Concept to Code: Deep Learning for Fashion Recommendation

Omprakash Sonie
Flipkart
omprakash.s@flipkart.com

Muthusamy Chelliah
Flipkart
muthusamy.c@flipkart.com

Shamik Sural
IIT Kharagpur
shamik@cse.iitkgp.ac.in

ABSTRACT

Deep Learning has shown significant results in various domains. In this tutorial, we provide conceptual understanding of embedding methods, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNNs). We present fashion use case and apply these techniques for modeling image, text as well as sequence data to figure out user profiles, give personalized recommendations tailored to changing user taste and interest. Given the image of a fashion item, recommending complementary matches is a challenge. Users' taste evolves over time and depends on persona. Humans relate objects based on their appearance and non-visual factors of lifestyle merchandise which further complicate recommendation task. Composing outfits in addition necessitates constituent items to be compatible - similar in some but different in other aspects.

CCS CONCEPTS

• **Applied computing → Online shopping**.

ACM Reference Format:
Omprakash Sonie, Muthusamy Chelliah, and Shamik Sural. 2019. Concept to Code: Deep Learning for Fashion Recommendation. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308560.3320100

1 INTRODUCTION AND EMBEDDING

Embedding is one of the key components in deep learning based systems. Hence, we go deeper in explaining Embeddings [15] [10] for learning vector representation of products, users and other types of data. Will deep dive into Skip-gram technique (with necessary foundation) which will be used for hands-on. We summarize the key concepts of the tutorial and the coverage in the following sections.

2 BASIC DEEP NEURAL NETWORK ARCHITECTURES

Convolutional Neural Network (CNN): A CNN is a multi-layer network having convolutional layers. Each layer consists of a number of filters. A filter is applied to the various local sub-regions of the image giving rise to a feature map. A CNN requires fewer parameters as the features are directly extracted from content such as product image or product review text. In a deep multi-layer CNN, the different layers learn different image features [14]. A common technique is to extract features from an image of a product using a pre-trained CNN.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3320100

Recurrent Neural Network (RNN): RNNs model sequential data, capture memory of the past and can be trained by unfolding and doing back propagation through time. However, simple RNNs often fail to capture long distance temporal dependencies. Long Short Term Memory network (LSTM) and Gated Recurrent Unit (GRU) have been proposed to address this issue and have performed very well [6]. They enable recommender systems to model content sequences or temporal dynamics. RNNs have been used for session-based recommendations. [20] uses LSTMs to capture temporal dependencies of users and items.

3 STYLE DETECTION

In this section we describe what style is, its importance and how it is learnt. Style prediction involves analyzing the degree to which a style is exhibited either for items within same category or across different categories and finding which style is general and which one indicates a person's specific style. o address representation and learn inter-dependencies between features of fashion item we need to harness large fashion image corpus and rich meta [9]. There is a need to model simultaneously visual appearances of product as well as their evolution over time to uncover the complex and evolving visual factors that people consider when evaluating products [4]. An item consists of both Style and Category. Style information indicates preferences of users and has a significant effect in visual recommendation. Conventional methods model only categorical information. The model learns style features for understanding user preference [11].

4 RECOMMENDATION

In this section we cover various recommendation techniques. People make choices based on which items are alternative to each other (such as two T-shirts) and which items are complementary (T-Shirts and matching jeans). This uncovering human notions of visual relationships of items is modeled as network inference problem on graphs of related images [14]. An effective recommendation is not just recommending one item but a set of items that fit user's personal preferences. Instead of suggesting individual items, the model suggests a set of personalized items [7]. Aesthetic features (e.g., clean design, elegant style etc.) play important role in user's decision making process and varies from user to user and by time. The models extracts aesthetic features and includes in recommender system [21]. Users are more likely to buy outfits recommended based on high Attractive Quotient. The model suggests outfit based on Attractive Quotient and removes the outfits which are judged unattractive and suggests partial replacement [1].

It may be noted that, unlike other kinds of recommender systems, in the context of outfit recommendation, there are a very large number of possible outfits that are available on any online shopping site although for each such item, the number of actual customers could only be a small fraction of the total number of existing or potentially new customers. Also, customers are often less guided by the choice of other customers than by their own personal preferences. Further, unlike movies and books, different users can have different budgetary constraints and would prefer that the recommender system meets such constraints specified either explicitly or implicitly.

5 OUTFIT COMPATIBILITY

In this section we cover various outfit compatibility approaches. In the fashion domain, discovering items that are functionally complementary or visually compatible is important. The model is designed to handle a set of complex relationships among items, as well as high-dimensional and semantically complicated features [5]. Users need to match clothes to make a suitable outfit. The model integrates deep neural networks and rich fashion domain knowledge for recommending outfit [17]. For suggesting an item that matches items in a stylish outfit (a collection of fashion items) the model learns compatibility by considering a fashion outfit fit to be a sequence and each items as a step. The visual embedding of items is included in the model [3]. Outfit recommender system need to discriminate substitutable products (that are interchangeable) and complementary products (which might be purchased together by users). The model leverages not just direct paths but complex dependencies - represented as sequence of relations and items in a product graph [19].

6 MISCELLANEOUS WORK

In this section we cover other approaches for fashion recommendation. In search, learning to rank of items is traditionally done over few hand-constructed features from item's text. The model combines traditional features with visual semantic features from images for ranking [12]. To help designer, brands and consumers to comprehend importance of different attributes in an outfit composition we need to include explainability in the deep learning based models. The model recommends a ranked list of outfit composition with interpretable matching scores [2]. Users are unhappy with wrong product size in popular categories (e.g., apparel, shoes) which results in high return rate for supplier. The model determines how a product of certain size fits a customer based on physical true sizes, sizes that are learnt from past product purchase and returns data [16]. Fashion domain is primarily visual. Traditional systems recommend similar looking 'top' for a 'bottom' query. To provide diversified items the model learns and recommends items based on generated complementary items [8].

7 CODE WALKTHROUGH AND CONCLUSION

We walk-through the code for a fashion recommender system on e-commerce dataset, summarize these models, parameters and understand what is going on behind the scene with various visualizations [13] [18]. We will use Jupyter notebook with already executed code for walkthrough. Fashion recommendation is an intricate problem that needs out of the box algorithm design. This is more due to the fact that there are various factors that need to be estimated from user interaction. For example, whether the user wants complementary or substitutable items, items similar to or completely different from previous purchases, intends to go with the current trend or deviate from the same, etc. There are also difficulties arising out of the need for making the recommendation personalized. While some of these necessitate quite a bit of interaction with the user, at the same time, too much of interaction might put off the potential customer. Profitability and competitiveness considerations also add to the overall complexity of the system design.

REFERENCES

[1] D. Banerjee, N. Ganguly, S. Sural, and K. S. Rao. 2018. One for the Road: Recommending Male Street Attire. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. 571–582.

[2] Z. Feng, Z. Yu, Y. Yang, Y. Jing, J. Jiang, and M. Song. 2018. Interpretable Partitioned Embedding for Customized Multi-item Fashion Outfit Composition. In *International Conference on Multimedia Retrieval (ICMR)*. 143–151.

[3] X. Han, Z. Wu, Y. Jiang, and L. S. Davis. 2017. Learning Fashion Compatibility with Bidirectional LSTMs. In *ACM International Conference on Multimedia (MM)*. 1078–1086.

[4] R. He and J. McAuley. 2016. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. In *International Conference on World Wide Web (WWW)*. 507–517.

[5] R. He, C. Packer, and J. McAuley. 2016. Learning Compatibility Across Categories for Heterogeneous Item Recommendation. In *International Conference on Data Mining (ICDM)*. 937–942.

[6] S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[7] Y. Hu, X. Yi, and L. S. Davis. 2015. Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach. In *ACM International Conference on Multimedia (MM)*. 129–138.

[8] C. P. Huynh, A. Ciptadi, A. Tyagi, and A. Agrawal. 2018. CRAFT: Complementary Recommendations Using Adversarial Feature Transformer. *CoRR* abs/1804.10871 (2018).

[9] V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di, and N. Sundaresan. 2014. Large Scale Visual Recommendations from Street Fashion Images. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. 1925–1934.

[10] Q. Le and T. Mikolov. 2014. Distributed Representations of Sentences and Documents. In *International Conference on International Conference on Machine Learning (ICML) (ICML'14)*, Vol. 32. II–1188–II–1196.

[11] Q. Liu, S. Wu, and L. Wang. 2017. DeepStyle: Learning User Preferences for Visual Recommendation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 841–844.

[12] C. Lynch, K. Aryafar, and J. Attenberg. 2016. Images Don'T Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. 541–548.

[13] L. van der Maaten and G. Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research (JMLR)* 9 (2008), 2579–2605.

[14] J. McAuley, C. Targett, Q. Shi, and A. van den H. 2015. Image-Based Recommendations on Styles and Substitutes. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 43–52.

[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems (NIPS)*. 3111–3119.

[16] V. Sembium, R. Rastogi, L. Tekumalla, and A. Saroop. 2018. Bayesian Models for Product Size Recommendations. In *International Conference on World Wide Web (WWW)*. 679–687.

[17] X. Song, F. Feng, X. Han, X. Yang, W. Liu, and L. Nie. 2018. Neural Compatibility Modeling with Attentive Knowledge Distillation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 5–14.

[18] O. Sonie, S. Sarkar, and S. Kumar. 2018. Concept to code: learning distributed representation of heterogeneous sources for recommendation. In *The ACM Conference Series on Recommender Systems (RecSys)*. 531–532.

[19] Z. Wang, Z. Jiang, Z. Ren, J. Tang, and D. Yin. [n. d.]. A Path-constrained Framework for Discriminating Substitutable and Complementary Products in E-commerce. In *ACM International Conference on Web Search and Data Mining (WSDM)*. 619–627.

[20] C. Wu, A. Ahmed, A. Beutel, A. J. Smola, and H. Jing. 2017. Recurrent Recommender Networks. In *ACM International Conference on Web Search and Data Mining (WSDM)*. 495–503.

[21] W. Yu, H. Zhang, X. He, X. Chen, L. Xiong, and Z. Qin. 2018. Aesthetic-based Clothing Recommendation. In *International Conference on World Wide Web (WWW)*. 649–658.