Statistical and Generative Models with Subtitle Extraction for Next Product Title Generation

Team: We Bare Bears

Honghee Lee, Youngjae Chang, Kyuri Choi, Jongwuk Lee, Youngjoong Ko*

SungKyunkwan University

*corresponding author

KDD2023 LONG BEACH, CA — AUGUST 6 - 10

SUNG KYUN KWAN UNIVERSITY 1398

Introduction

Abstract

- Session-based recommendation aims to predict the next item from the user's actions in the ongoing session. It mainly suffers from the cold start item problem, referring to the difficulty in providing accurate recommendations for items with little or no previous interactions. The KDD Cup 2023 Task 3 (next product title generation) addressed this challenge to improve session-based recommendation. This paper proposes an effective solution for the next product title generation using statistical and generative models. In this process, we optimize a model combination strategy that selects the optimal prediction model for each session based on predefined conditions. The title of the last product serves as a fallback when the session does not meet any conditions. We also devise subtitle extraction techniques to identify a common element among multiple predicted titles. Consequently, our team, **We Bare Bears**, has achieved third place in the KDD Cup Task 3 with a BLEU score of 0.26998, demonstrating the effectiveness of our proposed solution.

Challenges

- How to solve cold start products, addressing items with little or no previous interactions.
- How to accurately generate product titles that does not exist in the train sessions.

Methods

- We utilize both the **Markov Chain model** and the **Generative model** to secure a stable prediction performance while addressing the cold start problem.
- We enhance the title search process by predicting multiple title candidates and integrating them using a **subtitle extraction method**.

Data Analysis

Baseline

```
Previous item title
    Kilt Pins | Large Safety Pins 2.5 inch (10 Pack) Heavy Duty Blanket Pins
    Kilt Pins | Large Safety Pins 2.5 inch (10 Pack) Heavy Duty Blanket Pins
    Kilt Pins | Large Safety Pins 3 inch (10 Pack) Heavy Duty Blanket Pins
Next item title
    Kilt Pins | Large Safety Pins 3.5 inch (5 Pack) Heavy Duty Blanket Pins
```

- Some products show up more than once in a session, and items with similar titles often appear one after the other. This pattern helps the system make good predictions about what a user will pick next just by looking at the last item's title.
- So, we set the last item title of the session as a high performing baseline.

Statistical Properties of the Session Dataset

<Highest transition-probability histogram>   <Item frequency histogram>
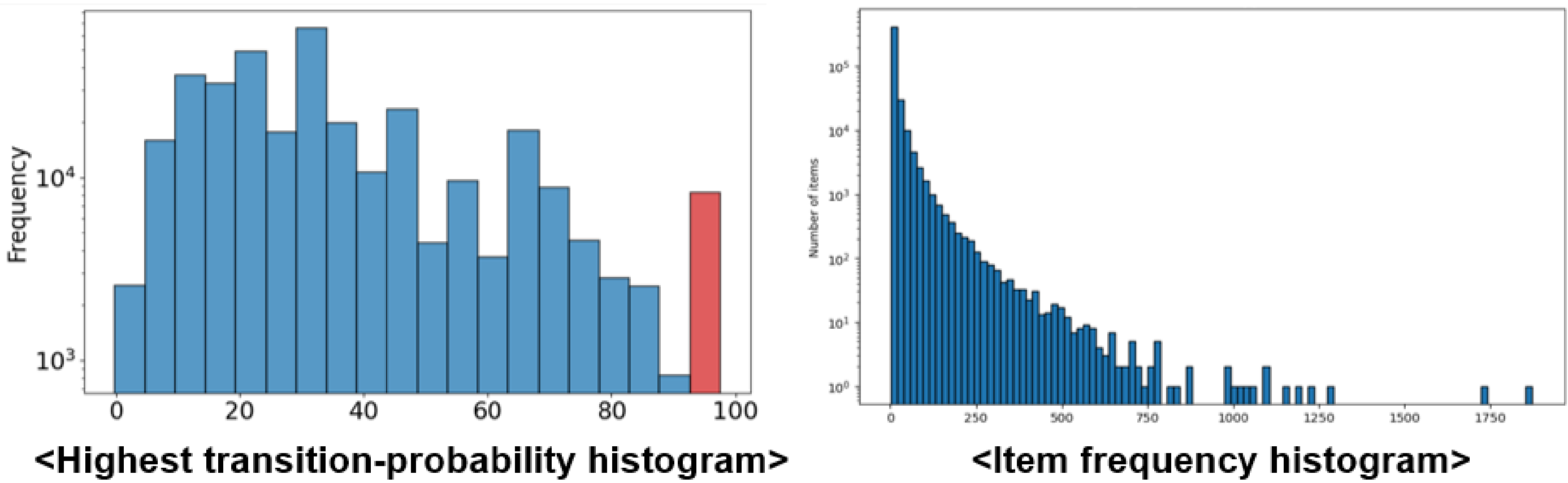
- Users often choose a certain next item for some items, creating a pattern that can be predicted. With the right settings, this pattern can be used to help guess what item a user will choose next, making predictions more accurate.
- We also find that items appearing less than 5 times take about 30% of data for the UK locale. So, we exclude less frequent items as they introduce noise.

Results

Id	Method	BLEU	BP
Base	Last item title	0.26553	1
Stat1	Base + 1st-order MC	0.26584	1
Stat2	Base + 1st-order MC + LCstr	0.26878	1
Stat3	Base + 2nd-order MC + LCstr	0.26764	1
Gen1	Base + gen 'UK'	0.26093	1
Gen2	Base + gen 'UK' + LCpre	0.26699	1
Gen3	Base + gen 'UK' + LCsub	0.26616	0.97870
Best	Base + Stat2 + Gen3	**0.26998**	0.99995

- We find that 1st order Markov Chain is better than the 2nd order version. Also, The Longest Common Subsequence method provides better results than the Longest Common Prefix method. We get best results from integrating the models with the control statements and conditioning.
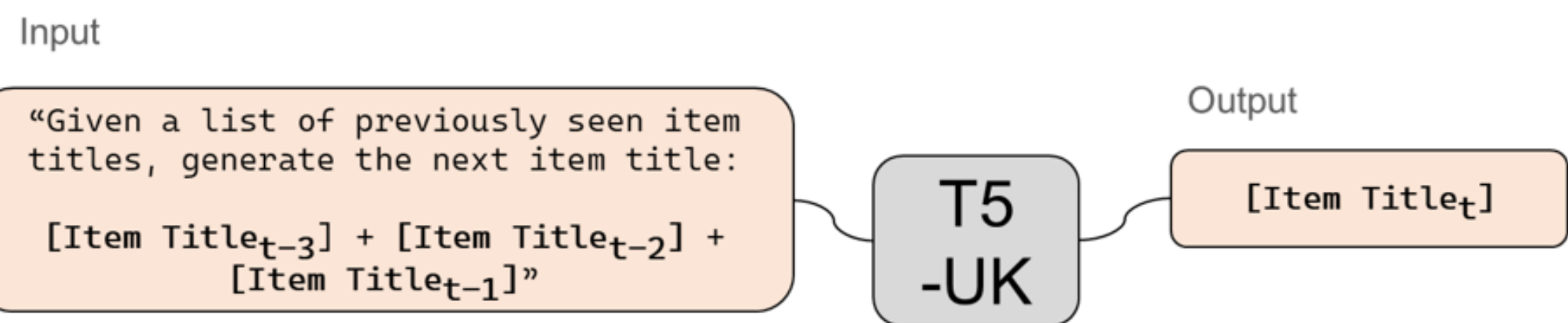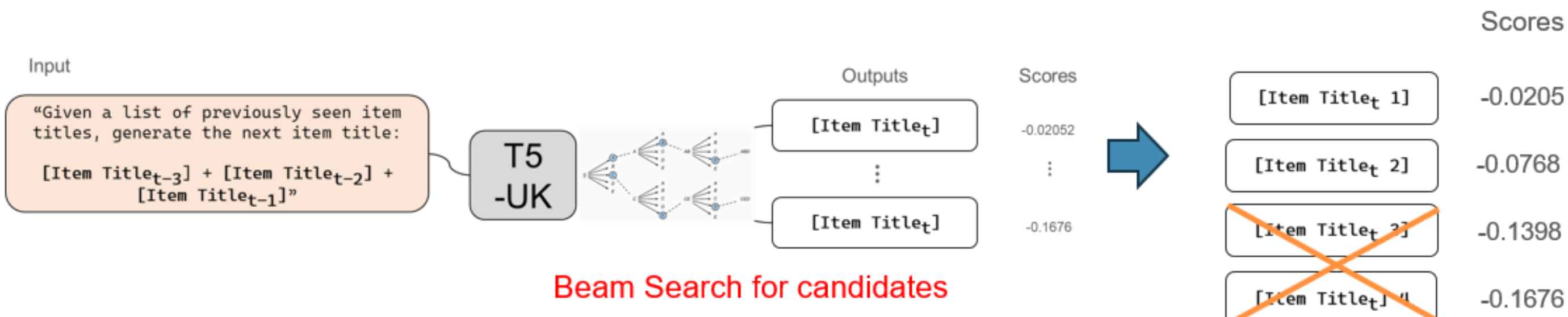
Our Model

Markov Chain Model
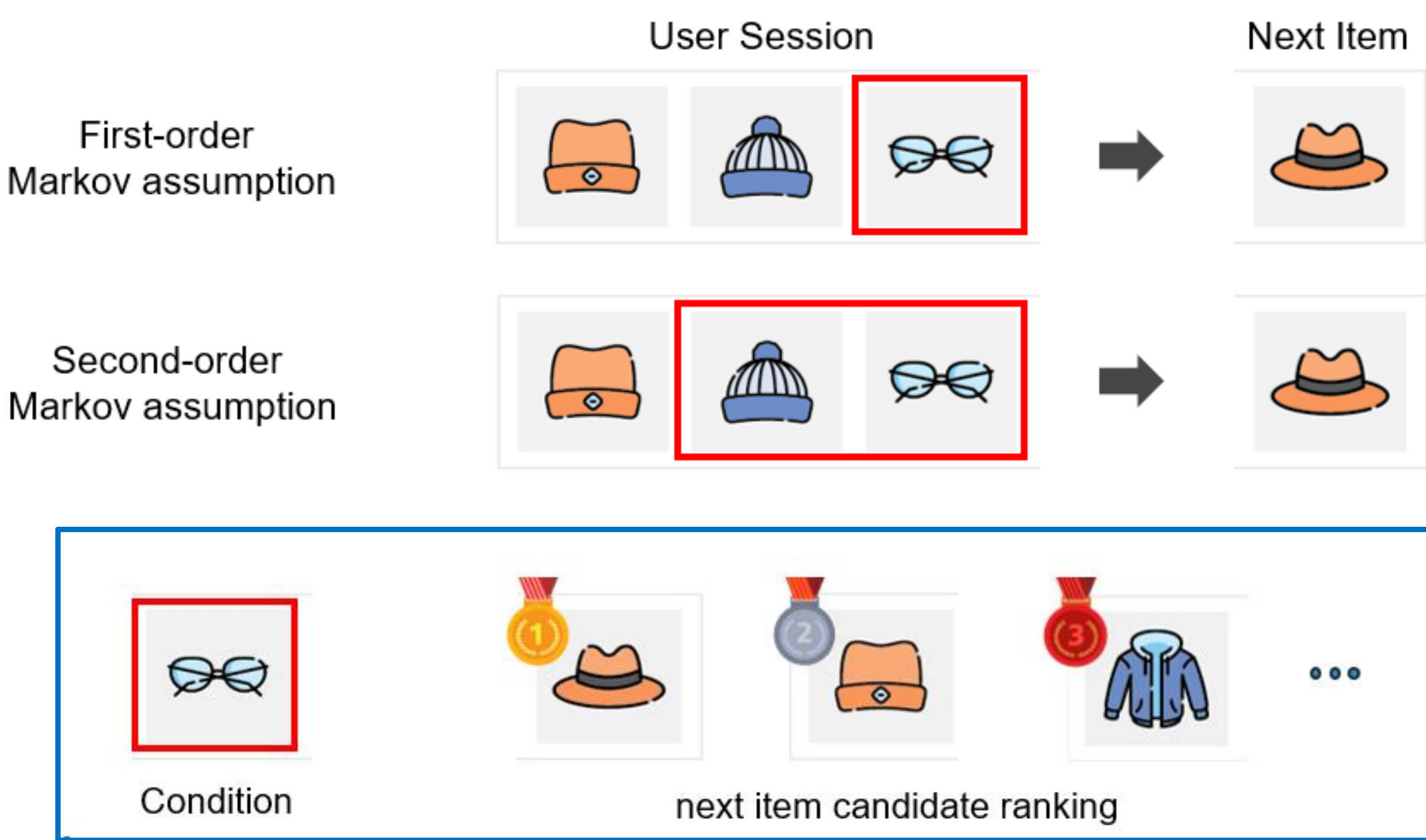
- The Markov Chain Model addresses the characteristic of highly biased item transition probability. For a given condition, the most probable item will be selected as prediction.

Generative Model

Subtitle Extraction Method

Title Candidates
Kilt Pins | Safety Pins 2.5 inch (10 Pack) Heavy Blanket Pins
Kilt Pins | Safety Pins 3 inch (10 Pack) Heavy Blanket Pins

Longest Common Prefix
Kilt Pins | Safety Pins

Longest Common Substring
inch (10 Pack) Heavy Blanket Pins

Longest Common Subsequence
Kilt Pins | Safety Pins inch (10 Pack) Heavy Blanket Pins

- Predicting the complete title is **difficult** so we focus on subtitles.

Integration
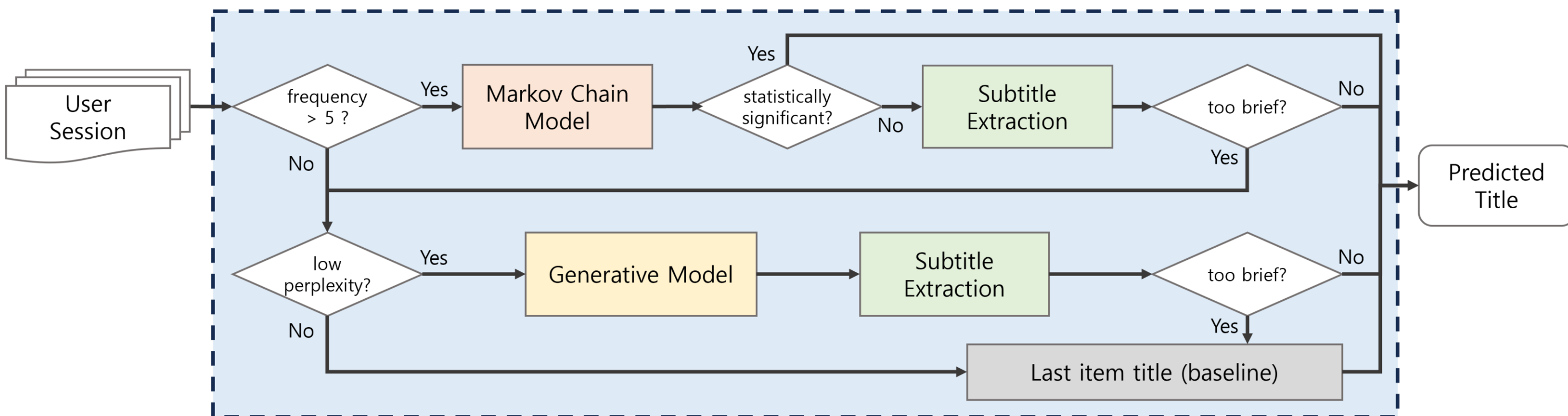
- We use **5 control statements** to direct each session to the correct methodology for prediction.

Conclusion

This paper presents our approach to the Next Product Title Generation task for the KDD Cup 2023. Our solution utilizes statistical and generative models to predict product title candidates and applies subtitle extraction techniques to generate the final predicted title. Our method is based on a thorough analysis of the given dataset. It employs a model combination approach that selects the appropriate model according to the characteristics of the input session. Our solution effectively tackles the cold-start problem and the product repetition within a session. By applying all the proposed methods, we have improved the performance from a baseline BLEU score of 0.26558 to 0.26998. As a result, our solution achieved 3rd place in the KDDCup'23 Challenge Task 3.