Translation-based Recommendation

October 28, 2017

1 Problem being addressed

Sequential recommendation with third order interaction: $< u, i_t, i_{t+1} >$

2 Key ideas and contributions

- Unified method *TransRec* to capture the third order interactions in one space (contrary to PRME that has a sequential transition space and user preference space, which are inherently correlated.)
- Uses metric embeddings (TransE) just like PRME, but incorporates knowledge graph concepts to unify as one space. Each *user* is now a translation vector operating on item sequences. The personalized translation operation is:

$$\vec{\gamma_i} + \vec{T_u} \approx \vec{\gamma_j}$$

over a transition space $\Phi = \mathbb{R}^K$, $\vec{\gamma_i}$ is the latent (embedding) item vector, $\vec{T_u} = \vec{t} + \vec{t_u}$ as a sum of global translation vector and translation vector associated with user u.

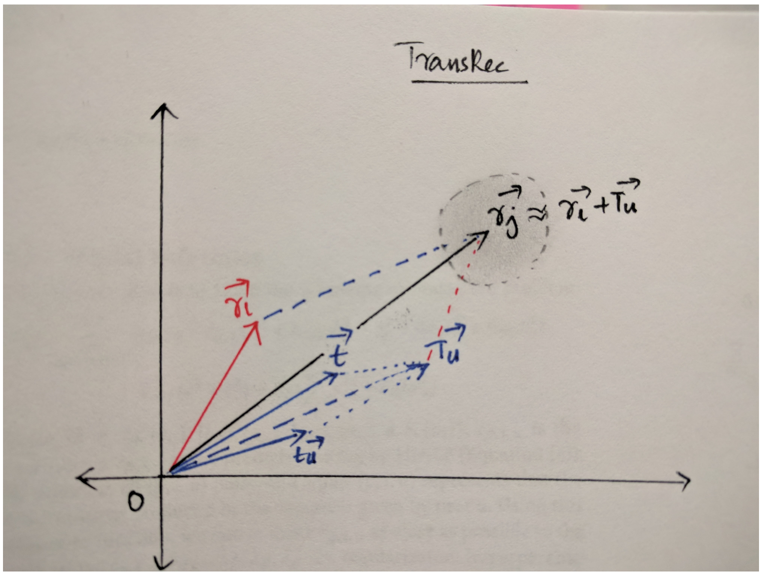

Figure 1: User transition from item i to another item j

- $\vec{\gamma_j}$ is the nearest neighbor of $\vec{\gamma_i} + \vec{T_u}$ in Φ. If user transitions from A to B,C, TransRec puts B close to C.

- $P(j|u,i) \propto \beta_j - d(\vec{\gamma_i} + \vec{T_u}, \vec{\gamma_j})$,
  $s.t.\ \vec{\gamma_i} \in \psi \subseteq \Phi$ and β_j is the item popularity.
  If the item j is more popular, and the distance to the previous item i is small, probability of predicting increases massively. If the item j is less popular, and the distance to the previous item is still small, this may depend on relative ranking with the rest of products.

- ψ could be a unit $\mathcal{L}_2$ ball used as a regularizer i.e. reduce dimensionality.

- **Optimization** - S-BPR with stochastic gradient ascent. Needs renormalization to be within the constrained space.
  Training data - Amazon, Epinions, Google, FourSquare, Flixter
  Evaluation metric - AUC, Hit Rate @50 (@10 for item-to-item recommendation)
  Baselines PopRec, BPR-MF, FMC, FPMC, PRME, HRM

- Incorporate content like item features (eg: review text) through an additional embedding to project it to the embedding space for the problem of *item-item recommendation*. Baselines include WNN, LMT, Monomer.

3 Strong points

- Implicit metricity assumption of triangular inequality is maintained
- Single space handles the third order interaction.
- Handles large sequences simplistically, but limited to first order item interaction in the sequence.
- Handles cold-start with some global translation vector (averaged behaviour).
- Handles vast variability wrt datasets like Google Local that has large number of categories.
- Gives empirical evidence as to the strength of the model for sparse datasets. The transrec model only performs subpar in the case of dense datasets where PRME performs supremely with the large number of parameters in the model.

4 Weak Points

- *Directionality* : The authors mention that the model correctly predicts a tripod after a sequence of camera-related products. However, this does not prevent the occurrence of predicting a camera after a tripod, as the distance metric is Euclidean and considers the items as points.
- Does not perform strongly with dense datasets unless the dimensionality of the model is increased. (as it uses fewer parameters than PRME to uncover the dense relationships.)

5 Extensions

- Weak/ Absent notion of directionality can be improved. Check out other metric spaces like torus? Maybe manifolds?
- Explicit handling of short term and long term dynamics, though in one space.

- Study the use of different types of embeddings including MLPs.
- Temporal dynamics and content-based recommendation.
- Incorporating category in a more direct way (for better cold-start recommendation)

6 Questions

1. Is directionality handled by the vector addition? $\vec{\gamma_i} + \vec{T_u} \approx \vec{\gamma_j}$ But now, $\vec{\gamma_i} \approx \vec{\gamma_j} - \vec{T_u}$.
   It does not matter because with item recommendation, we are looking at the nearest of neighbor of $\vec{\gamma_i} + \vec{T_u}$ using Euclidean distance metric. Item i can fall within this region as well.
2. Why is the sequential transition space and the user preference space correlated in PRME?
   User interacts with items, and also moves from one item to another item. The same user crawls through two different spaces that model these separately. They are interdependent.
3. How strong is the case authors make about TransRec handling short term dynamics?
   $\vec{T_u}$ takes the user from the current item to the next item (short term dynamics) as well as maintains the user's long term preferences. Being the same vector, I feel either one of them is compromised to an extent.

7 Misc

Easy read. Uses various metric space concepts for efficiency.
ACM ref:
Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based Recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17). ACM, New York, NY, USA, 161-169. DOI: https://doi.org/10.1145/3109859.3109