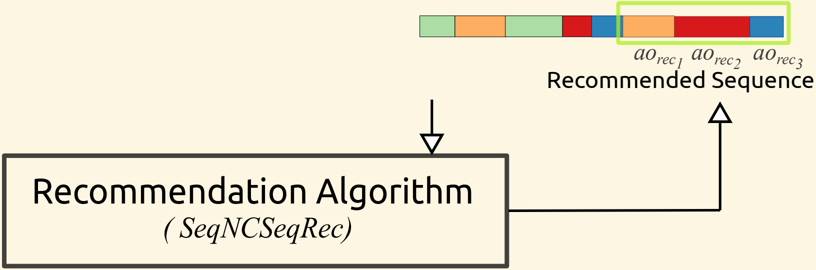
Recommending



Recommending

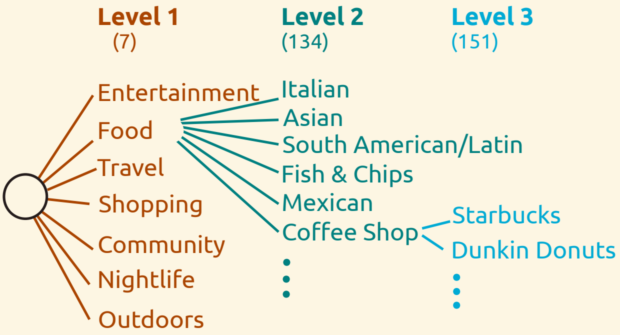


Recommending



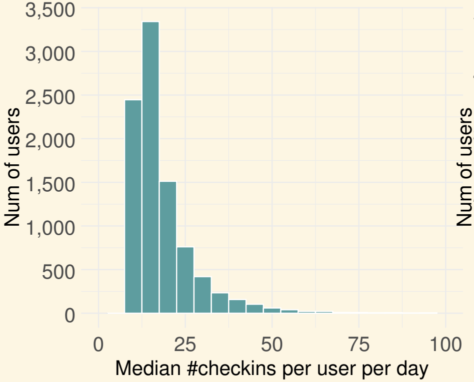
Dataset

- Gowalla checkins dataset [Liu et al., 2013]
- Every checkin is bound to a specific location and timestamp.
- Features: latitude, longitude, number of users checking in to it, number of photos taken at the location, etc.
- Locations have categories assigned to them, such as, ‘Italian Food’, ‘Museum’, ‘City Park’, etc.

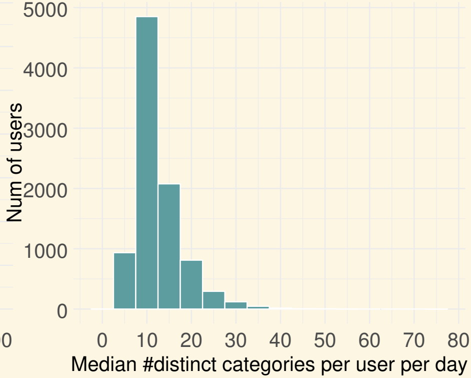


Dataset

- Sampled dataset: 9.2k users , 6 million checkins in total.
- median # checkins per user per day: 10 - 270
- median # of distinct categories per user per day : 2 - 76

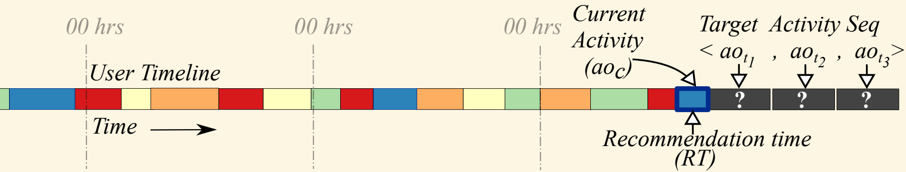


(a)



(b)

Methodology



- **Leave-one-out evaluation:** Each user's complete timeline is split into training and test timelines by time.
- **Agreement @ k** ($k = 1, 2, 3$): % of RTs for a user where the first k categories in the recommended sequence and the actual sequence are an *exact* match.
- **Recommendation algorithms:**
  - *N*-count recommendation algorithm (*SeqNCSeqRec*)
  - Bi-gram-based sequence recommender (*BiGramSeqRec*)
  - Popularity-based sequence recommender (*PopSeqRec*)

Recommendation Performance (Level 2)

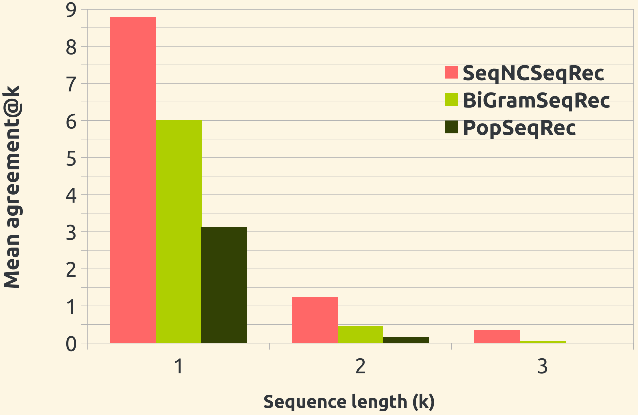
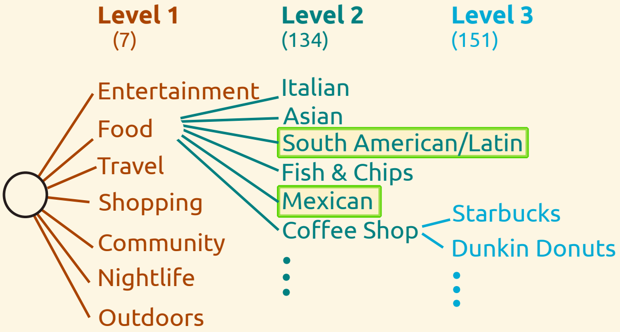


Figure: Mean percentage agreements for recommended sequences for *SeqNCSeqRec* (top 10% neighbours) and baseline algorithms using timelines constructed from categories at level 2 in the hierarchy.

Recommendation Performance

- Some level 2 activities are semantically closer than others.
- ‘true’ performance between those at level 2 and level 1.



Recommendation Performance

- Some level 2 activities are semantically closer than others.
- ‘true’ performance between those at level 2 and level 1.

Recommendation Performance (Level 1)

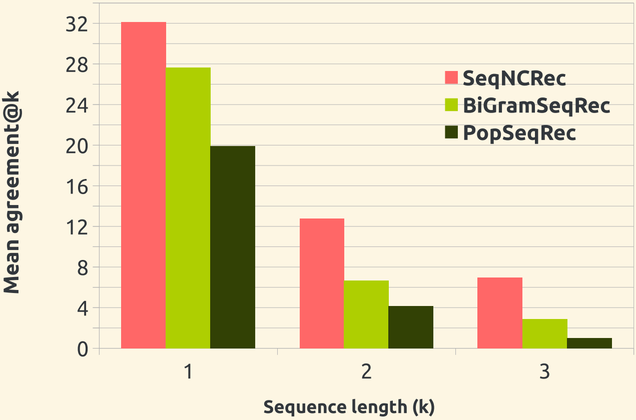


Figure: Mean percentage agreements for recommended sequences for *SeqNCSeqRec* (top 10% neighbours) and baseline algorithms using timelines constructed from categories at level 1 in the hierarchy.

Conclusion

- A generic activity recommendation framework to recommend the next sequence of activities to users based on past activity patterns and context (extending [Kumar et al., 2014, 2016]).
- Experiments demonstrate the efficacy of our approach in recommending sequences given a diverse variety of activities and user activity patterns.

Future Work

- Consider alternative approaches to suggest sequences of activities (for example, using RNNs).
- Introduce new evaluation metrics for evaluating sequence recommendation.
- Investigate the recommendation of context (for example, where, when, with whom etc.) associated with each of the suggested sequence of activities.
- Consider socio-economic characteristics, user demographics, and travel variables.

Centre for
Data Analytics

Insight

Thank You

gunjan.kumar@insight-centre.org

@gunjanthesystem

References I

G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, June 2005. ISSN 1041-4347. doi: **10.1109/TKDE.2005.99**. URL http://dx.doi.org/10.1109/TKDE.2005.99.

G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-aware recommender systems. *AI Magazine*, 32(3), 2011.

C. Baccigalupo and E. Plaza. Case-based sequential ordering of songs for playlist recommendation. In *Proceedings of the 8th European Conference on Advances in Case-Based Reasoning*, ECCBR'06, pages 286–300, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-36843-4, 978-3-540-36843-4. doi: **10.1007/11805816_22**. URL http://dx.doi.org/10.1007/11805816_22.

T. Bohnenberger and A. Jameson. When policies are better than plans: Decision-theoretic planning of recommendation sequences. In *Proceedings of the 6th International Conference on Intelligent User Interfaces*, IUI '01, pages 21–24, New York, NY, USA, 2001. ACM. ISBN 1-58113-325-1. doi: **10.1145/359784.359829**. URL http://doi.acm.org/10.1145/359784.359829.

M. Braunhofer, M. Elahi, F. Ricci, and T. Schievenin. *Context-Aware Points of Interest Suggestion with Dynamic Weather Data Management*, pages 87–100. Springer-Verlag, Cham, 2013. ISBN 978-3-319-03973-2. doi: **10.1007/978-3-319-03973-2_7**. URL https://doi.org/10.1007/978-3-319-03973-2_7.

References II

S. Chen, J. L. Moore, D. Turnbull, and T. Joachims. Playlist prediction via metric embedding. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 714–722, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1462-6. doi: **10.1145/2339530.2339643**. URL http://doi.acm.org/10.1145/2339530.2339643.

M. Deshpande and G. Karypis. Selective Markov models for predicting web page accesses. *ACM Transactions on Internet Technology*, 4(2):163–184, May 2004. ISSN 1533-5399. doi: **10.1145/990301.990304**. URL http://doi.acm.org/10.1145/990301.990304.

G. Dong and J. Pei. *Sequence Data Mining (Advances in Database Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007. ISBN 0387699368.

G. Kumar, H. Jerbi, C. Gurrin, and M. P. O'Mahony. Towards activity recommendation from lifelogs. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '14, pages 87–96, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3001-5. doi: **10.1145/2684200.2684298**. URL http://doi.acm.org/10.1145/2684200.2684298.

G. Kumar, H. Jerbi, and M. P. O'Mahony. Personalised recommendations for modes of transport: A sequence-based approach. *The 5th ACM SIGKDD International Workshop on Urban Computing (UrbComp 2016)*, 2016.

References III

Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '08, pages 34:1–34:10, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-323-5. doi: **10.1145/1463434.1463477**. URL http://doi.acm.org/10.1145/1463434.1463477.

X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 733–738, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2263-8. doi: **10.1145/2505515.2505639**. URL http://doi.acm.org/10.1145/2505515.2505639.

G. Shani, D. Heckerman, and R. I. Brafman. An MDP-based recommender system. *Journal of Machine Learning Research*, 6:1265–1295, Dec. 2005. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1046920.1088715.

Y. Sun, N. J. Yuan, X. Xie, K. McDonald, and R. Zhang. Collaborative nowcasting for contextual recommendation. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 1407–1418, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4143-1. doi: **10.1145/2872427.2874812**. URL http://dx.doi.org/10.1145/2872427.2874812.

References IV

P. Symeonidis, A. Papadimitriou, Y. Manolopoulos, P. Senkul, and I. Toroslu. Geo-social recommendations based on incremental tensor reduction and local path traversal. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 89–96, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1033-8. doi: **10.1145/2063212.2063228**. URL http://doi.acm.org/10.1145/2063212.2063228.

C.-H. Tai, D.-N. Yang, L.-T. Lin, and M.-S. Chen. Recommending personalized scenic itinerary with geo-tagged photos. In *IEEE International Conference on Multimedia and Expo*, pages 1209–1212, June 2008. doi: **10.1109/ICME.2008.4607658**.

C.-Y. Wang, Y.-H. Wu, and S.-C. T. Chou. Toward a ubiquitous personalized daily-life activity recommendation service with contextual information: A services science perspective. *Information Systems and E-Business Management*, 8(1):13–32, January 2010.

Z. Xing, J. Pei, and E. Keogh. A brief survey on sequence classification. *SIGKDD Explor. Newsl.*, 12(1):40–48, Nov. 2010. ISSN 1931-0145. doi: **10.1145/1882471.1882478**. URL http://doi.acm.org/10.1145/1882471.1882478.

H. Yoon, Y. Zheng, X. Xie, and W. Woo. Smart itinerary recommendation based on user-generated GPS trajectories. In *Proceedings of the 7th International Conference on Ubiquitous Intelligence and Computing*, UIC'10, pages 19–34, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-16354-8, 978-3-642-16354-8. URL http://dl.acm.org/citation.cfm?id=1929661.1929669.

References V

H. Yoon, Y. Zheng, X. Xie, and W. Woo. Social itinerary recommendation from user-generated digital trails. *Personal Ubiquitous Comput.*, 16(5):469–484, June 2012. ISSN 1617-4909. doi: **10.1007/s00779-011-0419-8**. URL http://dx.doi.org/10.1007/s00779-011-0419-8.

V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI 2010*. Association for Computing Machinery, Inc., July 2010. URL http://research.microsoft.com/apps/pubs/default.aspx?id=143146.

V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Towards mobile intelligence: Learning from GPS history data for collaborative recommendation. *Artif. Intell.*, 184-185: 17–37, June 2012. ISSN 0004-3702. doi: **10.1016/j.artint.2012.02.002**. URL http://dx.doi.org/10.1016/j.artint.2012.02.002.

Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 791–800, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. doi: **10.1145/1526709.1526816**. URL http://doi.acm.org/10.1145/1526709.1526816.

Some Extra Bits

Trend across matching-unit N

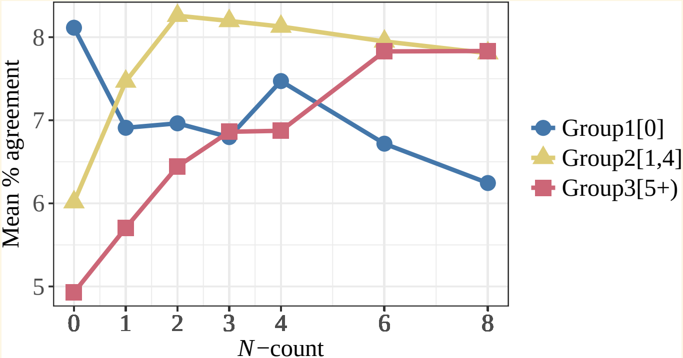
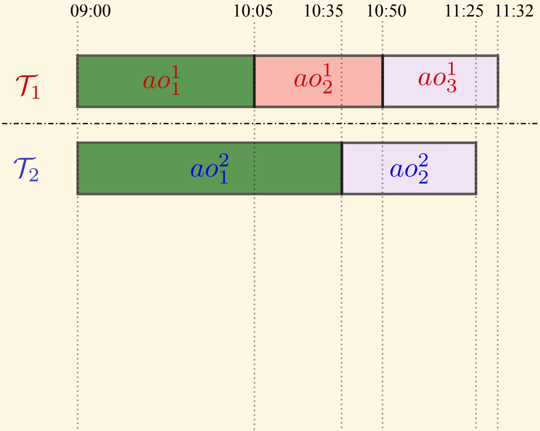


Figure: MRR versus matching unit for three user groups for first sequence index.

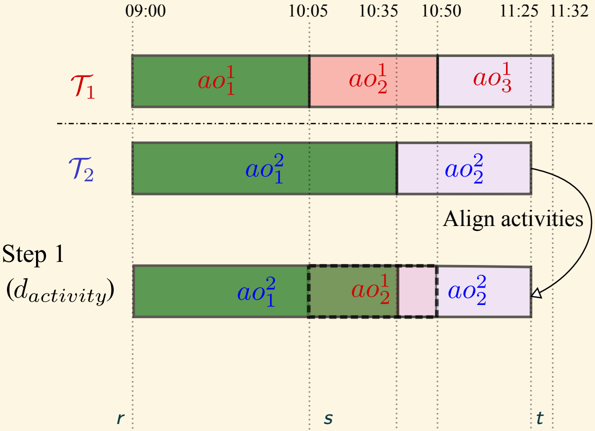
Two-level Distance

- Inspired by edit distance
- Adapted for sequence of objects



Two-level Distance

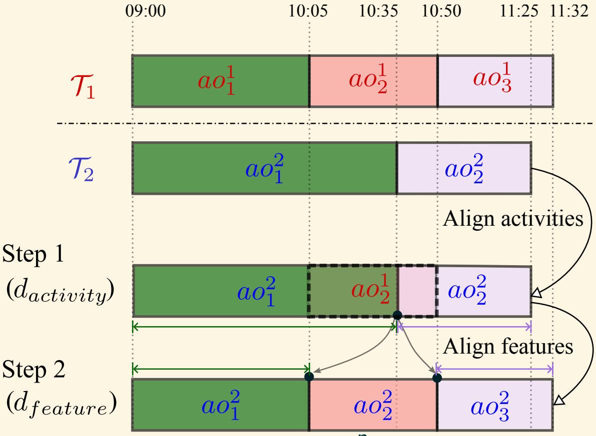
- Inspired by edit distance
- Adapted for sequence of objects



$$d_{activity}(\mathcal{T}_1, \mathcal{T}_2) = \sum_{i=1}^{r} w_{obj} \times c_{ins} + \sum_{j=1}^{s} w_{obj} \times c_{del} + \sum_{k=1}^{t} w_{obj} \times c_{sub}$$

Two-level Distance

- Inspired by edit distance
- Adapted for sequence of objects



$$d(\mathcal{T}_1, \mathcal{T}_2) = d_{activity}(\mathcal{T}_1, \mathcal{T}_2) + \sum_{i=1}^{n} d_{feature}(ao_i^1, ao_i^2) \ .$$

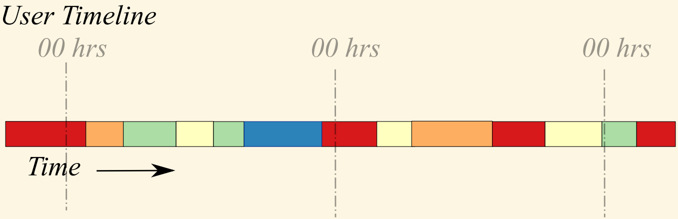
Methodology

Learning Optimal Matching Unit range

- **Wrapper attribute selection**: C4.5 algorithm, greedy backward search and area under ROC curve as evaluation measure.
- **Classification**: pruned attribute vectors for each user fed into a C4.5 induction algorithm to predict optimal matching unit range.

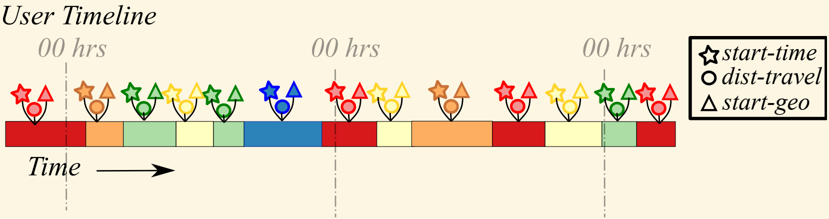
Attribute Extraction: Timeline Decomposition

- Each user represented by an attribute vector.
- For attribute extraction:
  timelines are decomposed into features-sequence :



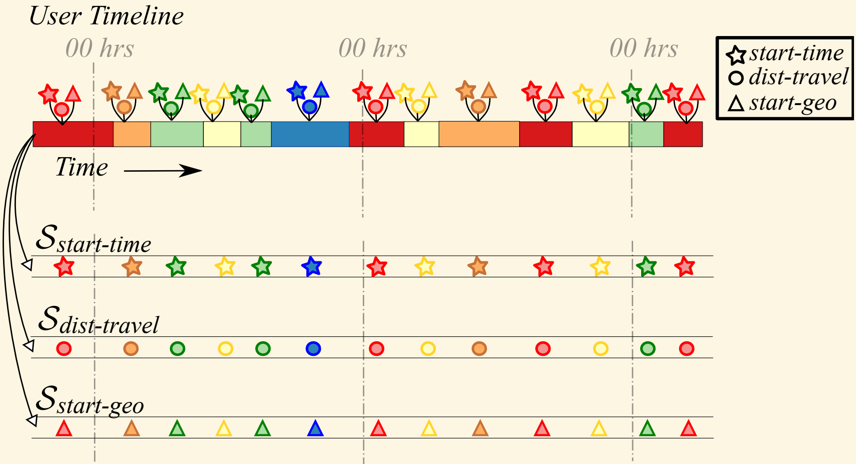
Attribute Extraction: Timeline Decomposition

- Each user represented by an attribute vector.
- For attribute extraction:
  timelines are decomposed into features-sequence :



Attribute Extraction: Timeline Decomposition

- Each user represented by an attribute vector.
- For attribute extraction:
  timelines are decomposed into features-sequence :



Timeline Attributes

Timeline Attributes

- Regularity
- Repetition

Timeline Attributes

Regularity Attributes: Sample Entropy

1. $SampEn_z^p$: **sample entropy** of a feature sequence $\mathcal{S}_z$ for epoch length p,
2. $\mu SampEn_{\mathcal{T}}^p$: **mean sample entropy** over all feature sequences $\mathcal{S}_z$, $z = 1, 2, ..., m$ of timeline $\mathcal{T}$ for epoch length p,
3. $\sigma SampEn_{\mathcal{T}}^p$: **standard deviation of sample entropy** over all feature sequences $\mathcal{S}_z$, $z = 1, 2, ..., m$ of timeline $\mathcal{T}$ for epoch length p.

Timeline Attributes

Repetition Attributes: k-gram attributes

Previously used for sequence classification, biological sequence analysis and text classification [Xing et al., 2010; Dong and Pei, 2007].

1. η_z^k: **total number of distinct k-grams** in feature sequence $\mathcal{S}_z$, normalised by total number of k-grams occurring in $\mathcal{S}_z$,
2. μf_z^k: **mean frequency of occurrence** of distinct k-grams in feature sequence $\mathcal{S}_z$, normalised by total number of k-grams occurring in $\mathcal{S}_z$,
3. σf_z^k: **standard deviation of frequency** of occurrence of distinct k-grams in feature sequence $\mathcal{S}_z$, normalised by length of $\mathcal{S}_z$.