Research on Time-sync Danmaku data

Background: Crowdsourced Time-sync Video comment

- Millions of videos generated everyday;
- Political debate organizers adjust their strategies by analyzing users' comments.
- A director of a TV show can optimize his plays by analyzing the topics and feedback about the videos
- …and more.

Background: Crowdsourced Time-sync Video comment

Movie captions

comments/twitter

Time-sync comments

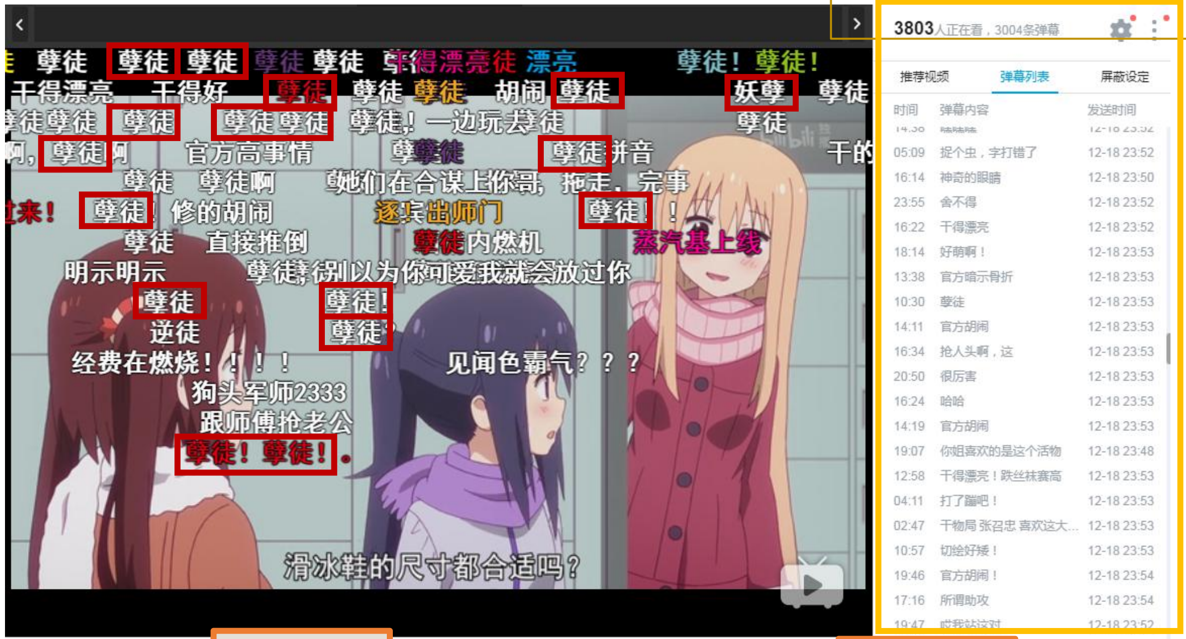
- Traditional methods utilize low-level features (comments, twetter, etc.);

Drawbacks:

- (1) it is difficult to capture the segment information since the comments are based on entire video;
- (2) we do not know what specific content of the video causes the point of view.

Background: Key observations of Danmaku comments

History Comments



History Comments

10.30''

User A: Sinners（孽徒）
User B: tomfoolery（胡闹）

10.34''

User B: Sinners
User C: Sinners
User D: Sinners

14.04''

User A: Well done
User B: Well done

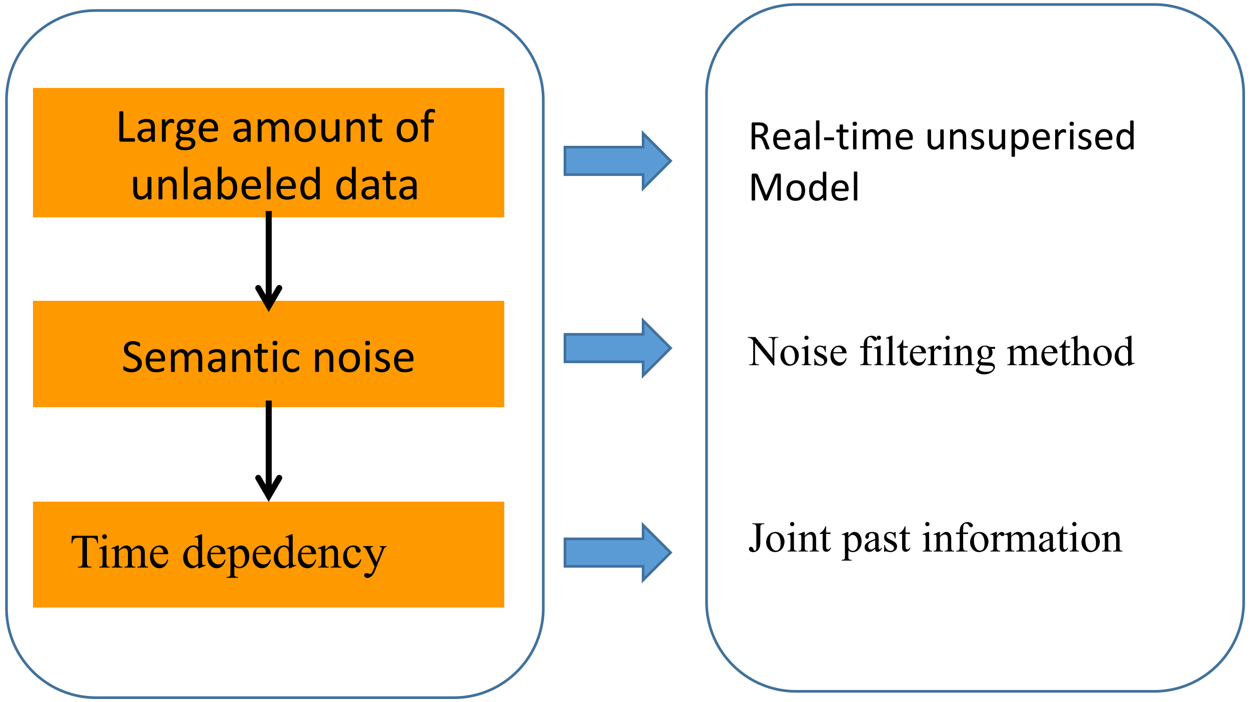
14.13''

User C: rewards
User D: Well done

Problems with danmaku

- Large amount of unlabeled data: comment is usually short; large amounts of comments;
- Semantic dependency;
- Semantic noise : some comments are irrelevant to the shots; Slang lauguage; follower brush phenomenon

Summary


Large amount of unlabeled data
Semantic noise
Time depedency
Real-time unsuperised Model
Noise filtering method
Joint past information

Related work: Video Tagging

Crowdsourced Time-sync Video Tagging using Temporal and Personalized Topic Modeling

KDD 14

Bin Wu[1], Erheng Zhong[1], Ben Tan[1], Andrew Horner[1], Qiang Yang[1,2]
[1]Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong
[2]Noah's Ark Lab, Huawei, Hong Kong

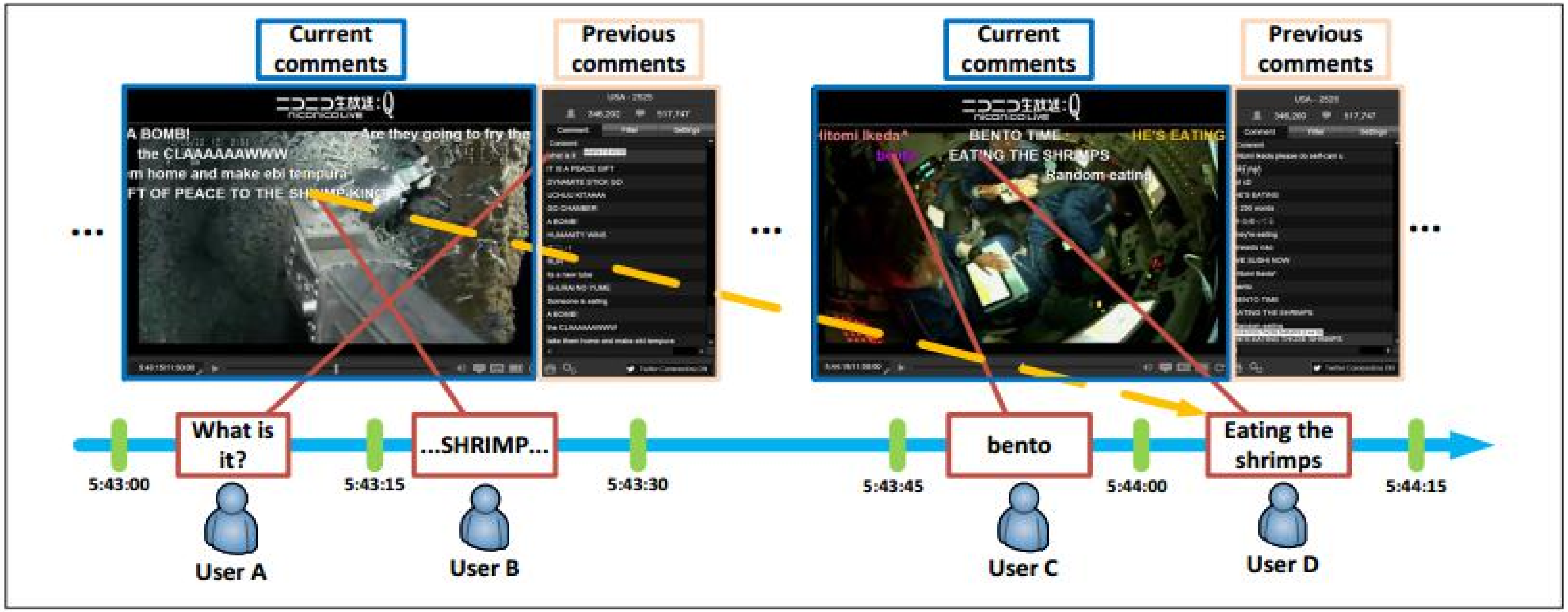


Figure 2: Snapshot of an example Time-Sync Commented (TSC) video. Users share their watching experience by providing time-sync comments that appear on the screen.

temporal and personalized topic model

Related work: Video Summarization

Bridging Video Content and Comments: Synchronized Video Description with Temporal Summarization of Crowdsourced Time-Sync Comments

AAAI 2017

Linli Xu, Chao Zhang

Table 2: The average F-measures of ROUGE-1 and ROUGE-2. A bold number indicates the highest ROUGE score.

	QPS-EP06		QPS-EP07		QPS-EP12		LYB-EP07		LYB-EP26		LYB-EP40	
	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2
Random	0.2353	0.0526	0.2374	0.0583	0.2443	0.0616	0.2694	0.0583	0.2611	0.0657	0.2573	0.0547
ClusterHITS	0.2808	0.0985	0.2816	0.0993	0.3033	0.1045	0.3108	0.1013	0.3213	0.1167	0.3087	0.0973
LexRank	0.2787	0.0968	0.2932	0.1082	0.2908	0.1026	0.3076	0.0966	0.3142	0.1128	0.2985	0.0926
DSDR	0.3356	0.1172	0.3407	0.1218	0.3346	0.1288	0.3386	0.1207	0.3511	0.1202	0.3353	0.1198
TopicDSDR	0.2793	0.0977	0.2976	0.1116	0.3067	0.1102	0.3150	0.1064	0.3052	0.1097	0.3043	0.0954
SJTTR	**0.3682**	**0.1375**	**0.3761**	**0.1463**	**0.3874**	**0.1404**	**0.3775**	**0.1413**	**0.3913**	**0.1475**	**0.3704**	**0.1385**
T-SJTTR	**0.3758**	**0.1402**	**0.3895**	**0.1525**	**0.3982**	**0.1486**	**0.3969**	**0.1517**	**0.4095**	**0.1583**	**0.3843**	**0.1447**

Table 3: Selected comments and the corresponding video plots along the timeline of a segment in “LYB-EP26”

 22:11	 23:06	 24:20	 25:27	 25:51
Nian Nian has slanted eyebrows.	Mr. Su used to be Brother Su. How sad.	I feel sorry for both Jing Rui and Su from their conversation.	Mr. Su lost a good friend forever!	Jing Rui leaves the place where his heart broke and dreams faded away.
keywords by LDA: ‘Su’ , ‘say’ , ‘really’ , ‘Jing Rui’ , ‘leave’ , ‘eyebrow’ , ‘Nian Nian’ , ‘come’ , ‘friends’ , ‘love’				

Related work: Video Summarization

LiveBot: Generating Live Video Comments Based on Visual and Textual Contexts

AAAI 2019

Shuming Ma[1]*, Lei Cui[2], Damai Dai[1], Furu Wei[2], Xu Sun[1]
[1]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2]Microsoft Research Asia

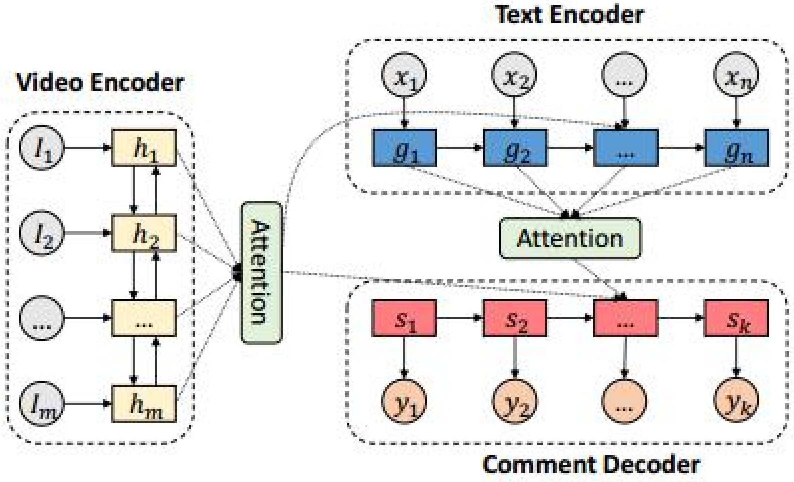


Figure 5: An illustration of Fusional RNN Model.

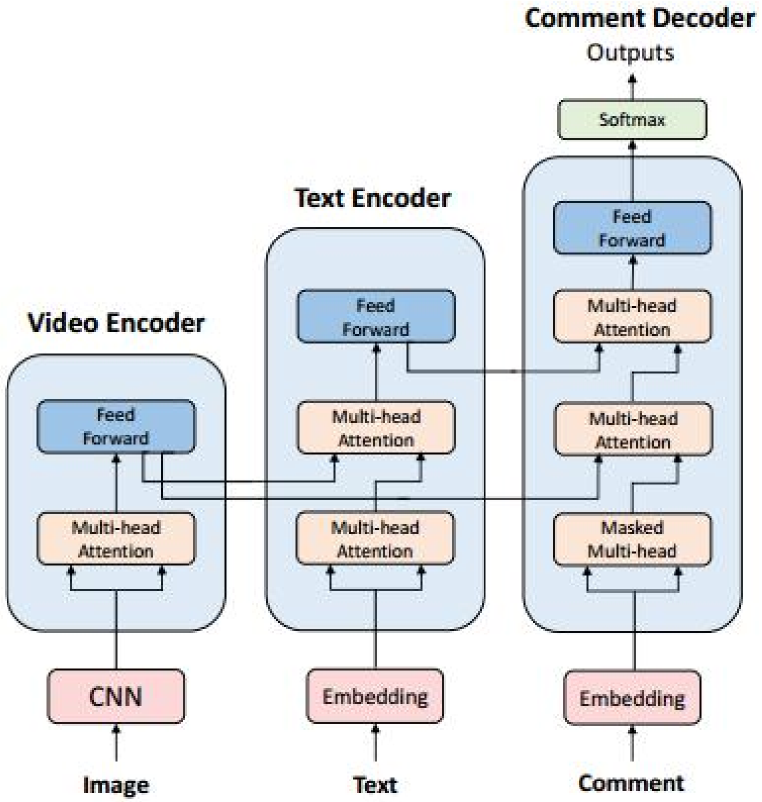


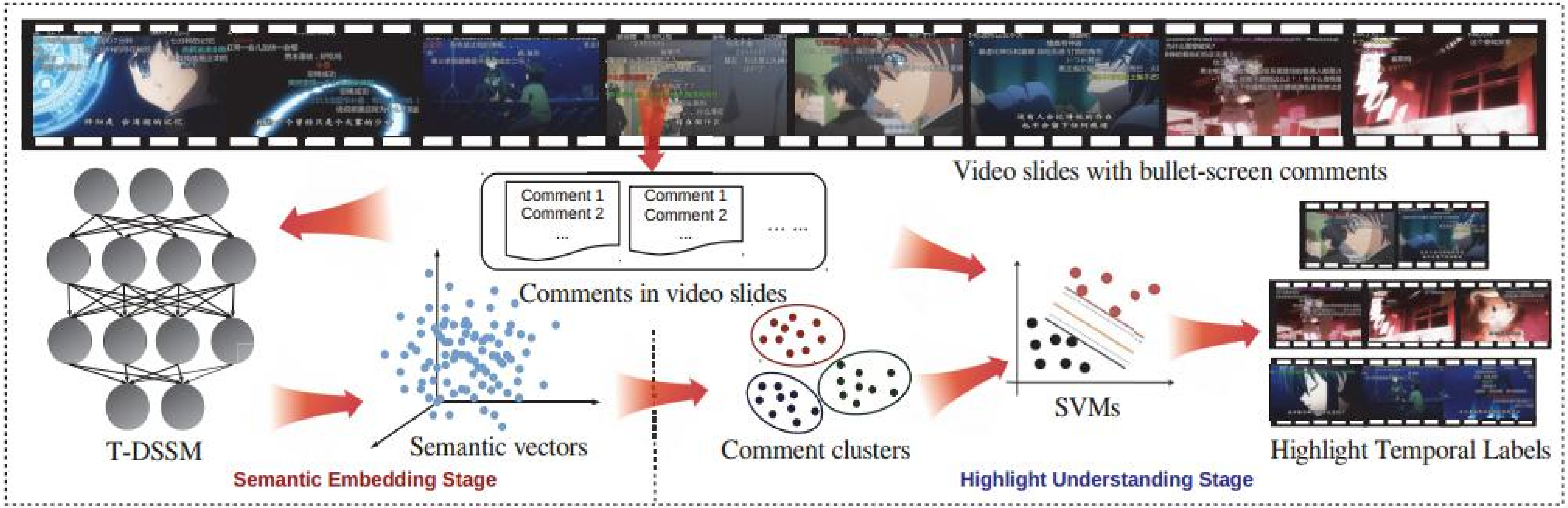
Figure 6: An illustration of Unified Transformer Model.

Related work: Video Highlight

Reading the Videos: Temporal Labeling for Crowdsourced Time-Sync Videos Based on Semantic Embedding

AAAI-16

Guangyi Lv†, Tong Xu†, Enhong Chen†*, Qi Liu†, Yi Zheng‡
†School of Computer Science and Technology, University of Science and Technology of China
gylv@mail.ustc.edu.cn, tongxu@mail.ustc.edu.cn, cheneh@ustc.edu.cn, qiliuql@ustc.edu.cn
‡Ant Financial Services Group
aliguagua.zhengy@alipay.com



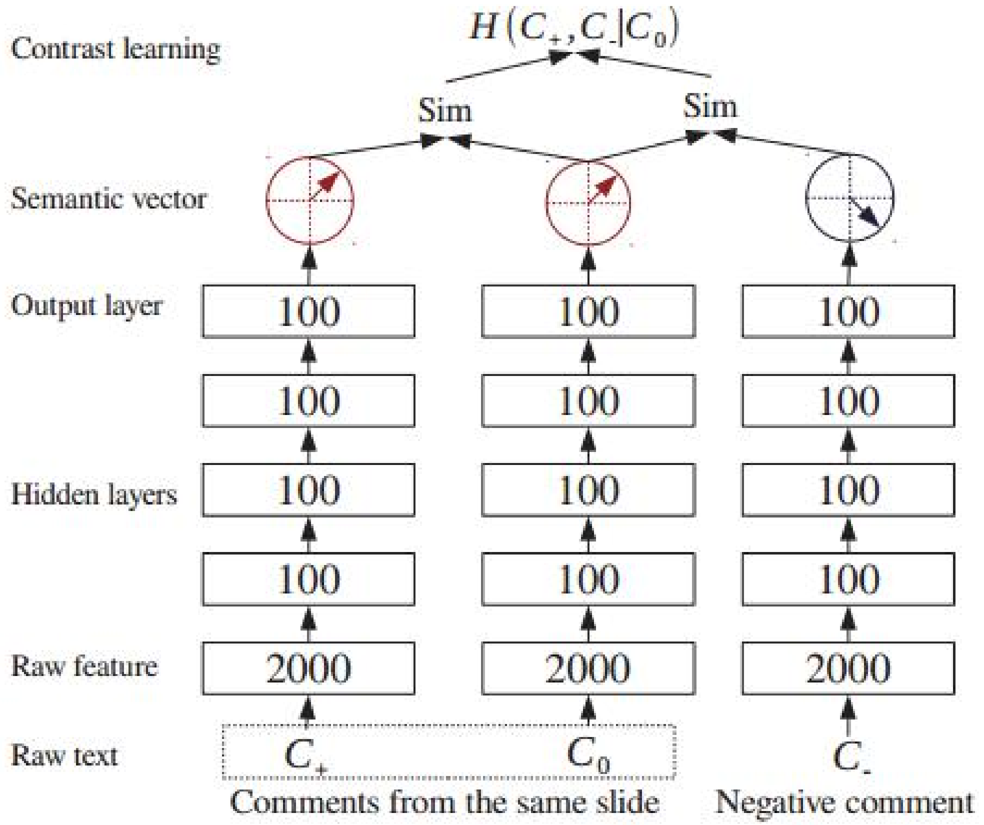


Figure 1: The architecture of T-DSSM with 3 hidden layers.

Our work：Time-sync Danmaku Dataset

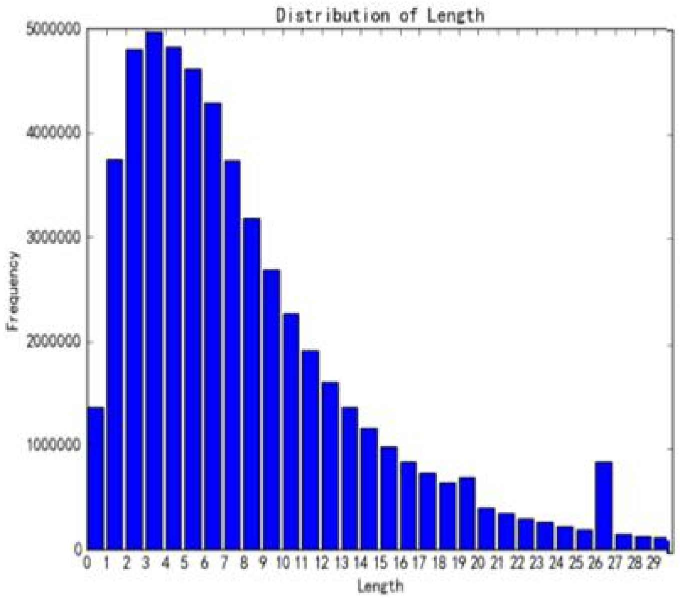
Time-sync Danmaku Dataset

Table 1: The statistics of danmaku datasets. (1) #Seasons,#Videos, #Danmaku, #Users, #Tags: the number of seasons,videos, Danmaku reviews , uses and tags respectly; (2)#Comments: the number of traditional comments about the video; (3) #UCS : the number of users of traditional comments; (4) Summarization: video summarization; (5)Sentiment Tag: the labeled sentiment class.

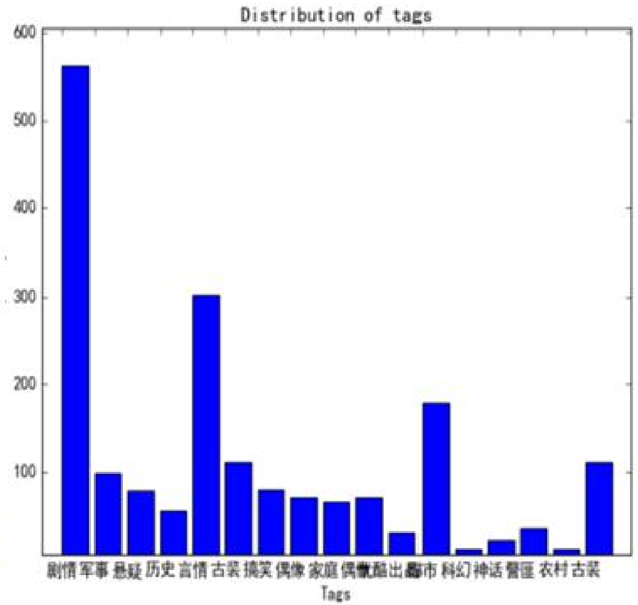
	Paper [3]	Paper [14]	Paper [2]	Paper [11]	Paper [15]	Paper [15]
Source	acfun.tv	acfun.tv	iqiyi	bilibili	acfun.tv	acfun.tv
#Seasons	×	×	2	×	×	×
# Videos	6,506	6	7166	6	120	120
# Danmaku	1,704,930	234,003	11,842,166	52,174	227,780	227,780
# Users	320,000	×	1,133,750	×	×	×
#Tags	×	×	×	×	3	3
Likes	×	×	×	×	√	√
#Comments	×	×	×	×	×	×
#UCS	×	×	×	×	×	×
Summarization	×	×	×	×	×	×
Sentiment Tag	×	×	×	×	×	×

	Paper [5]	Paper [12]	Paper [8]	Paper [4]	DR_Four	DR_E
Source	bilibili	acfun.tv	bilibili	bilibili	youku	youku
#Seasons	64	×	×	×	×	612
# Videos	716	16,414	×	3,623	4,517	**8,156**
# Danmaku	7,413,517	1,103,884	133,250	60,956	32,950,000	**57,176,457**
# Users	1,482,120	382,752	×	278,520	5,412,000	**6,259,558**
#Tags	42	×	×	×	×	**17**
Likes	√	×	×	×	×	√
#Comments	×	×	×	×	×	**2,170,033**
#UCS	×	×	×	×	×	**1,264,811**
Summarization	×	×	×	×	×	√
Sentiment Tag	×	×	×	×	×	√

Statistics information



a) Distribution of danmaku length.



b) Distribution of Chinese tags.

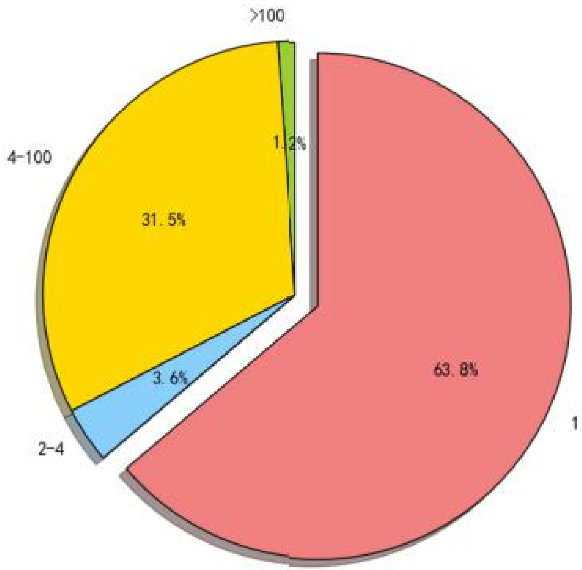
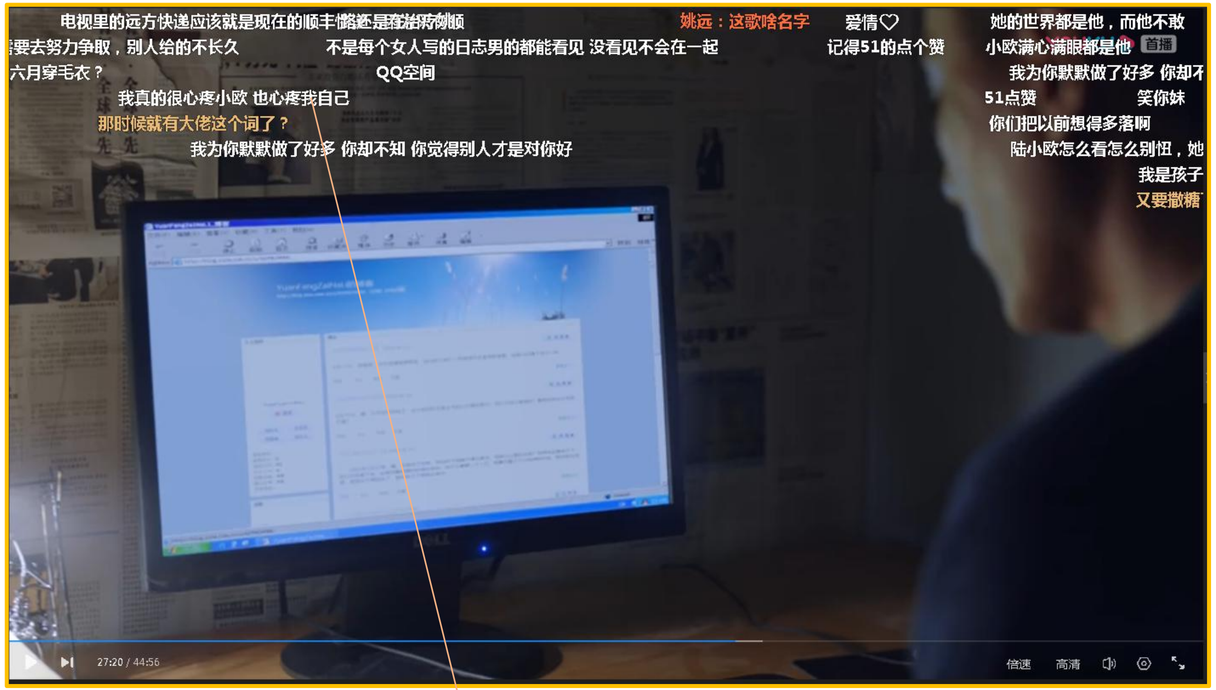


Fig. 4: Distribution of danmaku users.

Plan：Matching Video Storyline and Comments



10.30''

路队看到远方的简陋环境还有给晓欧住的房间，便立刻对阿远发起火来，他觉得姚远是在利用和欺骗晓欧为他自己卖命，阿远知道自己委屈了晓欧，心里也十分愧疚。

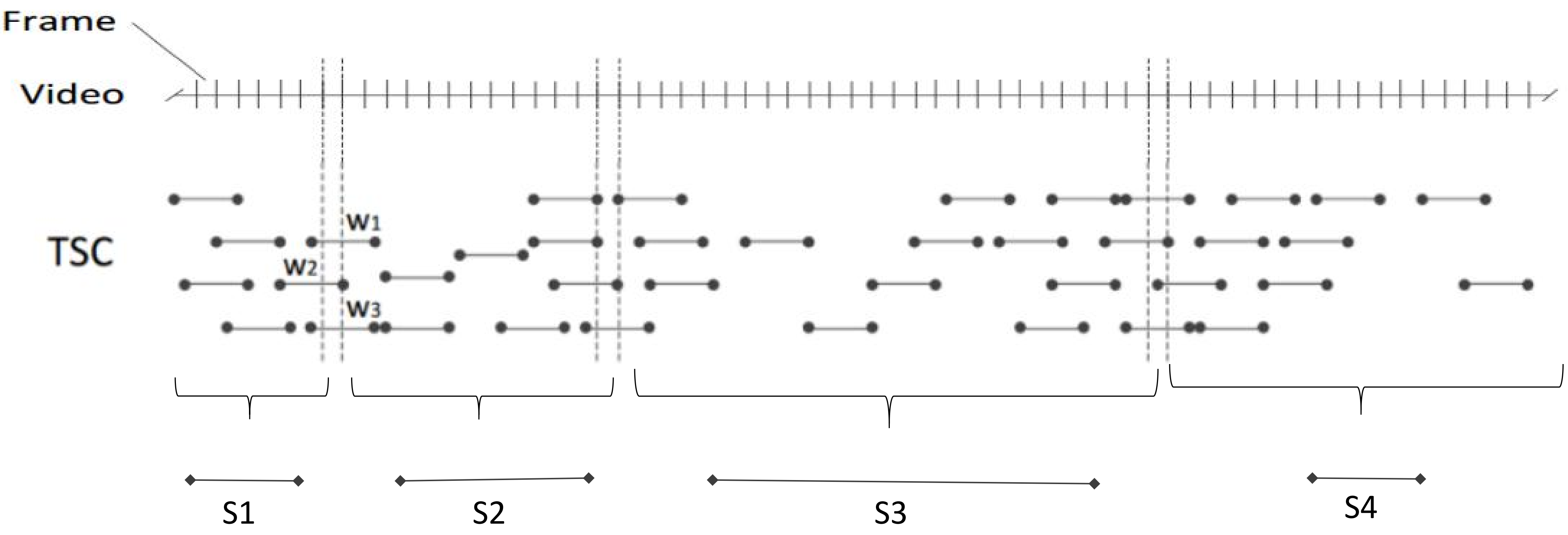
20.34''

阿远连忙跑到了火车站，他不明白爱莲为什么要走，爱莲只对他说自己想回去过自己生活。

Large amount of unlabeled data → Real-time unsuperised Model

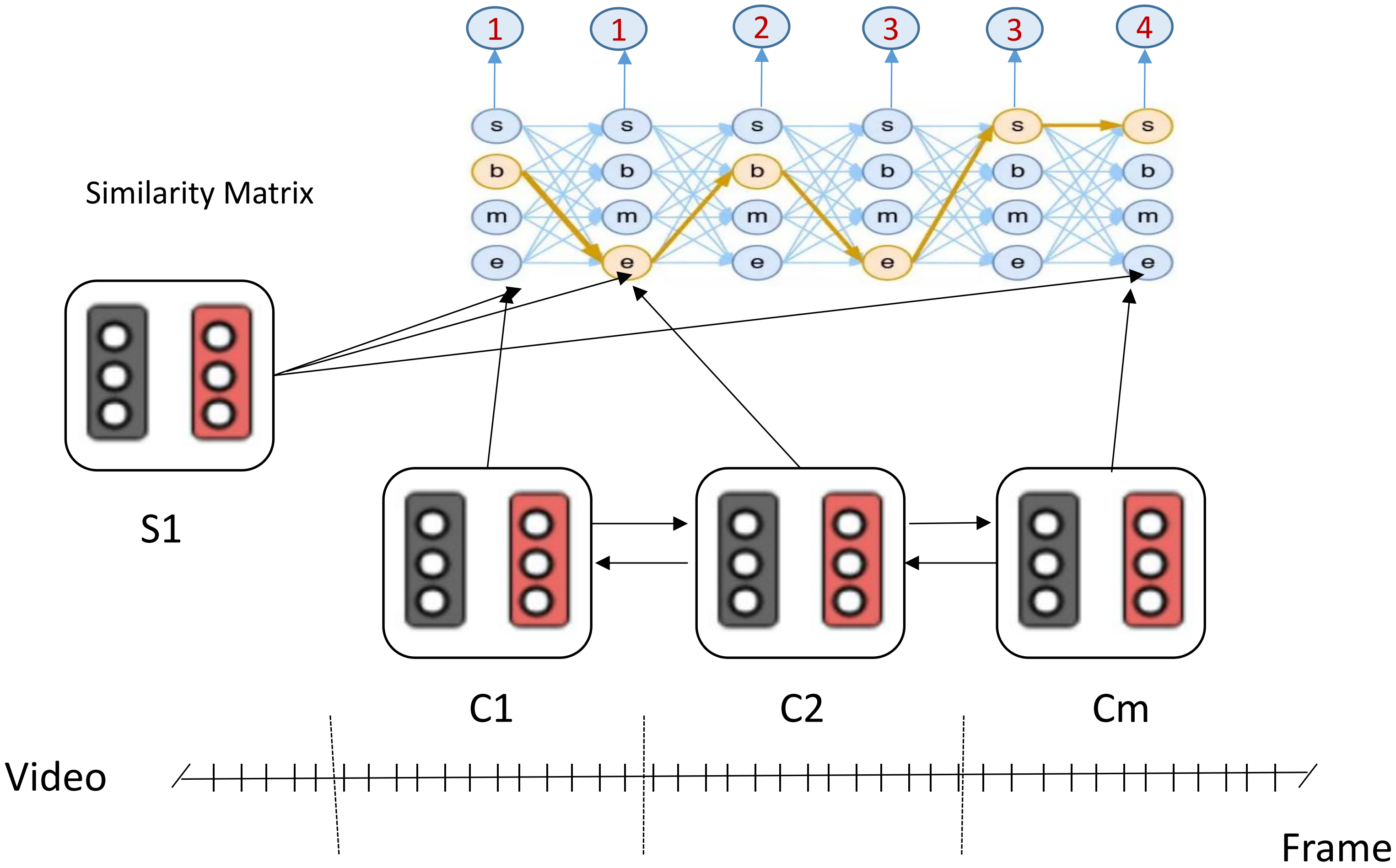
Semantic noise → Noise filtering method

Time depedency → Joint past information



In the unpaired Video Storyline and Comments matching tasks, we have a dataset of Storyline $S = \{s_1, s_2, ..., s_N\}$, and a dataset of danmaku comment sentences $C = \{c_1, c_2, ..., c_M\}$, where M and N are the total numbers of storyline and comments, respectively. In this setting, there is no alignment between C and S. Our goal is to train an Matching model to bridge the relation between storyline and comments.

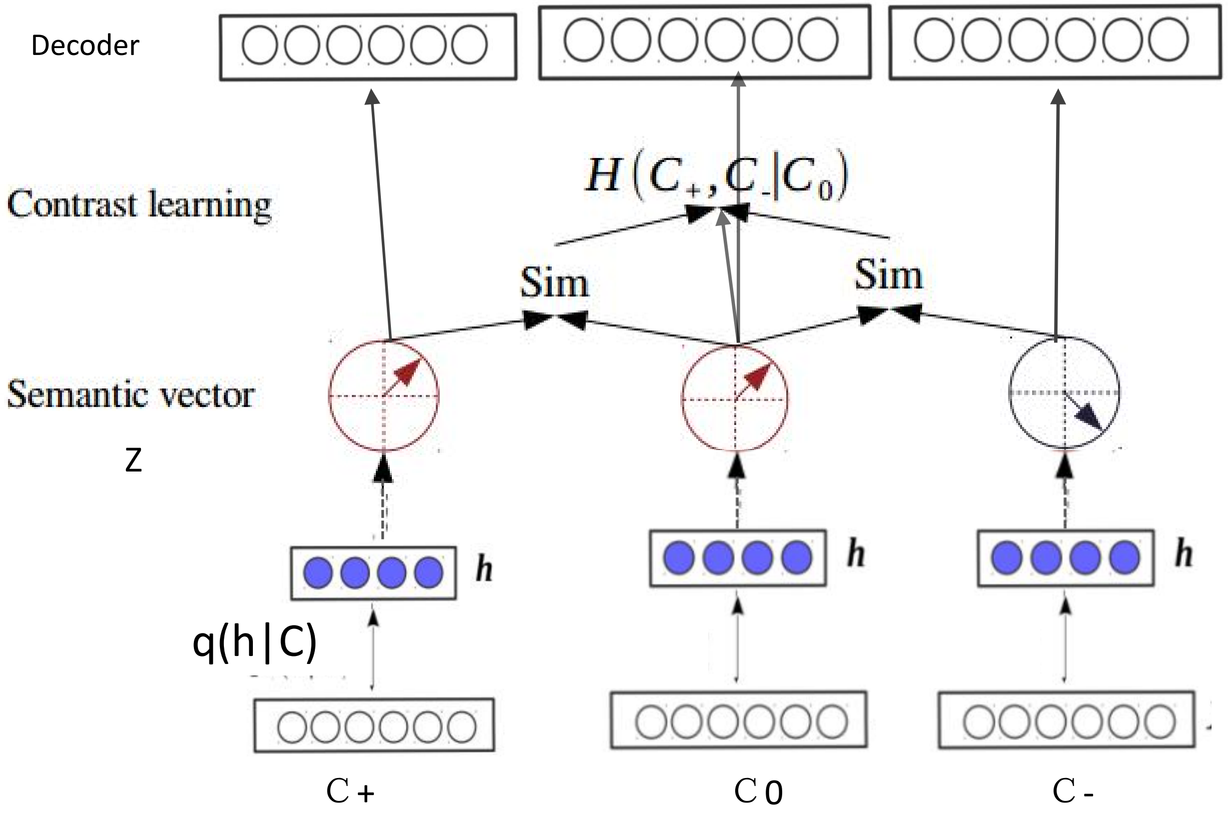
Plan：Matching Video Storyline and Key Frame Comments



Step 1: Temporal Semantic Segmention
Step 2. Computing Semantic similarity between storyline and danmu comment
Step 3. Finding Optimal Segmention by Dynamic Programming

Cosin-Sim-TF-IDF   Cosin-Sim-WordEmbeding

Plan：Matching Video Storyline and Key Frame Comments



e n coder

$$
\begin{aligned}
\boldsymbol{c}_i &= \text{counts}(\boldsymbol{x}_i) && (6)\\
\boldsymbol{h}_i &= \text{MLP}(\boldsymbol{c}_i) && (7)\\
\boldsymbol{\mu}_i &= f_\mu(\boldsymbol{x}_i) = \boldsymbol{W}_\mu \boldsymbol{h}_i + \boldsymbol{b}_\mu && (8)\\
\boldsymbol{\sigma}_i &= f_\sigma(\boldsymbol{x}_i) = \exp(\boldsymbol{W}_\sigma \boldsymbol{h}_i + \boldsymbol{b}_\sigma) && (9)\\
\boldsymbol{z}_i^{(s)} &= \boldsymbol{\mu}_i + \boldsymbol{\sigma}_i \cdot \boldsymbol{\varepsilon}^{(s)}. && (10)
\end{aligned}
$$

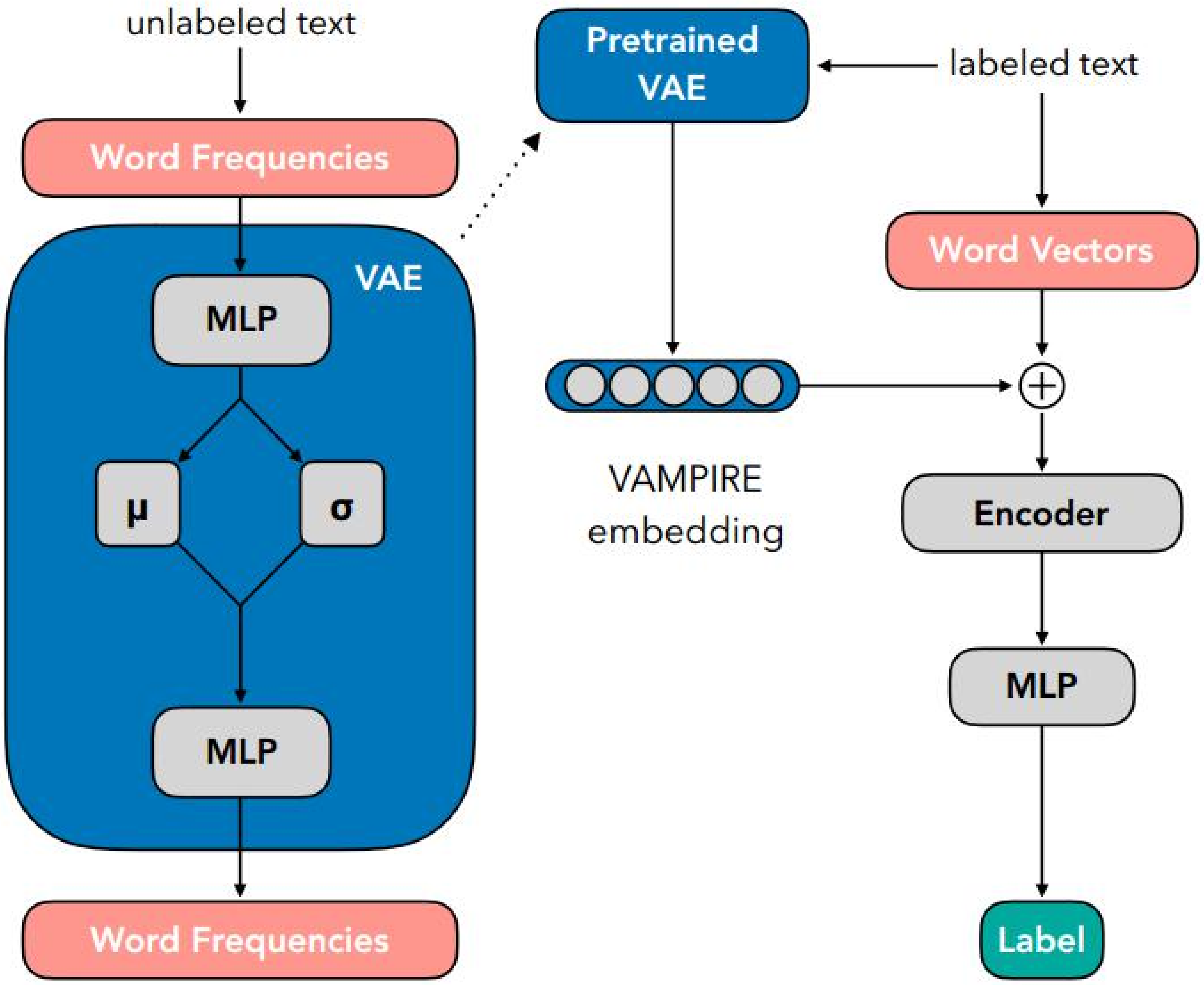
decoder

$$
\begin{aligned}
\boldsymbol{\theta}_i &= \text{softmax}(\boldsymbol{z}_i^{(s)}) && (11)\\
\boldsymbol{\eta}_i &= \text{softmax}(\boldsymbol{b} + \boldsymbol{B}\boldsymbol{\theta}_i) && (12)
\end{aligned}
$$

$$
\log p(\boldsymbol{x}_i \mid \boldsymbol{z}_i^{(s)}) = \sum_{j=1}^{V} \boldsymbol{c}_{ij} \cdot \log \boldsymbol{\eta}_{ij}, \quad (13)
$$

where j ranges over the vocabulary.

[ACL 2019] Variational Pretraining for Semi-supervised Text Classification



1. Unsupervised Pretraining

2. Model Selection via Topic Coherence

3. Using a Pretrained VAE for Text Classification

Figure 1: VAMPIRE involves pretraining a deep variational autoencoder (VAE; displayed on left) on unlabeled text. The VAE, which consists entirely of feed-forward networks, learns to reconstruct a word frequency representation of the unlabeled text with a logistic normal prior, parameterized by μ and σ. Downstream, the pretrained VAE's internal states are frozen and concatenated to task-specific word vectors to improve classification in the low-resource setting.

[ACL 2019]Variational Pretraining for Semi-supervised Text Classification.Suchin Gururangan, Tam Dang, Dallas Card and Noah A. Smith

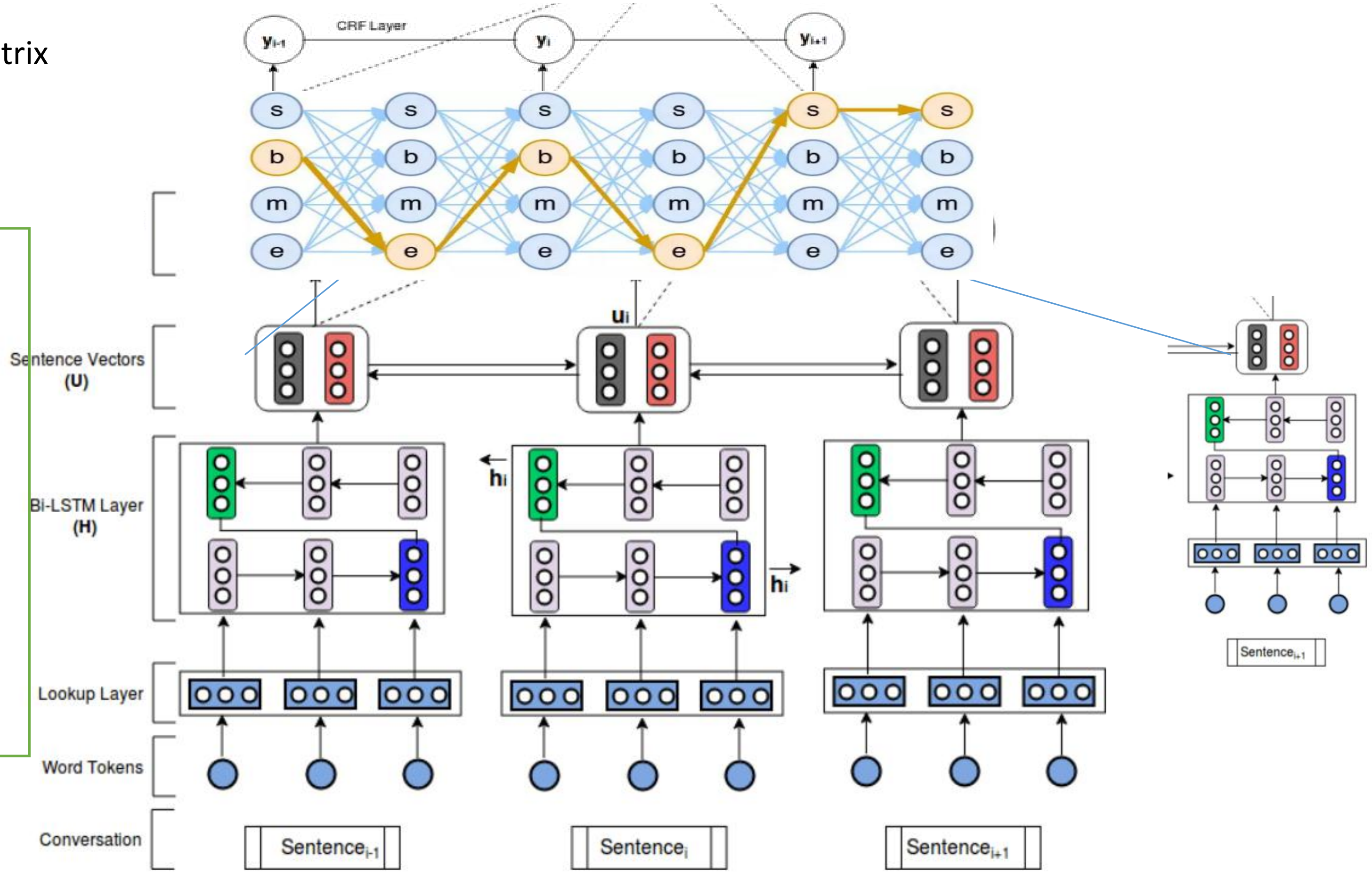
Plan：Matching Video Storyline and Key Frame Comments

Similarity Matrix

BiLSTM+CRF

BiLSTM+Cos+CRF

BiLSTM+Cos+DP



Thanks