Predykcja Rekomendacji Kosmetyków przy Użyciu Multimodalnych Sieci Neuronowych

Autorzy: Natalia Łączkowska

Abstrakt

Artykuł przedstawia porównanie trzech architektur sieci neuronowych do predykcji rekomendacji produktów kosmetycznych: modelu MLP (cechy tabelaryczne), sieci BiLSTM z mechanizmem uwagi (tekst recenzji) oraz modelu hybrydowego. Dataset zawierał 1M recenzji z platformy Sephora. Model MLP osiągnął najlepsze wyniki (96.24%% accuracy, AUC=0.983) przy najmniejszej złożoności, przewyższając model multimodalny o 1.93pp.

1. Wstęp

W branży kosmetycznej generowane są miliony recenzji produktów rocznie. Stanowią one źródło informacji dla konsumentów i producentów. Automatyczna klasyfikacja sentymentu pozwala na szybką identyfikację produktów polecanych przez użytkowników. Głównym wyzwaniem jest skuteczne przetworzenie wielomodalnych danych - tekstu recenzji oraz cech produktów (cena, kategoria, rating).
Clem pracy jest porównanie trzech architektur sieci neuronowych rozwiązujących problem klasyfikacji binarnej: czy użytkownik poleca produkt (klasa 1) czy nie (klasa 0). Problem charakteryzuje się znacznym niezrównoważeniem klas (84% pozytywnych recenzji).

2. Przegląd Stanu Nauki

Najnowsze modele transformerowe (BERT, RoBERTa) dominują w analizie sentymentu. Osiągają dokładność powyżej 95-97% na zbiorach benchmarkowych, dzięki self-attention. W scenariuszach z ograniczonymi zasobami obliczeniowymi sprawdzają się rekurencyjne sieci BiLSTM z mechanizmem uwagi. Oferuje to accuracy na poziomie 92-94% przy mniejszej liczbie parametrów w porównaniu do pełnych transformerów [1][2][3].

3. Metody

3.1. Dataset

- Źródło: Kaggle - "Sephora Products and Skincare Reviews" (nadyinky/sephora-products-and-skincare-reviews), zawierający ~1M recenzji produktów kosmetycznych.
- Próbki zaierały: tekst, tytuł, rating (1-5), is_recommended (0/1), helpfulness, metadata.
- Preprocessing: Czyszczenie tekstu, tokenizacja, feature engineering, StandardScaler, SMOTE balancing.
- Podział: 80/20 train/test stratified, validation 20% train.

3.2. Testowane architektury

Model 1 (MLP Baseline): Sieć feedforward przetwarzająca wyłącznie cechy tabelaryczne. Input(19) → Dense(128, L2) → BatchNorm → Dropout → Dense(64, L2) → BatchNorm → Dropout → Dense(32) → Dropout → Output. Adam (lr=0.001). 13,697 parametrów.

Model 2 (BiLSTM + Attention): Sieć rekurencyjna przetwarzająca tekst recenzji. Embedding(5000, 128) → SpatialDropout1D → BiLSTM(64) → BiLSTM(32) → Attention → GlobalPooling → Dense(128) → BatchNorm → Dropout → Dense(64) → BatchNorm → Dropout → Dense(32) → Dropout → Output. AdamW (lr=0.001). 820,097 parametrów.

Model 3 (Fusion): Text branch (BiLSTM+Attention) + Tabular branch (MLP) → Concat → Dense(128, L2) → BatchNorm → Dropout → Dense(64) → BatchNorm → Dropout → Dense(32) → Dropout → Output. 821,473 parametry.

Hiperparametry: batch_size=128 (64 dla Model 1), epochs=30, EarlyStopping(patience=8), ReduceLROnPlateau(patience=4, factor=0.5), class weights balanced.

4. Rezultaty

Do określenia wyników modeli na zbiorze testowym zastosowano metryki takie jak AUC-ROC, Accuracy, Precision, Recall, F1-score, Confusion Matrix.

Model	Parametry	AUC	Accuracy	Precision	Recall	F1
Model 1	13,697	0.9826	0.9624	0.9913	0.9637	0.9773
Model 2	820,097	0.9088	0.8712	0.8973	0.8970	0.9212
Model 3	821,473	0.9723	0.9431	0.9828	0.9489	0.9655

Confusion Matrix:
Model 1: TN=518, FP=24, FN=102, TP=2733 (Recall: 95.57% klasa 0, 96.41% klasa 1)
Model 2: TN=425, FP=117, FN=378, TP=2457 (Recall: 78.41% klasa 0, 86.67% klasa 1)
Model 3: TN=495, FP=47, FN=137, TP=2698 (Recall: 91.33% klasa 0, 95.17% klasa 1)

5. Dyskusja

Wyniki pokazują dominującą rolę cech tabelarycznych w predykcji rekomendacji. Model 1 przewyższył złożone modele multimodalne, ponieważ cechy strukturalne (rating, price) zawierają bezpośredni sygnał o jakości produktu - użytkownicy rekomendują głównie na ich podstawie, a tekst pełni rolę uzupełniającą.
Słaba wydajność Modelu 2 potwierdza, że semantyka recenzji kosmetycznych jest niejednoznaczna - użytkownicy mogą pisać pozytywne opisy, ale nie polecać produktu ze względu na cenę. Model 3 nie poprawił wyników vs Model 1, co wskazuje na curse of multimodality - gdy jedna modalność jest wystarczająca, dodanie słabszej wprowadza szum.
Error Analysis: Model 1 błędnie klasyfikował recenzje z niespójnością rating-recommendation (np. rating=4, is_recommended=0). Model 2 miał wysoką liczbę false positives - klasyfikował recenzje jako pozytywne na podstawie entuzjastycznego języka, mimo braku rekomendacji.
Ograniczenia: Dataset 20K samples (więcej mogłoby poprawić Model 2/3), vocabulary=5000 może być niewystarczające dla terminów kosmetycznych, brak testowania na innych domenach.

6. Konkluzje

Badanie wykazało, że w zadaniu predykcji rekomendacji kosmetyków proste modele tabelaryczne przewyższają złożone architektury multimodalne. Model MLP osiągnął AUC=0.98 i accuracy=96.2% przy 13,697 parametrach - 60x mniej niż Fusion.
Kluczowe wnioski: (1) Cechy strukturalne (rating, price) są najsilniejszymi predyktorami (+9pp accuracy vs text-only). (2) BiLSTM+Attention niewystarczający – tekst zawiera szum. (3) Fusion poprawił wyniki vs text-only (+7pp), ale nie dorównał MLP-only (-1.9pp). (4) SMOTE balancing skuteczny (84%→50%).
Implikacje praktyczne: Dla systemów e-commerce warto priorytetyzować cechy tabelaryczne zamiast kosztowne modele NLP. Proste MLP są wystarczające i łatwiejsze do wdrożenia.
Przyszłe prace: Pre-trained embeddings (BERT), analiza attention weights, klasyfikacja wieloklasowa (rating 1-5), deployment jako API.

Bibliografia

[1] Gibson Nkhata, Susan Gauch, Usman Anjum, Justin Zhan. *Fine-tuning BERT with Bidirectional LSTM for Fine-grained Movie Reviews Sentiment Analysis.* 2025. URL: https://arxiv.org/pdf/2502.20682

[2] Md Abrar Jahin, Md Sakib Hossain Shovon, M. F. Mridha, Md Rashedul Islam, Yutaka Watanobe. *A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets.* 2024. URL: https://www.nature.com/articles/s41598-024-76079-5

[3] Mahammed Kamruzzaman, Gene Louis Kim. *Efficient Sentiment Analysis: A Resource-Aware Evaluation of Feature Extraction Techniques, Ensembling, and Deep Learning Models.* 2023. URL: https://aclanthology.org/2023.socialnlp-1.2.pdf