Contexts Embedding for Sequential Service Recommendation

Benjamin Amankwata and Kenneth K. Fletcher

Computer Science Department

University of Massachusetts Boston

Outline

- Motivation
- Contributions
- Model
  - Higher Order Latent Interactional Context (HOLIC).
  - Deep Embedded Clustering
  - Neural Attention
- Experiments and Results
- Conclusion and Future Work

Motivation

1. Context labels must be pre-defined and require additional effort to augment the training dataset.

2. Context attribute labels may not be available for each data point resulting in data sparsity and dimensionality expansion problems.

3. Pre-defined context attributes may not reflect the optimal dynamics for the recommendation process.

Contributions

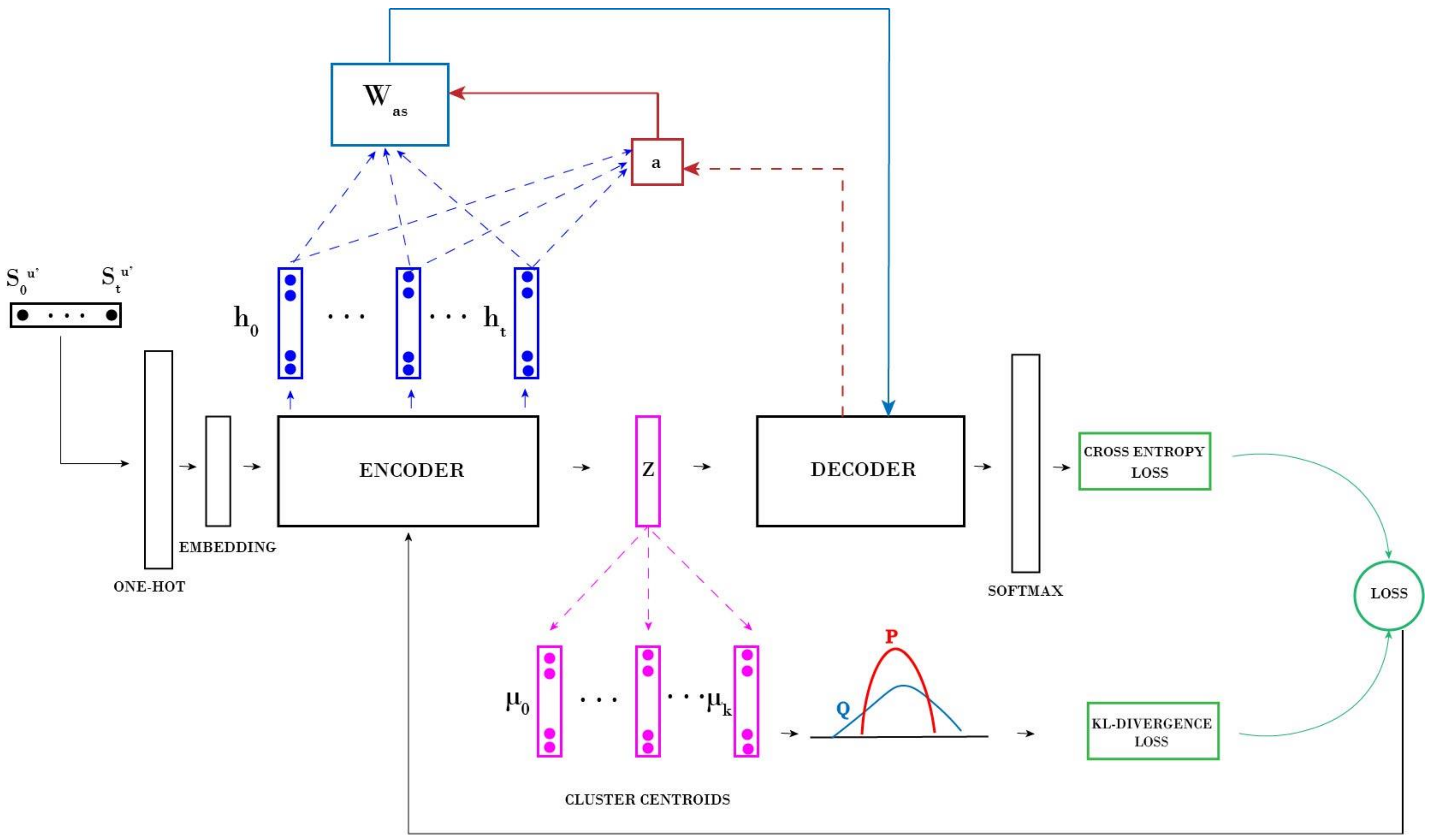
1. Inspired by HAC, we propose Higher-Order Latent Interactional Context (HOLIC), a novel formulation of interactional context.

2. We propose a novel model that leverages HOLIC, and neural attention for improved sequential service recommendation. [1]

3. We perform experiments to evaluate and validate our proposed models, using four real-world online services datasets.

1. To the best of our knowledge, this is the first work to unify interactional and representational context in a single model for sequential recommendation.

Model: Overview



Model: Hyperspace Analogue to Context (HAC)

Hyperspace Analogue to Context (HAC) :

- context as multi-dimensional space.
- captures continuously changing information from various context sources.

HAC = $\langle D_1, D_2, D_n \rangle$, is an n-dimensional space.
Each dimension D_i describes a data type and value set for a specific type of context
The dimensions in HAC are heterogeneous; i.e., values can be of different types e.g., categorical, discrete etc.

Our model is inspired by HAC which offers a valuable framework for thinking about modeling latent context where the context types may not be interpretable and structure unknown a priori.

Model: Higher Order Latent Interactional Context (HOLIC).

Given an input sequence and no additional context information:

1. We seek to learn a k-dimensional context vector **c**
   Such that each dimension $\mathbf{c}_i$ represents the relative contribution of a discrete context source

Our method is as follows:

1. Map input sequence to d-dimensional latent space
2. Learn a set of cluster centroids in latent space
3. Determine proximity of each input sequence to cluster centroids

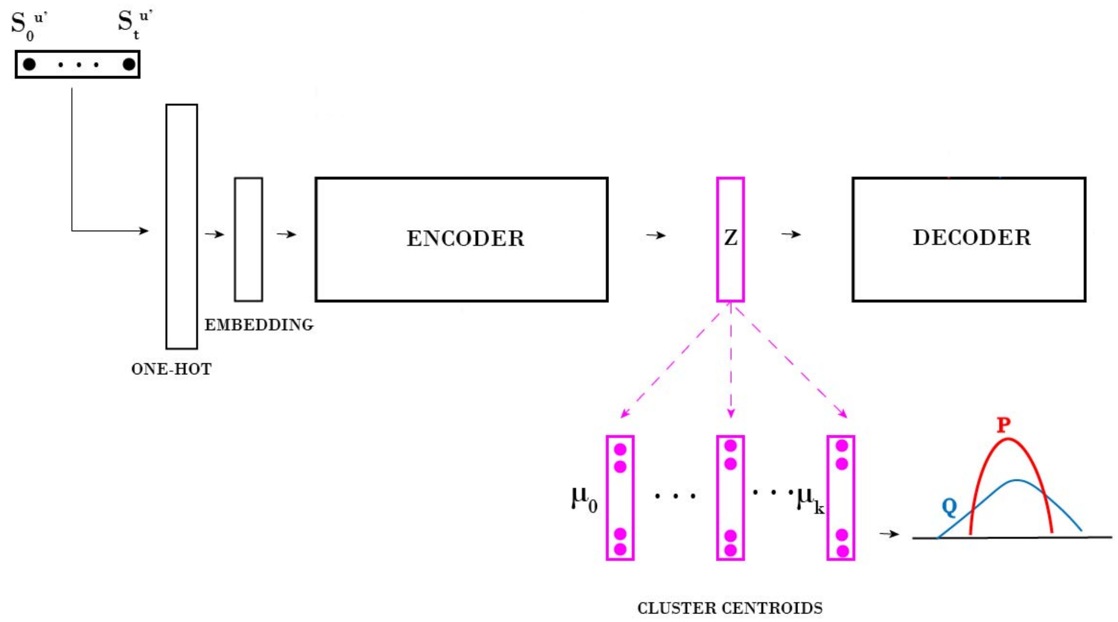
We call **c** a Higher Order Latent Interactional Context (HOLIC).

Hypothesis: We argue that for sequences which are contextually dependent, we expect that representations in latent space will cluster around significant context sources. In that vein, **c** is analogous to a point in HAC.

Model: Deep Embedded Clustering

Learn a suitable representation of an input with a DNN and subsequently apply a clustering technique to the learned input representation. Done in two stages:

Stage 1: Input X is transformed by a non-linear mapping
$f_\theta : X \rightarrow Z$ into a latent feature space Z, where θ are learnable parameters
Stage 2: A set of k centroids in the feature space Z are learned
The Student's t-distribution is used as a kernel to measure the similarity between an embedded point and a centroid



Model: Deep Embedded Clustering

We use the students t-distribution as a kernel to measure the similarity between the sequence representations and cluster centroids

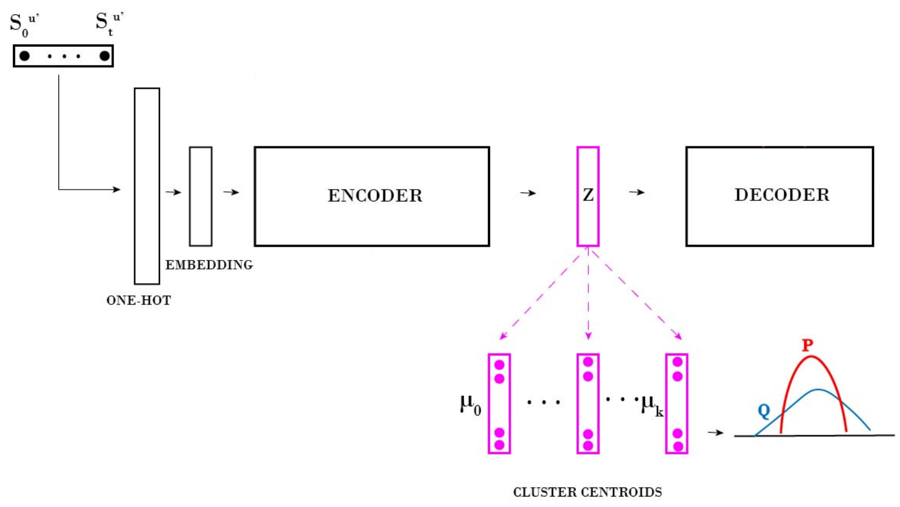
Given, $z_i \in Z$, an input representation in latent space

$\mu_j \in \{\mu_j\}_{j=1}^{k}$, a set of cluster centroids

Where $z_i, \mu_j \in R^l$ and $l < d$.

l is the latent space dimension

k is the number of cluster centroids,



$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'}(1 + \|z_i - \mu_{j'}\|^2/\alpha)^{-\frac{\alpha+1}{2}}},$$

$$f_j = \sum_i q_{ij}$$

$$p_{ij} = \frac{q_{ij}^2/f_j}{\sum_{j'} q_{ij'}^2/f_{j'}},$$

$$L = \mathrm{KL}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Model: Decoder (Neural Attention)

Decoder conditional probability

$$p(s^{u'}_{t+1} \mid H, z) = \omega(e_{sos}, z, w_{as})$$

Weighted alignment score

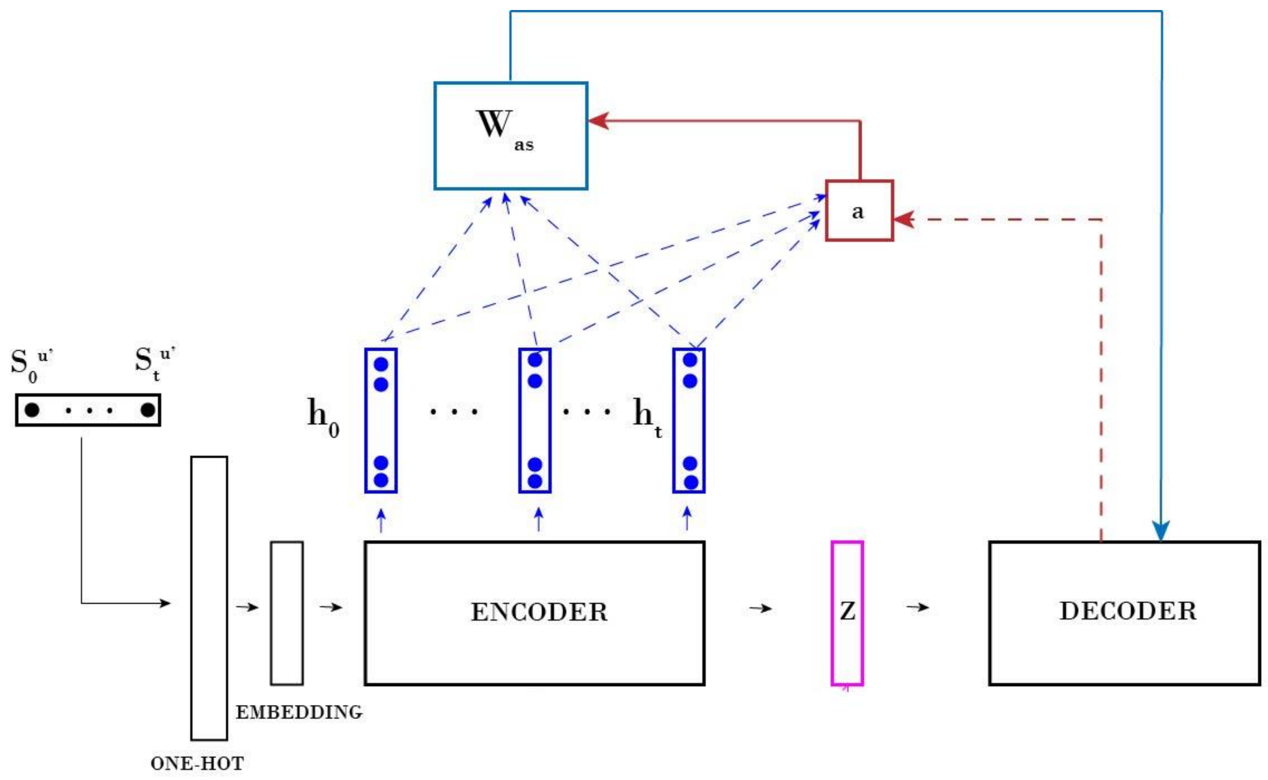
$$w_{as} = \sum_{j=1}^{t} \beta_j h_j .$$

Weight of each encoder hidden state

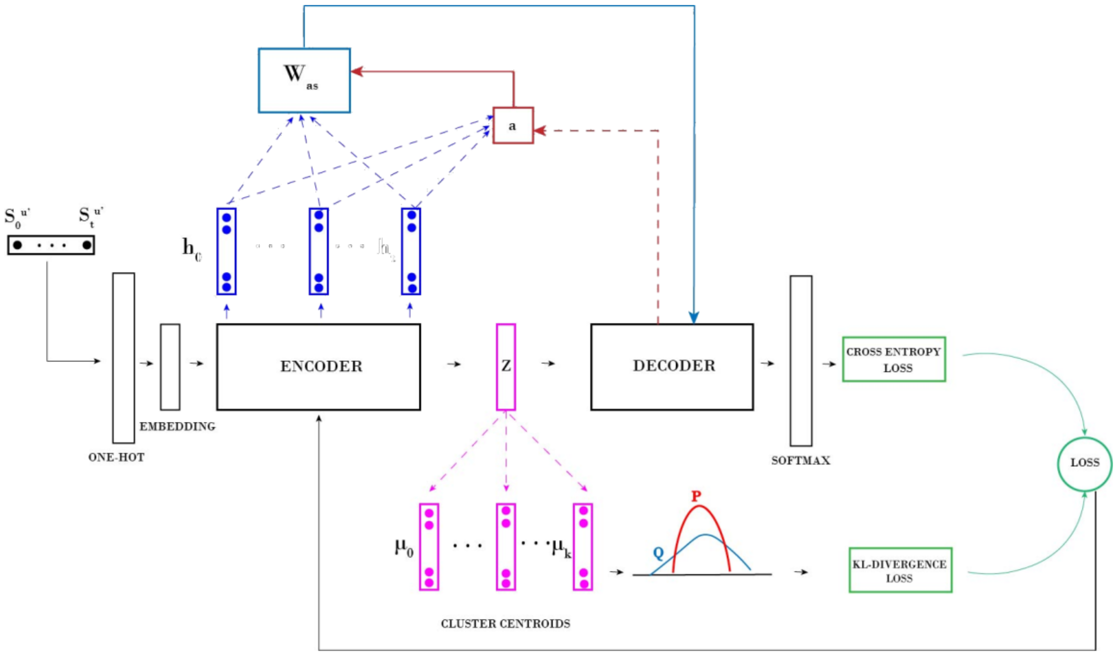
$$\beta_j = \frac{\exp(\Phi_j)}{\sum_{k=1}^{t} \exp(\Phi_k)}$$

$$\Phi_j = a(z, h_j)$$

Where *a* is a Feed Forward Network



Model: Network Training



$$\max_{\theta} \frac{1}{N} \sum_{n=1}^{N} \left[\log W_{\theta} \left(s_{t+1}^{u^n} \mid H_n, z_n \right) + \gamma \sum_{i} \sum_{j} p_{n_{ij}} \log \frac{p_{n_{ij}}}{q_{n_{ij}}} \right]$$

Experiments: Datasets

TABLE I: Meta-Data of our Datasets used in our Experiments.

Dataset	No. of users	No. of items	No. of interactions
YELP	478,630	7,239	2,714,935
ML25M	130,004	7,984	8,895,228
4SQUARE	227,301	14,153	7,183,758
LASTFM-1k	414,854	13,430	5,242,483

YELP: A subset of Yelp's businesses, reviews, and user data across 8 metropolitan areas in the USA and Canada

4SQUARE: Global check-in data from Foursquare. It contains check-ins users on cities across 77 countries.

MovieLens 25M: User-Movie interaction activity from a movie recommendation service

LASTFM 1K: Whole listening habits for nearly 1,000 users.

Experiments: Hyperparameters

- **K**: Number of cluster centroids
- γ : scaling factor of KL divergence loss
- **α** : degree of freedom of Student's t-distribution, α,

TABLE IV: Optimal K across datasets

Dataset	Optimal K
YELP	1000
LASTFM-1K	250
ML25M	500
4SQUARE	250

Experiments: Results

TABLE II: Results for MRR@20

	MRR@20			
Dataset	Transformer	RNN	ATTENTION	Our Model
YELP	0.028	<u>0.036</u>	0.034	**0.041**
LASTFM-1K	0.122	0.223	0.224	**0.244**
ML25M	0.030	<u>0.057</u>	0.043	**0.061**
4SQUARE	0.096	0.201	<u>0.203</u>	**0.217**

TABLE III: Results for RECALL@20

	RECALL@20			
Dataset	Transformer	RNN	ATTENTION	Our Model
YELP	0.111	<u>0.122</u>	0.117	**0.132**
LASTFM-1K	0.273	0.374	<u>0.378</u>	**0.402**
ML25M	0.114	<u>0.18</u>	0.15	**0.183**
4SQUARE	0.316	0.469	<u>0.473</u>	**0.486**

Conclusion

- In this work, we proposed:
  - Higher-Order Latent Interactional Context (HOLIC), a novel formulation of interactional context, for a context-aware sequential recommendation.
  - A sequential service recommendation method that leverages HOLIC, representational context, and neural attention mechanism for improved sequential recommendation.
- We compare our model to other Recurrent Neural Network and transformer architectures on four benchmark datasets for a next-item recommendation.
- We demonstrate the superiority of our approach to the baselines in next-item recommendation.

Future Work

- Investigate the internal structure of interactional context with a particular focus on delineating probable hierarchical relationships between context sources within the HOLIC framework.

- Explore analytical approaches to optimizing the number of cluster centroids in our formulation of HOLIC.

References

[1] E. Smirnova and F. Vasile, “Contextual sequence modeling for recommendation with recurrent neural networks,” Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, 2017.
[2] M. Unger, A. Bar, B. Shapira, and L. Rokach, “Towards latent contextaware recommendation systems,” Knowl. Based Syst., vol. 104, pp. 165– 78, 2016.
[3] A. Livne, M. Unger, B. Shapira, and L. Rokach, “Deep context-aware recommender system utilizing sequential latent context,” ArXiv, vol. abs/1909.03999, 2019.
[4] M. Unger, L. Rokach, A. Bar, E. Gudes, and B. Shapira, “Contexto: lessons learned from mobile context inference,” Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, 2014.
[5] F. Li, K. Rasch, S. Sehic, S. Dustdar, and R. Ayani, “Unsupervised context-aware user preference mining,” in AAAI 2013, 2013.
[6] K. Rasch, F. Li, S. Sehic, R. Ayani, and S. Dustdar, “Context-driven personalized service discovery in pervasive environments,” World Wide Web, vol. 14, pp. 295–319, 2011.
[7] J. Xie, R. B. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” ArXiv, vol. abs/1511.06335, 2016.
[8] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” CoRR, vol. abs/1409.0473, 2015.
[9] B. Kwapong and K. Fletcher, “A knowledge graph based framework for web api recommendation,” in 2019 IEEE World Congress on Services (SERVICES), vol. 2642. IEEE, 2019, pp. 115–120.
[10] B. A. Kwapong, R. Anarfi, and K. K. Fletcher, “Collaborative learning using lstm-rnn for personalized recommendation,” in International
Conference on Services Computing. Springer, 2020, pp. 35–49.
[11] R. Anarfi and K. K. Fletcher, “A reinforcement learning approach to web api recommendation for mashup development,” in IEEE World Congress on Services (SERVICES), vol. 2642. IEEE, 2019, pp. 372–373.