Middle East Technical University
Department of Computer Engineering

Hierarchical Context Modeling Using Incremental Deep Boltzmann Machines

Fethiye Irmak Doğan, Sinan Kalkan

METU-CENG-TR-2017-01

March 2017

Department of Computer Engineering
Middle East Technical University
Dumlupınar Bulvarı, 06531, Ankara
TURKEY

Technical Report

This page contains a Turkish translation of the title and the abstract of the report. The report continues on the next page.

Artırımlı Derin Boltzmann Makineleri ile Bağlamın Hiyerarşik Modellenmesi

Fethiye Irmak Doğan, Sinan Kalkan

Bilgisayar Mühendisliği Bölümü
Ortadoğu Teknik Üniversitesi
Dumlupınar Bulvarı, 06531, Ankara
TÜRKİYE

Öz

Bu raporda, bağlam üzerine yaptığımız (i) bağlamın hiyerarşik doğasına yönelik irdeleme, (ii) bağlamı modelleme çalışmaları hakkında inceleme ve (iii) bu bilgiler doğrultusunda tasarlamakta olduğumuz artırımlı bir derin mimari kullanarak bağlamı hiyerarşik modelleme araştırmalarımızı sunmaktayız.

Abstract

In this report, we describe our research on (i) one of the unsolved issues in context: Whether context is hierarchical or not, (ii) methods that are used for modeling context and (iii) our prospective incremental deep architecture to model context hierarchically.

1 Introduction

Context is very important for human cognition affecting many of our capabilities, including perception, recognition, reasoning, problem solving, action, language etc. [25, 30].

Even though context is very significant for human and artificial cognition, its structure remains unsolved and the results of studies which analyze its structure will make crucial contributions to studies in many areas. Therefore, we started our studies by examining whether its structure is hierarchical or not.

Since our analyses stated that the context has a hierarchical structure, we continued with observing existing studies to model context and its hierarchical structure. We noticed that topic modeling methods are very suitable for modeling context and it is commonly used in the literature [3]. Moreover, context is modeled in an incremental way in the study of Çelikkanat et al., since robots encounter each scene one by one. We intend to take the same approach as we are also planing to use context in robotics.

After analyzing the methods that are used for modeling context, we continue to design our prospective algorithm, where we are planing to use incremental Deep Boltzmann architecture. We claim that the hierarchical and generative architecture of deep Boltzmann Machines is suitable for modeling hierarchical nature of context and some existing studies already used Deep Boltzmann Machines (DBM) for topic modeling [22]. Moreover, we are planing to use the algorithm proposed by Çelikkanat et al. to introduce new contexts [3], which will correspond to neurons of the first hidden layer of our incremental Deep Boltzmann Machine.

2 Related Work

2.1 Hierarchical nature of context

To obtain conclusions about context structure, the properties of context should be examined. In a study where Zimmermann tried to provide a definition for context, the existence of social, spatial and temporal aspects are mentioned in the definitions of context [32]: Social aspects include e.g., being in a family or friendly environment, whereas spatial and temporal aspects correspond to contextual information owing to space and time.

Analyzing the social, spatial and temporal properties of context, one can deduce that contexts should include or be related to each other. For example, the context of "preparing breakfast" includes "family" context as a social aspect, "kitchen" as a spatial aspect and "morning" as a temporal aspect. Since contexts include each other, they have a hierarchical structure – see Figure 1 for an example hypothetical hierarchy of context.

One of the important pioneers who defined and used context in artificial intelligence is McCarthy. McCarthy, in addition to his other contributions in formulating and modeling context, described the following three properties of context [14]:

- Contexts are formal objects.
- Contexts can be completely described.
- There is a useful link within contexts and contextual functions, and a new context can be obtained from an old one by changing some conditions such as place, time or situation.

Based on McCarthy's descriptions, it can be stated that contexts can dynamically vary, one context can be reached from another context, one context can enter the scope of another context, and relational structures exist between contexts. In fact, such relational structures that link contexts of varying scopes might be the most efficient solution (if not the only one) for handling these complex structural descriptions and dynamic structures of context [10]. In short, more than

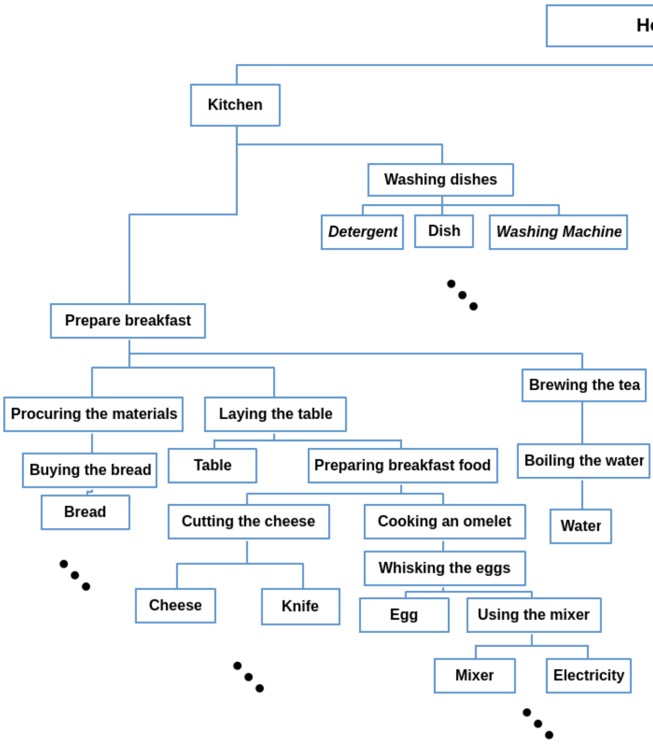

Figure 1: Hierarchic representation of home context.
(For the sake of space, a small portion of the hierarchy is shown)

20 years ago, McCarthy stated certain aspects and properties of context, suggesting containment of a context in another, i.e., a hierarchical structure.

Another important pioneer on context is Barsalou, though from a different discipline. Barsalou has highlighted the importance of context on human cognition (see, e.g., [30]): The effect of context on cognition is explained in terms of cognitive process examples such as episodic memory, object perception and language comprehension. Moreover, he stated that concept situations is affected by "grain size", which means that a situation can continue from a large space with huge amount of time to a tiny space with small amount of time. He also said that hierarchical combination of situations creates context by including many levels of grain size and this makes context hierarchical.

Nature, society and lots of phenomena in our world have multilevel structures, organized and structured hierarchically intrinsically or intuitively. Hierarchical working structure of the hippocampal system [15] is just an example of hierarchical working structure of the human body. Since it plays an important role for contextual formation [5, 17, 21], a support from neuroscience to the hierarchical structure of context can be concluded. Moreover, hierarchical nature of decision making processes [18] is also a demonstration of the hierarchical nature of human psychology.

Hierarchy is eminent in other disciplines as well. For example, Lane claim in his work on the hierarchy that human belonging social hierarchies which consist of society, culture, and economy has more ambiguous nature than physic-chemical (i.e. hierarchy between fundamental particle, nucleus, atom and molecule) and biological (i.e. hierarchy between organelle, cell, organ, multicellular creature, population, specify and ecosystem) hierarchies [11].

Since hierarchy is encountered at such an extent in many different disciplines and areas related to our world and humans, it is a natural result that context has a hierarchal structure owing to its social, spatial and temporal aspects linked to all these areas. This becomes more eminent considering that our lives are dominated with social, spatial and temporal hierarchies, and context having these aspects [14, 32] should be inevitably hierarchical.

Moreover, there are some hierarchical modeling efforts in computer vision for including context [4, 13, 23, 28, 29]. However, those efforts lack crucial aspects of context defined by McCarthy.

Table 1: Correspondence between context modeling and topic modeling

Context Modeling	Topic Modeling
a single scene	each document
all encountered scenes	corpus
each concept	each word in the vocabulary
context	topic

In spite of all these limitations, context has been shown to modulate many critical processes significantly, improving the performance of computational models in many challenging problems like object recognition and planning (see, e.g., [1, 3, 9]). We suggest that to make the best of contextual modulation in a world dominated by ambiguities, computational efforts on context should model context hierarchically.

2.2 Topic Modeling

Topic Modeling can be described as searching for short representations of huge collections without loosing their important statistical connections [2]. One of the most successful methods for topic modeling is Latent Dirichlet Allocation (LDA), which is a generative probabilistic model. In this approach, topics are represented by distributions over words and documents can be thought as distributions over topics.

Topic Modeling methods can be used for modeling context [3]. In that kind of approach, each scene should be regarded as a document, concepts can be seen as words in the vocabulary and contexts correspond to topics. The correspondence between modeling topics and context can be seen in Table 1.

Topic modeling approaches such as LDA needs to be given a predetermined number of topics. However, at the beginning, it is impossible for us to determine the number of topics/contexts. Some extensions of LDA such as Hierarchical Dirichlet Process (HDP) [24] handled the fixed number of topic problem by assuming infinite number of topics but they are still not suitable for modeling hierarchical data.

2.3 Hierarchical Topic Modeling

Since our observations state that context has a hierarchical structure, we need a topic modeling method which is suitable for modeling hierarchical documents.

Nested Chinese Restaurant Process (nCRP) is a model to learn topics and organize this topics in a hierarchical way [6]. In this model, each node corresponds to a topic and these topics can be considered as a distribution over words. A path from root to leaf is chosen to generate a topic. This model assumes fixed depth for hierarchy and it is almost impossible for us to know this depth at the beginning. Moreover, this model associates each document to a single path and documents could be associated with more than one path since they can be mixture of topics and do not have to belong a single topic.

Nested Hierarchical Dirichlet Process (nHDP) is also a hierarchical topic model to construct representation of text data by making the topics more general to more specific from root to leaf nodes [16]. This model is an extension of nCRP and it can handle the problem of following a single path for each document in nCRP by performing word-specific path clustering rather than document-specific paths. Therefore, a document has access to an entire tree and topics can be mixture of different paths. Path selection difference between nCRP and nHDP can be observed in Figure 2. However, this model still suffers from assuming the fixed depth for the hierarchy and fixed number of nodes for each layer.

In other study, Smith et al. proposed a hierarchical topic model that uses a top-down recursive approach which is based on splitting and re-modeling a corpus [20]. This recursive splitting and re-modeling process continues until documents and corpus become too small to model. However, this study assumes pre-determined number of nodes for each layer in the hierarchy.

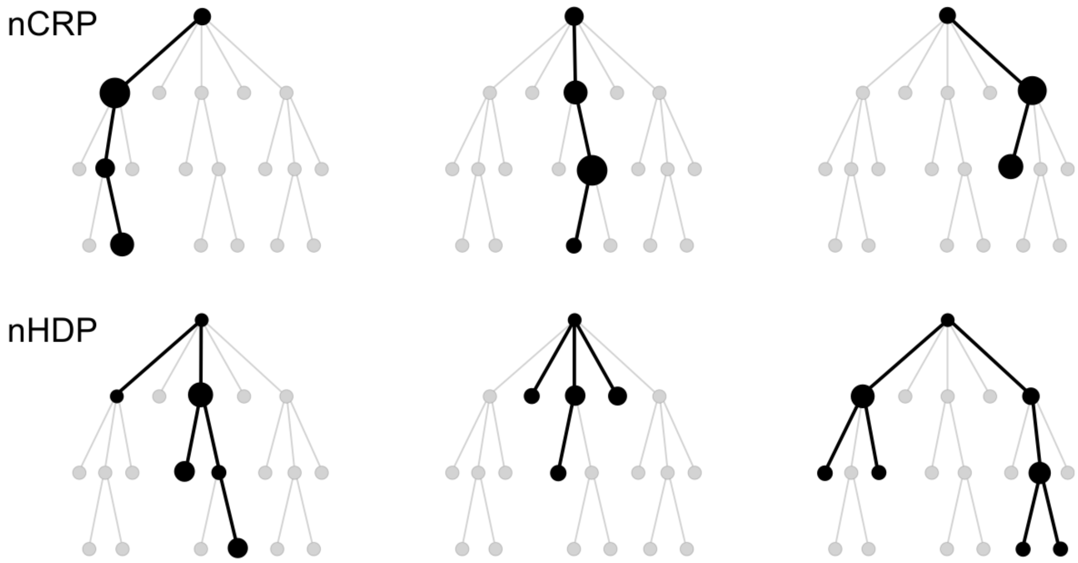

Figure 2: An example of path structures for the nested Chinese restaurant process (nCRP) and the nested hierarchical Dirichlet process (nHDP) for hierarchical topic modeling.
(Figure source: [16])

In another study, Seiter et al. proposed a hierarchical nonparametric topic modeling approach by using joint segmentation and activity recovery approach and they combined Chinese Restaurant process and distance dependent Chinese Restaurant process for this purpose [19]. This approach does not require any parameter segment size or number of topics selection. However, this method is not incremental, i.e., it is not adjustable for upcoming new data.

In an interesting study, Wang et al. proposed a hierarchical topic model which is based on fixed height and a node number, then after this fixed construction, hierarchy can be modified by user interactions [26]. It means that a top-down topical hierarchy can be manipulated by user operations such as merging, removing or branching topics which enables user to control the intermediate results and complexity of the hierarchy. However, this model is not suitable for constructing hierarchical topic model automatically in terms of height and node based on data.

Another interesting study belongs to Zavitsanos et al. [31], which proposes to construct a bottom-up hierarchy by modeling the vocabulary on leaf nodes and regarding topics in intermediate nodes as multinomial distributions over subtopics. This model does not require any assumption about the depth of hierarchy or number of nodes in each layer but it is not suitable for adjusting structure for new coming data, i.e. incremental learning.

2.4 Incremental Topic Modeling

Since we cannot anticipate all contexts that an agent is expected to encounter during its lifetime, context should be learned incrementally.

There are promising studies that models the context in an incremental way [3]. In the study of Çelikkanat et al., they proposed a model that computes number of topics for each encountered scene. They initialize topic count as 1, calculate topic probability distributions over words by using LDA and compute the confidence values for each topic. If confidence value of any topic is less than a threshold value (they call it C_{low}), they increment topic count and re-compute the probability distributions for topics. This process continues until no confidence value of a topic is less than a threshold (i.e until $C_{low} = \emptyset$). Then, algorithm waits for a new scene to continue the same process. Their incremental-LDA algorithm can be observed in Figure 1

Algorithm 1 Incremental Latent Dirichlet Allocation algorithm (Source: [3])

initialize context count $K \leftarrow 1$.
for all encountered scenes **do**
  run K-Incremental Gibbs sampler with K
  while $\mathbb{C}_{low} \neq \emptyset$ **do**
    increment context count $K \leftarrow K + 1$
    run K-Incremental Gibbs sampler with K
  end while
  output converged context assignments $\vec{z}_N$ for the scene
end for

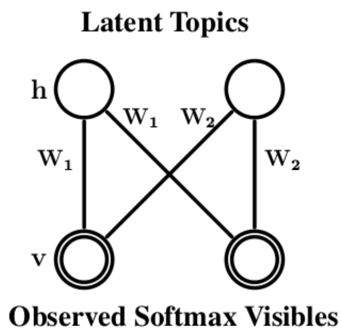

Figure 3: Replicated Softmax Model
(Figure source:[7])

2.5 Incremental Hierarchical Topic Modeling

Some studies already focused on building hierarchical structure in an incremental way.

In a study of Hu et al., they proposed two phased method for incrementally learning of topic hierarchies [8]. In the first phase, they recursively construct a topic hierarchy, then in the second phase, they applied an incremental hierarchical topic alignment algorithm to incrementally merge topic hierarchies through top-down hierarchical topic alignment. They compute a similarity matrix between subtopics and merge them if the similarity is larger than a threshold. However, in this study, they set set maximum level of hierarchy as three and the number of topics need should be specified which are the drawbacks of the model.

In another study, Wang et al. proposed an online hierarchical topic clustering algorithm called evolving hierarchical Dirichlet processes (EHDP) where clusters may be born, evolve, branch and die-out over time [27]. Moreover, their model is based on online learning so they can update the evolutionary traces of a cluster incrementally. Chinese Restaurant Processes and Hierarchical Dirichlet Processes are used in the construction phase of EHDP where the model inference is made by Gibbs sampling. Moreover, the algorithm does not require any prior knowledge about number of clusters or number of branches where they are determined dynamically in the model.

2.6 Neural Network architectures for Topic Modeling

There exist some studies which uses neural network architectures for topic modeling.

Replicated Softmax is a two-layer undirected graphical model which uses binary distributed representation of the documents rather than behaving documents as distribution over topics [7]. This model can be viewed as parameter sharing case of Restricted Boltzmann Machines which enables to model the different length of documents and makes learning easy and stable. The architecture of the model can be observed in Figure 3. In this figure, top layer corresponds to binary topic features and bottom layer shows the softmax visible units.

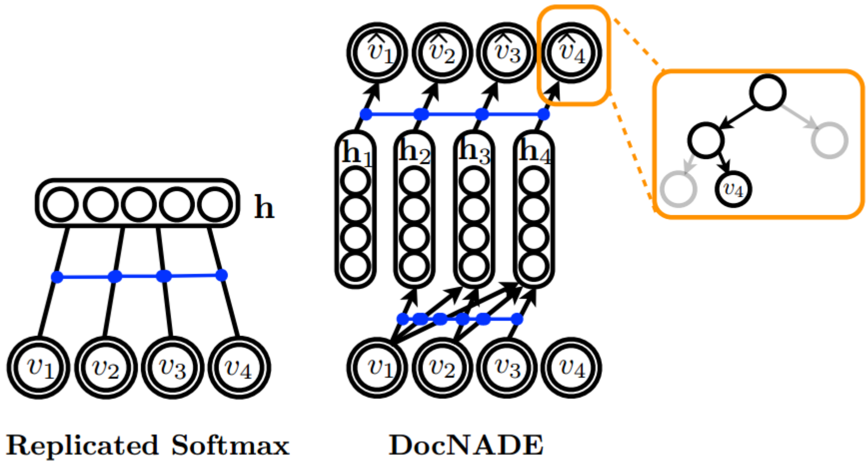

Figure 4: Replicated Softmax and DOcNADE models
(Figuresource:[12])

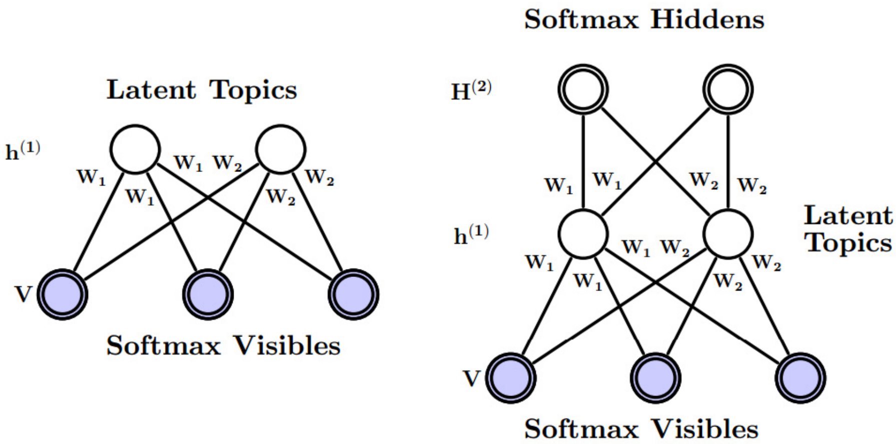

Figure 5: Deep Boltzman Machine for topic modeling
Left: Replicated Softmax Model **Right:** Over-Replicated Softmax Model
(Figuresource:[22])

Neural Autoregressive Density Estimators Neural (DocNADE) is an unsupervised neural network topic model of documents which is inspired by Replicated Softmax [12]. The model is generative and based on a binary tree that each leaf node can be seen as vocabulary. The structure difference between Replicated Softmax and DocNADE can be observed in Figure 4. In replicated Softmax Model and DocNADE models, each multinomial observation v_i correspond to a word and connections between v_i and hidden units are shared.

In another study, Deep Boltzman Machines (DBMs) with two hidden layers is used for modeling documents [22]. This model adds another hidden unit on top of the first hidden layer in Replicated Softmax structure to make it more flexible. This model is also called as Over-Replicated Softmax. Difference between these structures can be observed in Figure 5. The number of hidden softmax units should be pre-determined and fixed across all DBMs in this architecture.

3 Our Model: Incremental Hierarchical Topic Modeling with Incremental Deep Boltzmann Machines

Since we cannot anticipate all contexts that an agent is expected to encounter during its lifetime, we should develop mechanisms that allow the agent to discover a new context and place it appropriately in a hierarchy.

For this purpose, firstly we need a hierarchical structure. As already stated, models that are based on LDA behaves topics as distribution over words. Therefore, each topic contains a probability distribution of the words in a size of vocabulary. Since we need to measure the closeness of the topics, we need some metrics for distance calculation between topics. To calculate this distance, we can use Kullback-Leibler (KL) Divergence which is a measure of the non-symmetric difference between two probability distributions and proposed as an alternative way of measuring distances between topics, in the conclusion and further study section of the study of Hu et al. [8]. Therefore, we can calculate the KL distance between topics and merge them if this distance value is smaller than a threshold for each layer. This threshold values will be adjustable for each layer of incremental Deep Boltzmann Machine architecture and will specify the depth of our model.

Moreover, we need an incremental implementation where the number of topics can be increased automatically by the system when necessary upon each new encountered document/scene. Incremental LDA algorithm [3] seems very suitable for this purpose and we are planing to use the same approach to find the number of topics in an incremental way which will correspond to neurons of first hidden layer of our incremental Deep Boltzmann Machine architecture.

We are planing to extend DBMs in a way that suits for these requirements. Details of these extensions will be explained in the following section.

3.1 Incremental Deep Boltzmann Machines

To construct our incremental Deep Boltzmann machines, we will follow these steps:

- Start with $k = 1$.
- Receive/observe new scene/document.
- Assign the new topic to a suitable topic.
- Run incremental LDA [3], which introduces new context if necessary.
- Compare distances between contexts to check possibility for merging two contexts.
- Continue with step 2.

Our incremental DBM does not require any prior knowledge about depth of the hierarchy or nodes in each level. These parameters are calculated in terms of incremental LDA algorithm for the first hidden layer and distance between generated topics for deeper hidden layers. Therefore the model automatically estimates the depth and number of nodes.

Moreover, our DBM architecture is open for modification which means model can adjust its structure in terms of new coming documents.

The details of the construction process of incremental DBM can be observed in Figure 6.

4 Conclusion and Future Work

This report shows the main steps which we progressed so far:

- Examining the nature of the context especially for hierarchical properties
- Observing the existing architectures for modeling context
- Proposing a new model which is suitable for hierarchical and incremental structure of our problem

We need to construct the proposed architecture in Figure 6 and then apply the algorithm in Figure 7 to model context hierarchy as a future work of this project.

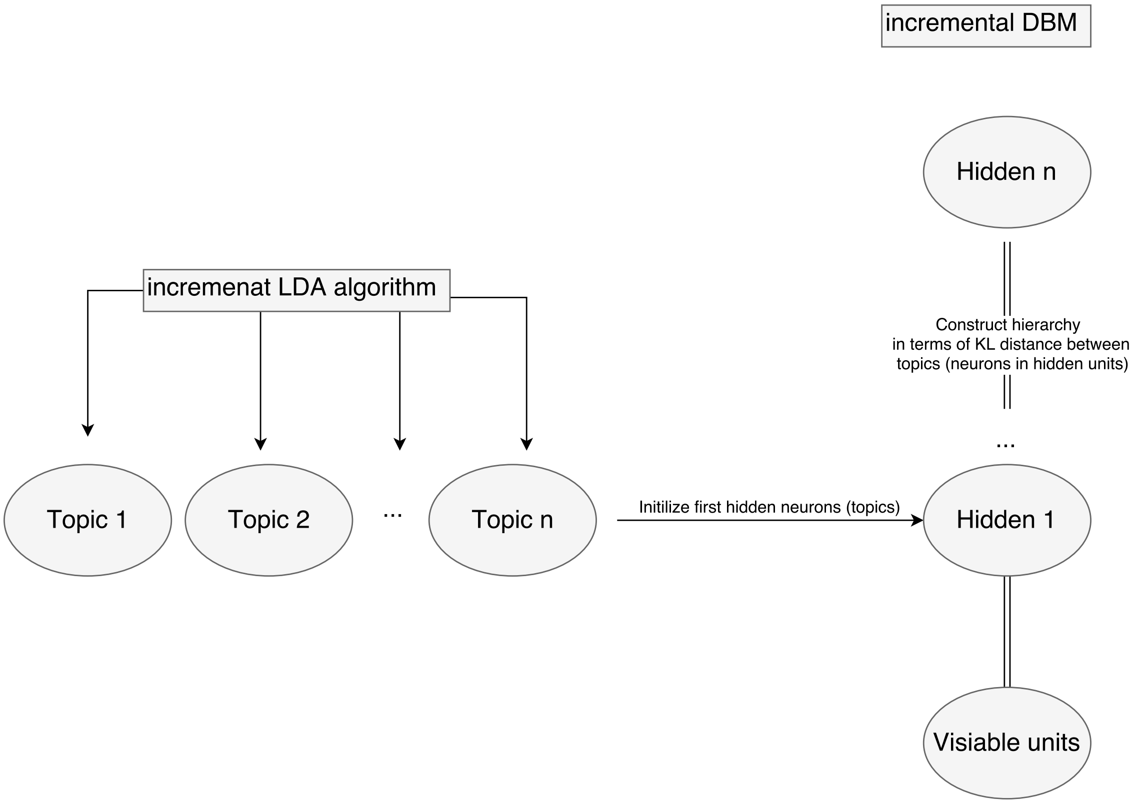

Figure 6: Construction of Incremental Deep Boltzmann Machines

Acknowledgments

For the experiments, we acknowledge the use of the facilities provided by the the Modeling and Simulation Center of METU (MODSIMMER). This work is funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) through project no 215E133.

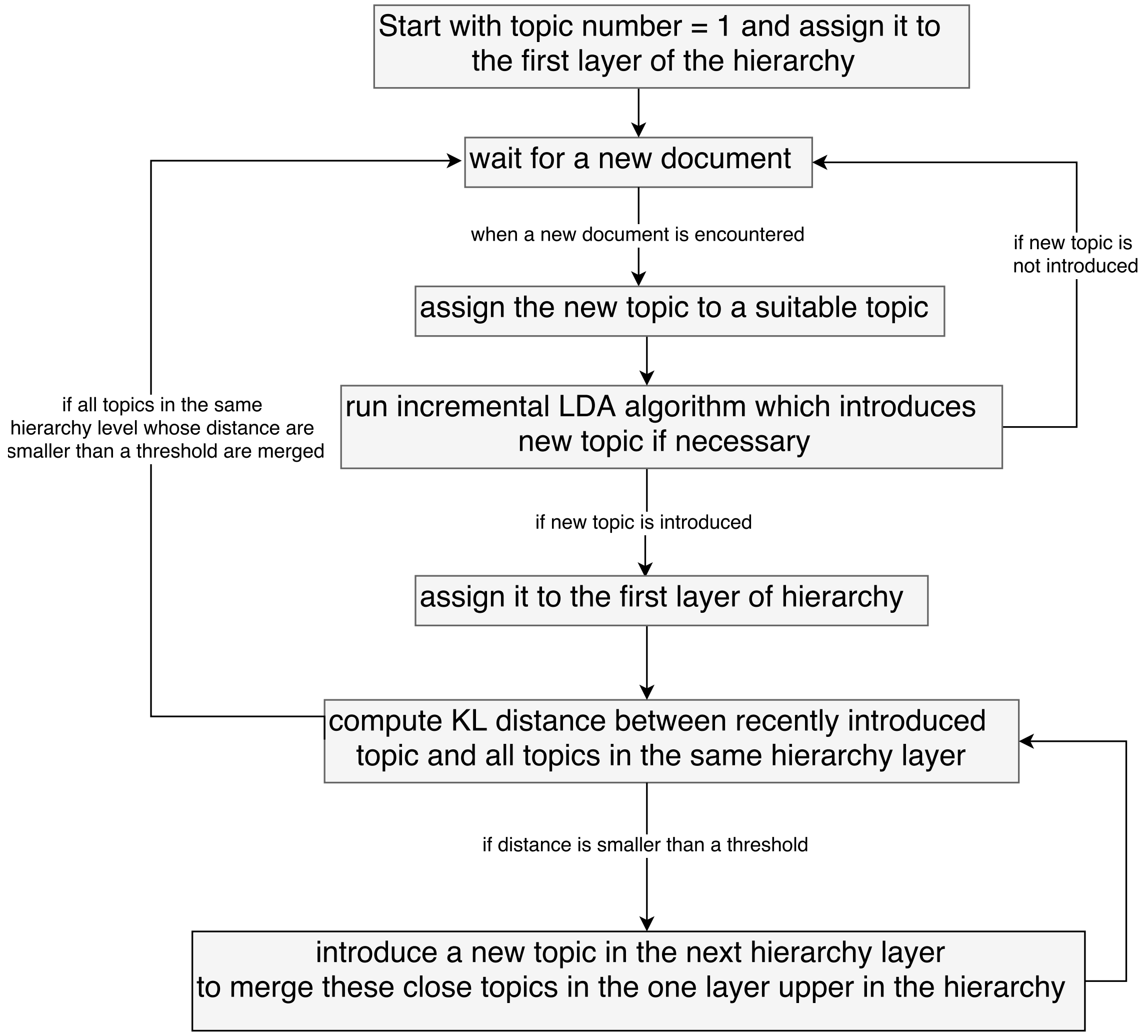

Figure 7: Algorithm for hierarchical topic modeling using Incremental Deep Boltzmann Machines

References

[1] Anand, A., Koppula, H. S., Joachims, T., and Saxena, A. (2012). Contextually guided semantic labeling and search for 3d point clouds. *The International Journal of Robotics Research*.

[2] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

[3] Celikkanat, H., Orhan, G., Pugeault, N., Guerin, F., Şahin, E., and Kalkan, S. (2016). Learning context on a humanoid robot using incremental latent dirichlet allocation. *IEEE Transactions on Cognitive and Developmental Systems*, 8(1):42–59.

[4] Choi, M. J., Lim, J. J., Torralba, A., and Willsky, A. S. (2010). Exploiting hierarchical context on a large database of object categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 129–136. IEEE.

[5] Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6):693–700.

[6] Griffiths, D. and Tenenbaum, M. (2004). Hierarchical topic models and the nested chinese restaurant process. *Advances in Neural Information Processing Systems*, 16:17.

[7] Hinton, G. E. and Salakhutdinov, R. R. (2009). Replicated softmax: an undirected topic model. In *Advances in Neural Information Processing Systems*, pages 1607–1614.

[8] Hu, L., Shao, C., Li, J., and Ji, H. (2015). Incremental learning from news events. *Knowledge-Based Systems*, 89:618–626.

[9] Jiang, Y., Koppula, H., and Saxena, A. (2015). Modeling 3d environments through hidden human context. *Tech Report*.

[10] Klarman, S. and Gutiérrez-Basulto, V. (2013). Description logics of context. *Journal of Logic and Computation*.

[11] Lane, D. (2006). Hierarchy, complexity, society. In *Hierarchy in Natural and Social Sciences*, pages 81–119. Springer.

[12] Larochelle, H. and Lauly, S. (2012). A neural autoregressive topic model. In *Advances in Neural Information Processing Systems*, pages 2708–2716.

[13] Li, Y., Huang, C., Loy, C. C., and Tang, X. (2016). Human attribute recognition by deep hierarchical contexts. In *European Conference on Computer Vision*, pages 684–700. Springer.

[14] McCarthy, J. (1993). Notes on formalizing context. *International Joint Conference on Artificial Intelligence*, pages 555–560.

[15] McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., and Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1):202–215.

[16] Paisley, J., Wang, C., Blei, D. M., and Jordan, M. I. (2015). Nested hierarchical dirichlet processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):256–270.

[17] Rudy, J. W. (2009). Context representations, context functions, and the parahippocampal–hippocampal system. *Learning & Memory*, 16(10):573–585.

[18] Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 48(1):9–26.

[19] Seiter, J., Chiu, W.-C., Fritz, M., Amft, O., and Tröster, G. (2015). Joint segmentation and activity discovery using semantic and temporal priors. In *IEEE Conference on Pervasive Computing and Communications*, pages 71–78. IEEE.

[20] Smith, A., Hawes, T., and Myers, M. (2014). Hiérarchie: Interactive visualization for hierarchical topic models. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 71–78.

[21] Smith, D. M. and Mizumori, S. J. (2006). Hippocampal place cells, context, and episodic memory. *Hippocampus*, 16(9):716–729.

[22] Srivastava, N., Salakhutdinov, R. R., and Hinton, G. E. (2013). Modeling documents with deep boltzmann machines. *arXiv*.

[23] Sun, J., Wu, X., Yan, S., Cheong, L.-F., Chua, T.-S., and Li, J. (2009). Hierarchical spatio-temporal context modeling for action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2004–2011. IEEE.

[24] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2004). Sharing clusters among related groups: Hierarchical dirichlet processes. In *NIPS*, pages 1385–1392.

[25] Turner, J. C., Oakes, P. J., Haslam, S. A., and McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, 20:454–454.

[26] Wang, C., Liu, X., Song, Y., and Han, J. (2015). Towards interactive construction of topical hierarchy: A recursive tensor decomposition approach. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1225–1234. ACM.

[27] Wang, P., Zhang, P., Zhou, C., Li, Z., and Yang, H. (2017). Hierarchical evolving dirichlet processes for modeling nonlinear evolutionary traces in temporal data. *Data Mining and Knowledge Discovery*, 31(1):32–64.

[28] Wang, X. and Ji, Q. (2014). A hierarchical context model for event recognition in surveillance video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2561–2568.

[29] Wang, X. and Ji, Q. (2015). Video event recognition with deep hierarchical context model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4418–4427.

[30] Yeh, W. and Barsalou, L. W. (2006). The situated nature of concepts. *The American Journal of Psychology*, pages 349–384.

[31] Zavitsanos, E., Paliouras, G., and Vouros, G. A. (2011). Non-parametric estimation of topic hierarchies from texts with hierarchical dirichlet processes. *Journal of Machine Learning Research*, 12(Oct):2749–2775.

[32] Zimmermann, A., Lorenz, A., and Oppermann, R. (2007). An operational definition of context. In *International and Interdisciplinary Conference on Modeling and Using Context*, pages 558–571. Springer.