Product Level Hierarchy Classification with Transformer-based Clustering

Hao-Che Hsu

Introduction

Price indexes provide a reference for the cost of living. But for each product, we need to first identify the level hierarchy. I adopt an unsupervised (cluster by EM algorithm) and a supervised (compute cosine similarity) method with Bidirectional Encoder Representations from Transformers (BERT) to assign products to COICOP 6 (Classification of Individual Consumption by Purpose Level 6).

Data

This data is obtained from the Austrian National Bank (Oesterreichische Nationalbank, OeNB).

- Products from online stores
- 418,537 product observations (mostly are clothing- and housing-related)
- 84,860 unique products
- Observed product names, descriptions, and URLs
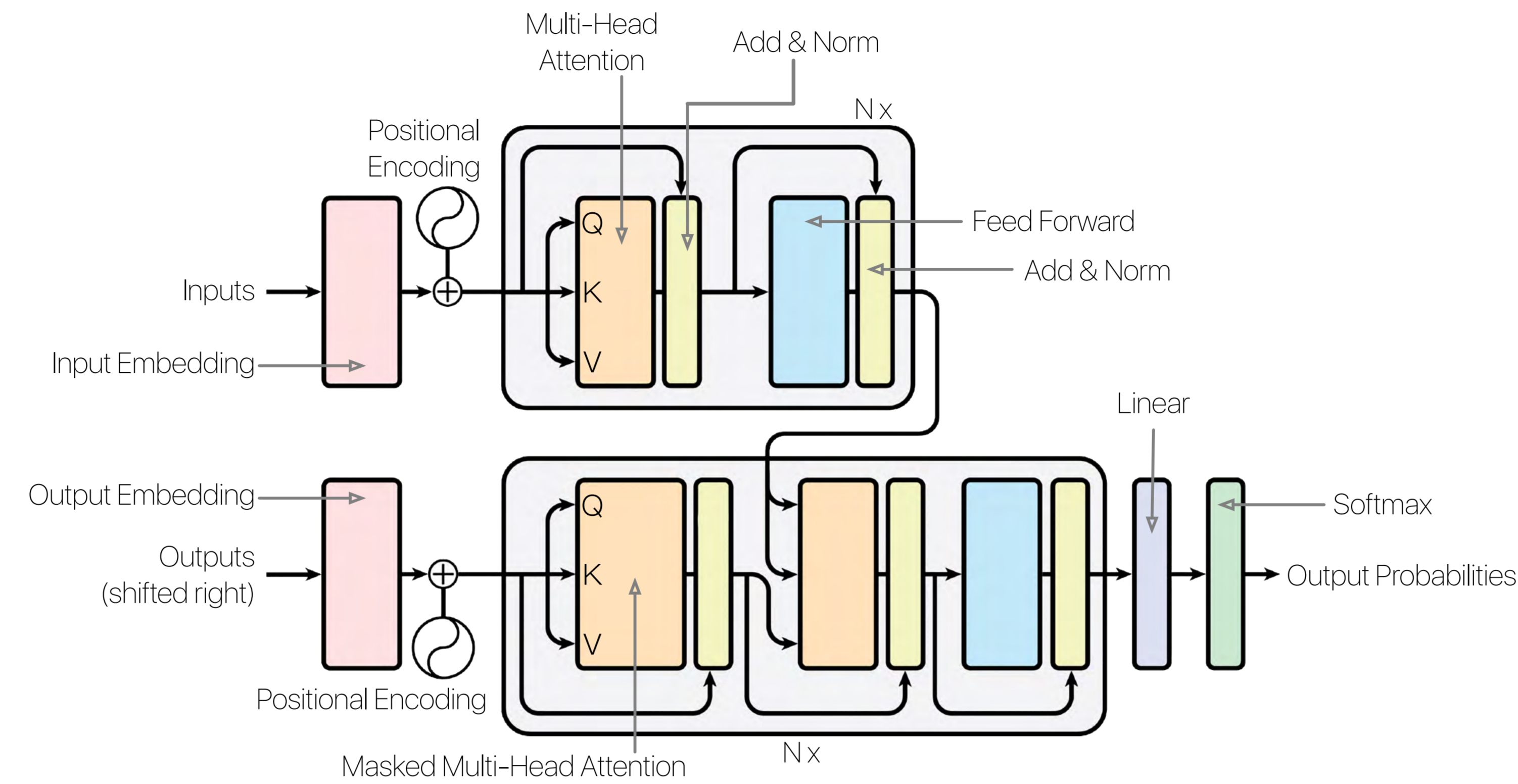- Product names in German or English

The Flow

Product Name → Batch Translation → Create Embeddings → Clustering & Classification → Price Analysis

NLP Models

- Bag of Words: lot of zeros, order does not matter
- N-grams: huge dimensionality
- RNN (plain): potential vanishing/exploding gradients
- LSTM: slow to train, gradient path grows in the number of words, need to be trained specifically for different dataset
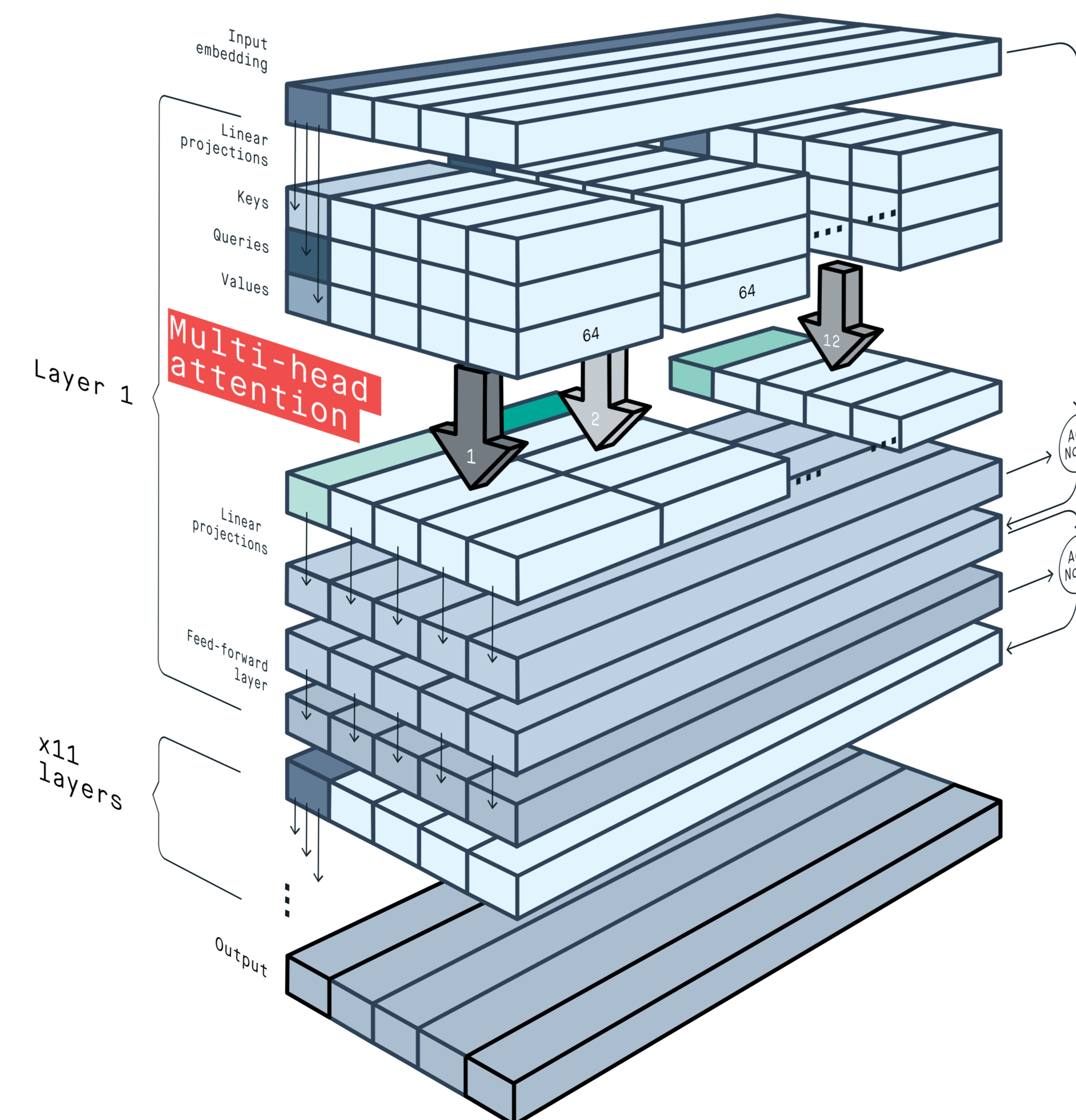
BERT Variations (pre-trained, fine tune later)

Models	$BERT_{base}$	$BERT_{large}$	DistilBERT
Number of parameters	110M	336M	66M
GPU usage	8	280	8
Embedding dimension	768	1024	768
# Attention heads	12	16	12

Transformer (Vaswani et al., 2017)

BERT (Devlin et al., 2018)

Product Name Clustering Results

Supervised Classification Results

- STRIPED JOGGER WAIST TROUSERS woman (Level 5: C03122)

$\implies$ (match with a Level 6) Womens tights, Womens jeans, Womens ski suit, Womens blazer, Womens jacket, Womens underpants, Womens sportswear, Womens skirt, Womens coat, **Womens pants**, Ladies shirt, Ladies blouse, Womens sweaters, Womens dress, Womens swimwear, Womens sleepwear

- RUNNING SNEAKERS WITH PERFORATED SOLES man

$\implies$ Men's casual shoes, Men's low shoes, Men's slipper, **Men's sports shoes**

- LEATHER OVERSHIRT woman

$\implies$ Womens blazer, Womens sweaters, **Womens jacket**, Womens sportswear, Womens coat, Womens tights, Ladies shirt, Ladies blouse, Womens pants, Womens ski suit, Womens jeans, Womens skirt, Womens dress, Womens swimwear, Womens sleepwear, Womens underpants, Brassiere

- SOFT-TOUCH SWEATSHIRT WITH PUFF SLEEVES

$\implies$ Womens tights, Womens jeans, Womens ski suit, Womens blazer, Womens jacket, Womens underpants, Womens sportswear, Womens skirt, Womens coat, Ladies shirt, Ladies blouse, **Womens sweaters**, Womens dress, Womens pants, Womens swimwear, Brassiere, Womens sleepwear

- 5-pack bodysuits kids

$\implies$ Children's jacket, Jeans for boys, Boys underpants, **Baby set**, T-shirt for girls