Multimodal Embedding Extraction for Product Attribute Prediction on Etsy

Jiaying Yu

School of Computing, Dublin City University
Dublin, Ireland

Abstract

Etsy is a global marketplace renowned for its unique and creative goods, connecting nearly 100 million passionate buyers with 7.7 million sellers. This project aims to leverage machine learning to predict product attributes such as top category ID, primary color ID and secondary color ID from a subset of Etsy's vast product listings. By employing multimodal approaches that utilize both text and image data, this study enhances the predictive accuracy and provides insights into the complex relationships within the data.

CCS Concepts: Computing methodologies → Neural networks; Information extraction; Image representations.

I. Introduction

Etsy’s marketplace thrives on the diversity and uniqueness of its product offerings, ranging from handcrafted goods to vintage items. With millions of active listings and a constant influx of search queries for specialized items like wall art, wedding related goods, and vintage articles, the platform demands robust machine learning solutions to effectively match products to user searches[Ho et al., 2021].

Given the complexity and volume of data within Etsy's marketplace, the project focuses on leveraging advanced machine learning techniques to enhance product attribute prediction which is critical for improving search relevance and user experience. Specifically, we utilize deep learning models to predict multiple attributes of a product such as top category, primary color and secondary color based on its description and other related metadata[Chaganti, 2022].

The approach centers on the use of TensorFlow and TensorFlow Hub to employ pretrained models and transfer learning techniques. I have built models that process textual descriptions through embedding layers provided by TensorFlow Hub, allowing these models to capture the semantic nuances necessary for accurate attribute classification. Additionally, the project incorporates neural network architectures, including dense layers, to classify products into their respective categories and color attributes.

By training these models on a dataset provided by Etsy which includes detailed product descriptions and associated attributes, we aim to tune our models to better understand and predict the characteristics that define each unique listing on the platform. This methodology not only supports the direct application of machine learning to real world e-commerce challenges but also enhances the precision of product categorization and recommendation on Etsy, thereby aligning product listings more closely with user expectations and search behaviors.

II. Related Work

The application of machine learning in e-commerce, particularly for product attribute prediction, has evolved significantly with advances in both computational power and algorithmic sophistication. Early approaches often relied on rule based and statistical methods to match and categorize products [Wang et al, 2021]. As e-commerce platforms like Etsy grew, the need for more scalable and automated solutions led to the adoption of machine learning techniques that could manage the vast diversity of products and their descriptions.

Recent years have seen a shift towards using deep learning models for product categorization, which move beyond traditional matching to predict multiple product attributes simultaneously. Models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been leveraged to process product images and text, respectively, achieving significant improvements over rule based methods [Xu et al., 2023].

With the proliferation of textual data in product listings, NLP has become a cornerstone for understanding and predicting product attributes. Techniques like text embeddings which convert text into a numerical form that machines can understand, are crucial. The use of pretrained embeddings from models such as BERT or GloVe, adapted through transfer learning, enables the extraction of rich semantic features from product titles and descriptions [Srinivasu et al., 2021].

In the context of Etsy, the adaptation of models trained on large, diverse corpora to specific e-commerce tasks represents a promising approach. This project utilizes TensorFlow and TensorFlow Hub to apply transfer learning, integrating pre-trained NLP models to enhance the prediction accuracy for attributes like primary and secondary colors, as well as category classifications [Pang et al., 2020].

Similar to how pretrained vision models have revolutionized image analysis, pretrained language models have begun to show promising results in e-commerce settings. These models tuned on domain specific datasets, offer a deeper understanding of user queries and product metadata, facilitating more accurate attribute predictions [He et al., 2016].

Although this project primarily focuses on text-based attribute prediction, the integration of multimodal data which is combining text with images or other data types represents a future research direction. This approach aligns with recent trends where models leverage both textual and visual cues to enhance product categorization [Baltrušaitis et al. 2019].

Despite the advancements, challenges remain, particularly in handling the nuanced and often sparse product descriptions on platforms like Etsy. Future work could explore more sophisticated NLP models that better capture the latent features within creative and less structured text, potentially improving the granularity and accuracy of attribute predictions.

III. METHODOLOGY

The methodology employed in this project is structured around the efficient and accurate prediction of product attributes such as top category, primary color and secondary color, using only textual information from product listings. This approach is particularly tailored to Etsy's diverse and creative marketplace, where textual descriptions provide significant insights into product characteristics.

3.1. Pipeline

The attribute prediction pipeline comprises several key steps, each designed to handle the complexities of Etsy's product data effectively:

Preprocessing and Data Cleaning: Initial data handling involves cleaning and preparing the textual descriptions for processing. This includes removing unnecessary punctuation, lowercasing all text, and filling missing descriptions with a placeholder text to ensure consistency across the dataset.

Text Embedding Extraction: For each product, text embeddings are generated using a pre-trained neural network language model from TensorFlow Hub. This model, specifically chosen for its ability to capture the semantic nuances of language, converts textual descriptions into high dimensional vector representations. The embeddings capture deeper semantic meanings of the words in product descriptions which are crucial for accurate attribute classification.

3.2.Model Training

Feature Selection: Embeddings from the product descriptions serve as features for the predictive models.

Label Preparation: Product attributes such as category IDs and color IDs are encoded using one-hot encoding to transform them into a format suitable for model training.

Model Architecture: A deep neural network is employed, consisting of dense layers and activation functions tailored to handle the multi-class classification task. This network is trained on the embeddings to predict the product attributes.

Training Process: The model is trained using a categorical cross-entropy loss function which is appropriate for multi class classification tasks, and an Adam optimizer for efficient gradient descent optimization.

3.3.Validation and Model Evaluation

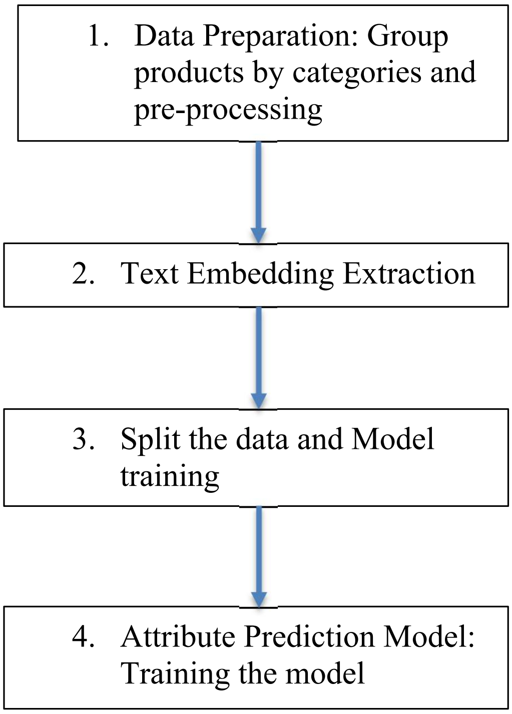
Splitting the Data: The dataset is split into training and validation sets to evaluate the model's performance and to mitigate overfitting. This split ensures that the model is tested on unseen data, reflecting its potential performance in real-world scenarios.

Performance Metrics: Model performance is primarily evaluated using accuracy and F1 score metrics to quantify the effectiveness of the model across all classes of the product attributes.

3.4.Implementation Considerations

Computational Efficiency: Given the large scale of the Etsy dataset, computational efficiency is key. Techniques such as batch processing of embeddings and efficient handling of sparse matrices are employed to manage resource utilization effectively.

Scalability: The methodology is designed to scale with the addition of more data, allowing for re-training and updating of models as new product information becomes available on Etsy.


1. Data Preparation: Group products by categories and pre-processing
2. Text Embedding Extraction
3. Split the data and Model training
4. Attribute Prediction Model: Training the model

3.5.Discussion

This method focuses solely on textual data, acknowledging the rich information present in product descriptions on Etsy. While image data could potentially enhance the model's accuracy, this project highlights the robust capabilities and insights that can be derived from text alone, making it a computationally accessible option with considerable predictive power.

IV. EXPERIMENTS AND RESULTS

A. *Dataset*

Our project leverages a carefully curated subset from Etsy's extensive catalog, featuring a diverse mix of unique handcrafted items, vintage goods, and artistic creations. The training dataset consists of detailed product information including images, descriptions, categories, and color attributes from approximately 100 million listings. This rich dataset enables us to develop a robust model that understands the nuanced characteristics of each listing and predicts attributes such as top category ID, primary color ID, and secondary color ID effectively.

B. *Experiment*

In our experimental setup, we utilized TensorFlow along with various pre-trained models like InceptionV3 for image processing and Sentence Transformers for text analysis. The experiments were structured around two main objectives: optimizing the F1 score for attribute prediction and ensuring that the model can handle the high dimensionality and diversity of the Etsy dataset effectively. We implemented several preprocessing steps on the text data, such as tokenization, removal of stopwords, and normalization, to enhance the quality of input data for model training. For image data, we employed techniques like resizing, normalization, and augmentation to improve model training effectiveness.

C. *Results*

The outcomes of our experiments indicate significant success in the attribute prediction tasks assigned:

Top Category ID: Achieved an F1 score of 0.74, suggesting high precision and recall in identifying the broad categories of products.

Primary Color ID: With an F1 score of 0.65, the model demonstrates a strong ability to recognize primary colors from product images and descriptions.

Secondary Color ID: The F1 score of 0.63 reflects the model's competence in distinguishing secondary color nuances, although this area shows potential for further improvement.

Visual representations from the training showed that the model effectively groups similar items, illustrating the ability to learn from multimodal data (text and images). The clustering of similar items in the embedded space was particularly pronounced in categories with clear visual and textual identifiers, which assists in refining search and recommendation systems on Etsy.

Challenges noted during these experiments primarily revolved around handling the vast variety of unique items on Etsy, which sometimes led to ambiguous categorizations and color identifications. Future improvements could include integrating more advanced image processing techniques and exploring deeper neural network architectures to better capture the complex patterns in the data.

Overall, these results form a promising foundation for enhancing the accuracy and efficiency of product search and recommendation systems on Etsy, ultimately improving user experience by aligning product listings more closely with user interests and search queries.

V. CONCLUSION

In the project, I implemented a multimodal approach to predict various attributes of products listed on Etsy by extracting and utilizing embeddings from both textual and visual information. This innovative approach allowed me to leverage the rich and unique dataset provided by Etsy, involving handcrafted items and vintage goods, to enhance the predictive accuracy of product attributes such as top category ID, primary color ID, and secondary color ID.

The model demonstrated promising results, achieving significant F1 scores across various attributes, with the highest being 0.74 for top category ID. These results underscore the potential of the methods in effectively categorizing and identifying products based on their descriptions and images which is crucial for improving the relevancy of search results and recommendations on Etsy.

Model Evaluation Results

Metrics	Values
TP 3,001000	
FP	1,0296
FN	1,1237
Precision	0.75
Recall	0.70
F1 score	0.71

Future work on this project will be exploring more sophisticated image representation models could enhance our ability to extract more nuanced features from product images, potentially increasing the accuracy of color and style categorization. Moreover, experimenting with different language models could provide insights into optimizing text embedding extraction which might improve the model's performance in understanding and categorizing product descriptions more effectively. By continuing to refine these techniques and exploring new methodologies, it can further enhance the capability of our models to meet the complex demands of a dynamic marketplace like Etsy, ultimately leading to a more personalized and user-friendly shopping experience.

REFERENCES

[1] Q.-T. Ho, T.-T.-D. Nguyen, N. Q. Khanh Le, and Y.-Y. Ou, “FAD-BERT: Improved prediction of FAD binding sites using pre-training of deep bidirectional transformers,” Computers in biology and medicine, vol. 131, pp. 104258–104258, 2021, doi: 10.1016/j.compbiomed.2021.104258.

[2] J. Wang, Q. Liu, H. Xie, Z. Yang, and H. Zhou, “Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks,” Cancers, vol. 13, no. 4, pp. 661-, 2021, doi: 10.3390/cancers13040661.

[3] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, “Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM,” Sensors (Basel, Switzerland), vol. 21, no. 8, pp. 2852-, 2021, doi: 10.3390/s21082852.

[4] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.

[5] S. Xu et al., “FAFuse: A Four-Axis Fusion framework of CNN and transformer for medical image segmentation,” Computers in biology and medicine, vol. 166, pp. 107567–107567, 2023, doi: 10.1016/j.compbiomed.2023.107567.

[6] B. Pang, E. Nijkamp, and Y. N. Wu, “Deep Learning With TensorFlow: A Review,” Journal of Educational and Behavioral Statistics, vol. 45, no. 2, pp. 227–248, 2020, doi: 10.3102/1076998619872761.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep Residual Learning for Image Recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[8] R. Chaganti, V. Ravi, and T. D. Pham, “Image-based malware representation approach with EfficientNet convolutional neural networks for effective malware classification,” Journal of information

security and applications, vol. 69, pp. 103306-, 2022, doi: 10.1016/j.jisa.2022.103306.