



南京大學  
NANJING UNIVERSITY



# W2KPE: Keypphrase Extraction With Word-Word Relation

ICASSP 2023 SPGC — MUG: Track 4

Wen CHENG<sup>1</sup>, Shichen DONG<sup>1</sup>, Wei WANG

**NANJING UNIVERSITY**

<sup>1</sup> Equally Contributed

# Introduction

- Target: extract key phrases from the meeting corpus
  - Keyphrases might not appear in their complete and contiguous form

## Keyphrase

开业筹备

Preparing opening

## Sentence

咱们今天讨论的话题是关于新店开业的筹备问题。

Today we're going to talk about preparing for the opening of our new store.

## Tokenized

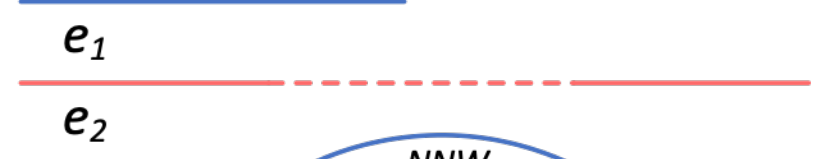
[咱们, 今天, 讨论, 的, 话题, 是, 关于, 新店, 开业, 的, 筹备, 问题]

[Today, we, ##'re, going, to, talk, about, preparing, for, the, opening, of, our, new, store]

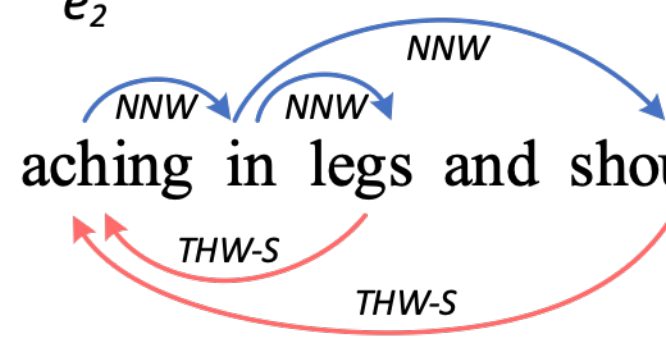
# Introduction

- Word-Word Relation
  - Next-Neighboring-Word(NNW)
  - Tail-Head-Word-\*(THW-\*)
- Apply word-word relation to KPE task
  - Designed for Named-Entity-Recognition task
  - Need extra data processing
  - Need adjustments of model design

(a) I am having aching in legs and shoulders

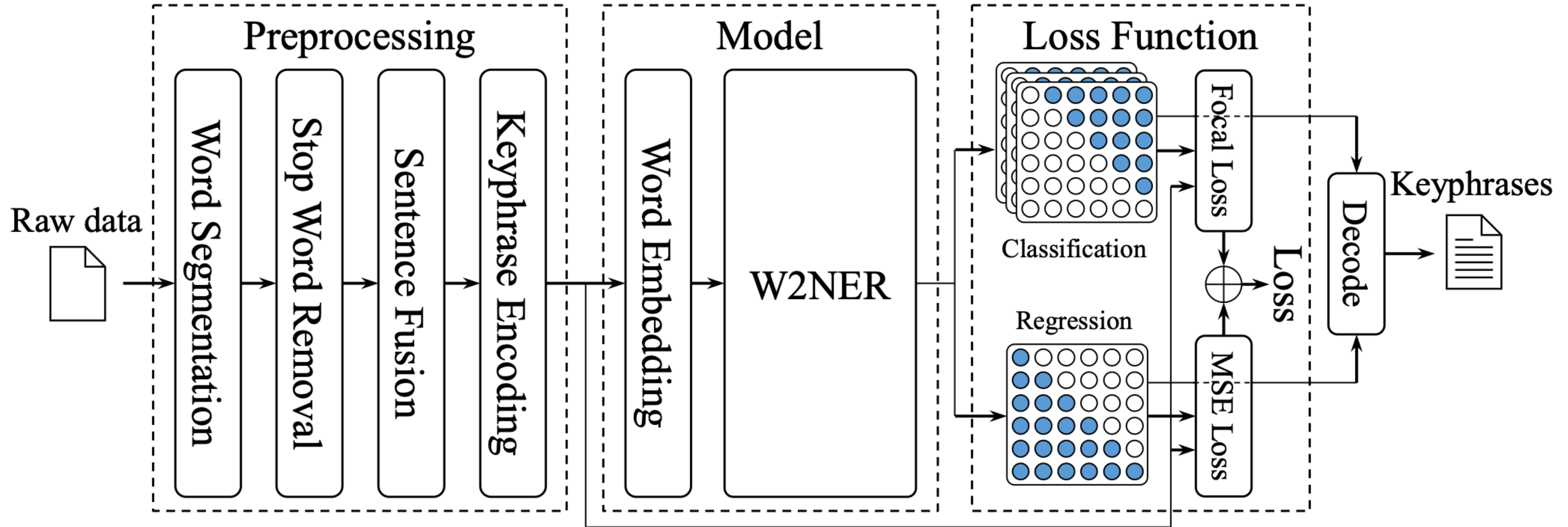


(b) I am having aching in legs and shoulders



	I	am	having	aching	in	legs	and	shoulders
I		■	■	■	■	■	■	■
am			■	■	■	■	■	■
having				■	■	■	■	■
aching					NNW	■	■	■
in						NNW	■	NNW
legs				THW-S			■	■
and								■
shoulders				THW-S				

# System Description



# Data Preprocessing

- Word Segmentation and Stop Word Removal
  - To increase information density

## Sentence

咱们今天讨论的话题是关于新店开业的筹备问题。

Today we're going to talk about preparing for the opening of our new store.

## Tokenized

[咱们, 今天, 讨论, 的, 话题, 是, 关于, 新店, 开业, 的, 筹备, 问题]

[Today, we, ##'re, going, to, talk, about, preparing, for, the, opening, of, our, new, store]

## SW removed

[咱们, 今天, 讨论, 话题, 新店, 开业, 筹备, 问题]

[Today, we, going, talk, preparing, opening, our, new, store]

# Data Preprocessing

- Sentence Fusion

- To increase the amount of information fed to the model within an iteration

## SW removed sentences:

- #1 [大家好, 咱们, 今天, 讨论, 话题, 新店, 开业, 筹备, 问题]
- #2 [这里, 问题, 比较, 多, 咱们, 一项, 一项, 讨论, 首先, 第一个, 问题]
- #3 [我们, 四个, 人, 如果, 开, 饭店, 大家, 觉得, 需要, 一些, 什么, 材料]
- #4 [哪方面, 着手, 准备, 这件事]
- #5 [开店, 有没有, 开, 饭店, 需要, 加盟, 属于, 自营]
- #6 [两个, 问题, 一个, 自营, 一个, 加盟]

## Fused segments:

- #1 [大家好, 咱们, 今天, 讨论, 话题, 新店, 开业, 筹备, 问题, 这里, 问题, 比较, 多, 咱们, 一项, 一项, 讨论, 首先, 第一个, 问题, 我们, 四个, 人, 如果, 开, 饭店, 大家, 觉得, 需要, 一些, 什么, 材料]
- #2 [哪方面, 着手, 准备, 这件事, 开店, 有没有, 开, 饭店, 需要, 加盟, 属于, 自营, 两个, 问题, 一个, 自营, 一个, 加盟]

# Data Preprocessing

- Keyphrase Encoding
  - To address the issue of keyphrase being incomplete or incontiguous

## Incomplete Example

**Keyphrase** 淘宝直播带货

**Tokens** [关于, 淘宝, 直播, 卖货, 方面, 看法]  
 $t_1$   $t_2$   $t_3$   $t_4$   $t_5$   $t_6$

**Encoded** [ $t_2, t_3$ ] (completeness=  $4/6 = 0.67$ )

## Incontiguous Example

**Keyphrase** 开业筹备

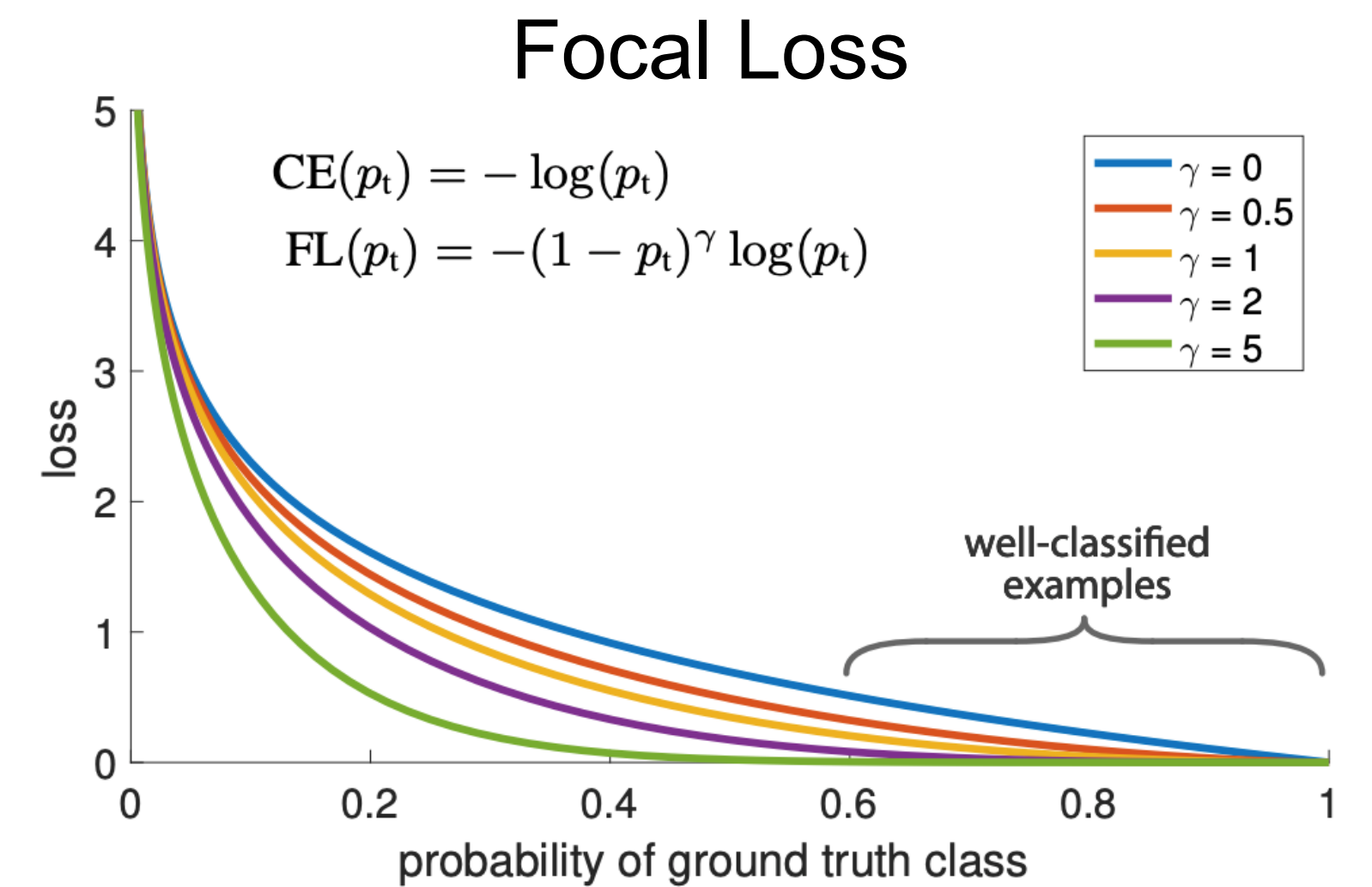
**Tokens** [关于, 新店, 开业, 的, 筹备, 问题]  
 $t_1$   $t_2$   $t_3$   $t_4$   $t_5$   $t_6$

**Encoded** [ $t_3, t_5$ ] (completeness=  $4/4 = 1.0$ )

# Model Design

- Focal Loss
  - To address the issue of imbalanced class distribution

$$\text{loss} = \alpha \underbrace{\left[ \sum_{c \in \text{classes}} -(1 - p_c^*)^\gamma \log(p_c^*) \right]}_{\text{Classification Loss}} + (1 - \alpha) \underbrace{\left[ \sum_{k \in \text{keyphrases}} (p_k - y_k)^2 \right]}_{\text{Regression Loss}}$$





# Model Design

- Keyphrase Scoring
  - To rank the output keyphrases

Tokens	[开, 饭店, 需要, 加盟, 属于, 自营, 两个, 问题, 一个, 自营, 一个, 加盟]
Output scores	0 0 0 0.8 0 0.7 0 0 0 0.6 0 0.9
Final scores	“加盟” = 0.8 + 0.9 = 1.7 “自营” = 0.7 + 0.6 = 1.3

# Evaluation & Conclusion

- Four main improvements
  - Data processing
    - Sentence Fusion
    - Keyphrase Encoding
  - Model design
    - Focal Loss
    - Keyphrase Scoring
- Final score
  - 45.88 on test set stage 1
  - 45.07 on test set stage 2
  - 47.69 on dev set

Experimental Config	Score
<b>W2KPE</b>	<b>47.69</b>
- Sentence Fusion	41.83(-5.86)
- Keyphrase Encoding	45.16(-2.53)
- Focal Loss	46.94(-0.75)
- Keyphrase Scoring	47.05(-0.64)
Baseline	41.48



南京大學  
NANJING UNIVERSITY

# Thanks for listening

## W2KPE: Keyphrase Extraction With Word-Word Relation

Wen CHENG<sup>1</sup>, Shichen DONG<sup>1</sup>, Wei WANG

**NANJING UNIVERSITY**