WI-IAT 2021

A Learning Framework with Disposable Auxiliary Networks for Early Prediction of Product Success

Chih-Ting Yeh[1], Sheng-Chieh Lin[1], Zhe-Li Lin[1], Jing-Kai Lou[3] Ming-Feng Tsai[2] and Chuan-Ju Wang[1]

Presenter: (Justin) Chih-Ting Yeh

[1] Research Center for Information Technology Innovation, Academia Sinica,
[2] Department of Computer Science, National Chengchi University
[3] KKStream Limited

Table of Contents

Introduction

Motivation

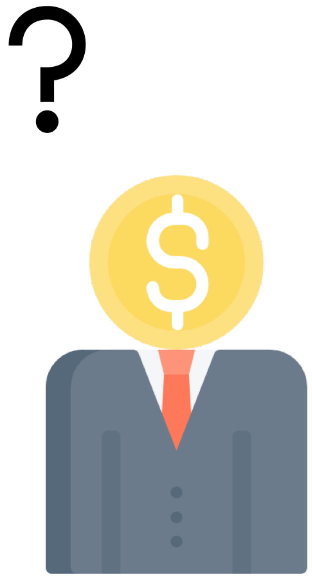
A common business issue: **Identify successful products <u>at an early stage</u>**

Motivation

A common business issue: **Identify successful products <u>at an early stage</u>**

What characteristic better represents a product?

What features make a product popular?



Motivation

A common business issue: **Identify successful products at an early stage**

What characteristic better represents a product?

What features make a product popular?

~~User feedbacks~~

Motivation

Several studies [1] [2] show that **descriptions are useful features for predicting product success.**

Royce's chocolate has become a **standard** Hokkaido **souvenir**. They are packaged one by one so your hands won't get dirty! Also, our **staff** recommends this product!
北海道のお土産で定番品となっているロイズ. 手が汚れないように1本ずつパッケージされているのもありがたい! 当店 スタッフもおすすめするロイズの自信作です！

Four types of nuts: almonds, cashews, pecans, macadamia, as well as cookie crunch and almond puff were packed carefully into each chocolate bar. This item is shipped with a refrigerated courier service during the **summer**.
アーモンド、カシュー、ペカン、マカダミアの4種類のナッツとクッキークランチやアーモンドパフを一本のチョコレートバーにぎっしり詰め込みました。こちらは夏期クール便発送商品です。

- In [1], two descriptions for the same product were compared.
  The item with the former description was preferred by customers.
- It demonstrate that product descriptions are vital factors for sales in Japanese e-commerce.

[1] Predicting Sales from the Language of Product Descriptions. SIGIR 2017
[2] Automatic generation of pattern-controlled product description in e-commerce. WWW 2019

Motivation

Several studies [1] [2] show that **descriptions are useful features for predicting product success.**

- We utilize descriptions as features to predict product success.
- Text → Ratings: a text regression problem

[1] Predicting Sales from the Language of Product Descriptions. SIGIR 2017
[2] Automatic generation of pattern-controlled product description in e-commerce. WWW 2019

Problem Setting

We formulate the early prediction of product success as follows:

Given a set of products $\mathcal{I} = \mathcal{I}_{\text{old}} \bigcup \mathcal{I}_{\text{new}}$

$\mathcal{I}_{\text{old}}$ denotes the set of mature products associated with user reviews

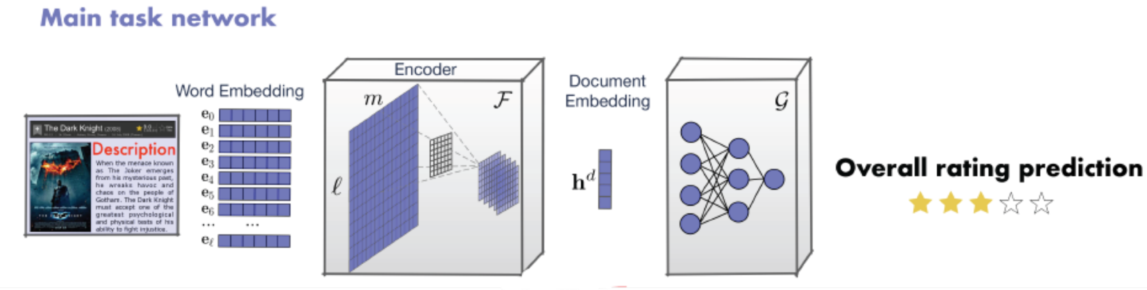
$\mathcal{I}_{\text{new}}$ denotes the set of upcoming products for which user reviews are unavailable

Task: Estimate the overall ratings for products in $\mathcal{I}_{\text{new}}$

Methodology

Proposed Method

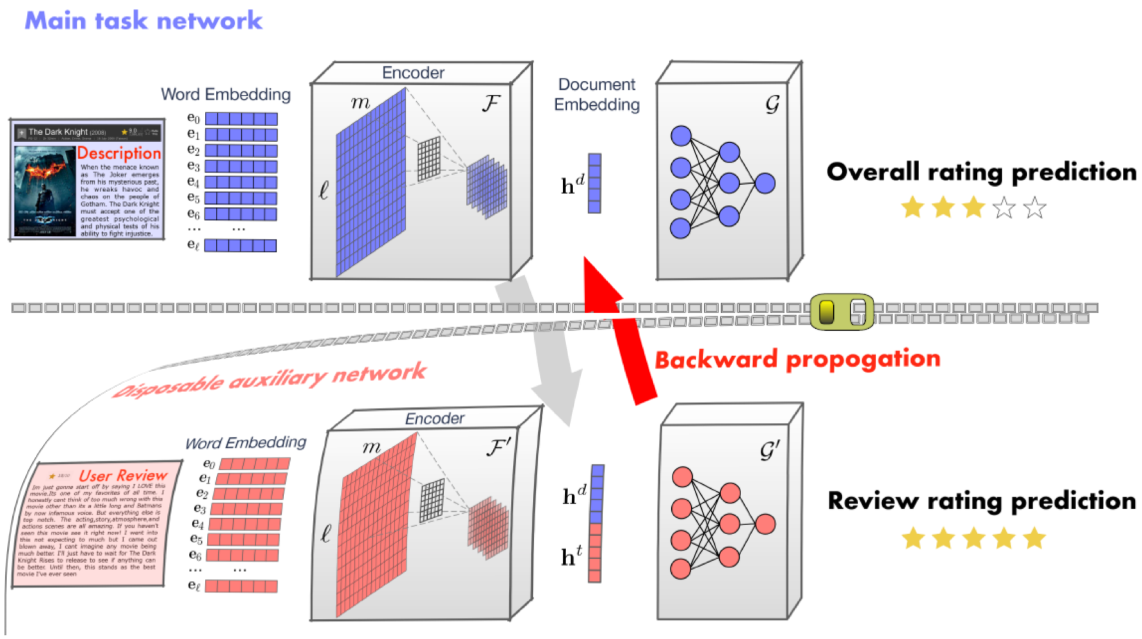
A joint learning framework that leverage the power of both descriptions and user feedbacks



Extract textual features from product descriptions

Proposed Method

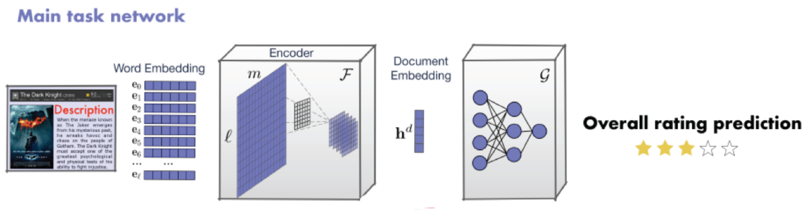
A joint learning framework that leverage the power of both descriptions and user feedbacks



Extract textual features from product descriptions

Auxiliary network popular trends from user reviews

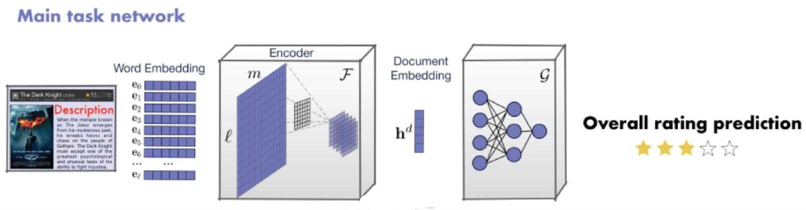
Proposed Method



Main Network

Text ➔ Word Embedding ➔ Document Embedding ➔ Rating

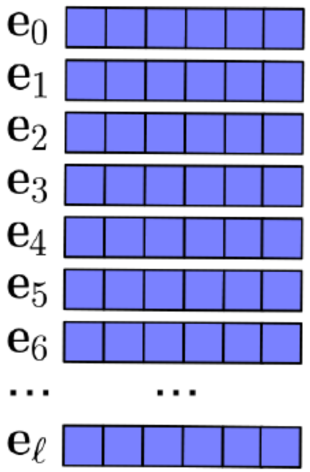
Proposed Method



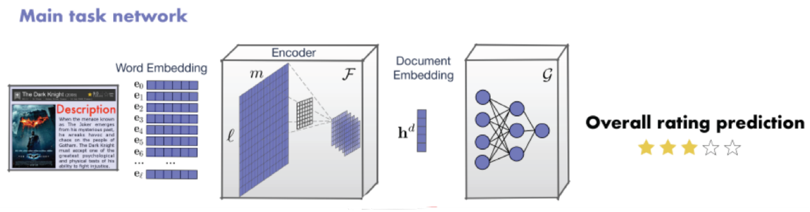
Main Network

Text ➔ Word Embedding ➔ Document Embedding ➔ Rating

Word2Vec

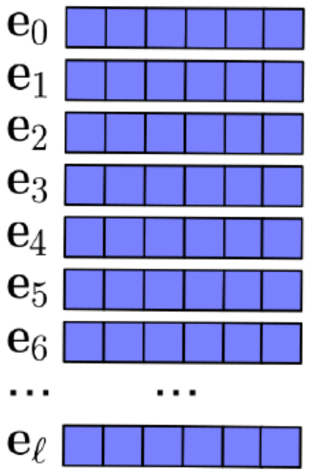


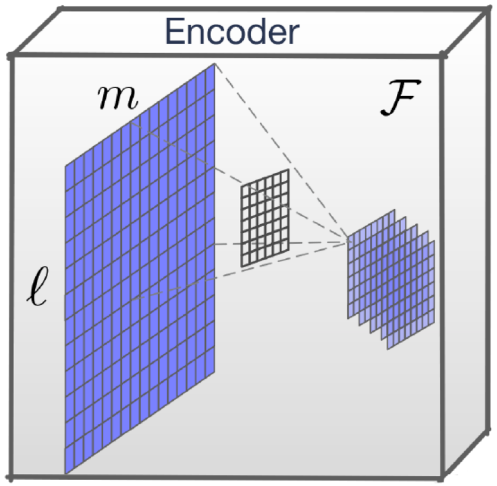
Proposed Method



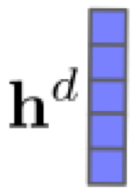
Main Network

Text ➔ Word Embedding ➔ Document Embedding ➔ Rating

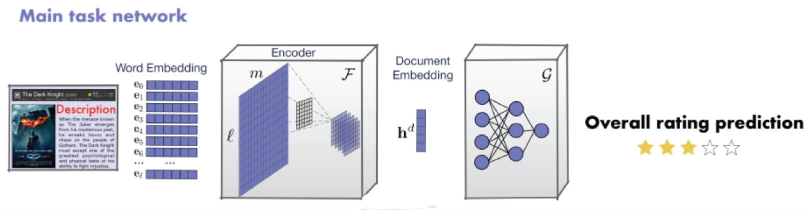




MLP
CNN
Self-Attention
BERT

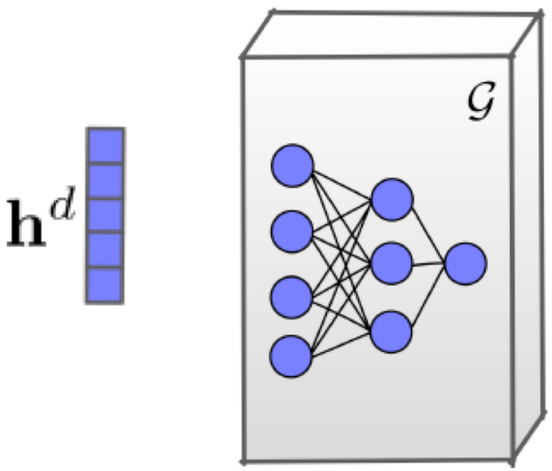


Proposed Method



Main Network

Text ➔ Word Embedding ➔ Document Embedding ➔ Rating



$$\hat{r}_i = \mathcal{G}\left(\mathcal{F}\left(\mathbf{e}_0^{d_i}, \cdots, \mathbf{e}_\ell^{d_i}\right)\right)$$

Overall rating prediction

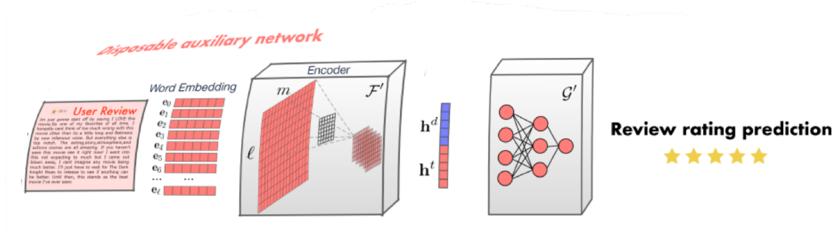
$$\mathcal{L}_{\text{main}}\left(\theta_{\mathcal{F}}, \theta_{\mathcal{G}}\right) = \sum_{i \in \mathcal{I}_{\text{old}}} \left(\hat{r}_i - r_i\right)^2$$

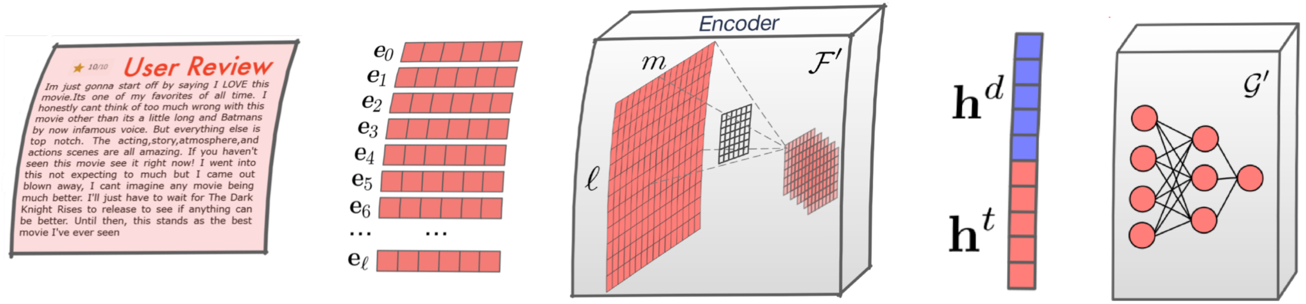
r_i

Proposed Method

Regression-based Auxiliary Network

Text → Word Embedding → Document Embedding → Rating

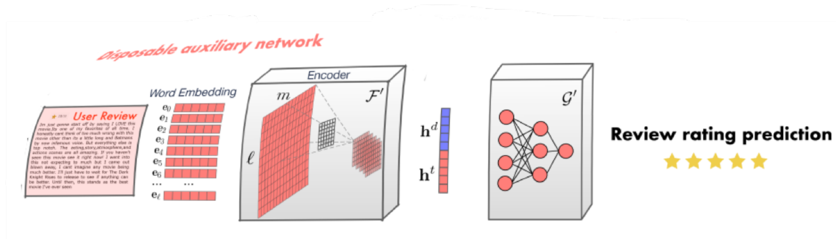




$$\hat{r}_{ij} = \mathcal{G}'_{\text{reg}}\left(\mathcal{F}'\left(e_0^{t_{ij}}, \cdots, e_\ell^{t_{ij}}\right), \mathcal{F}\left(e_0^{d_i}, \cdots, e_\ell^{d_i}\right)\right)$$

Review rating prediction

Proposed Method



Regression-based Auxiliary Network

Text ➔ Word Embedding ➔ Document Embedding ➔ Rating

$$\hat{r}_{ij} = \mathcal{G}'_{\text{reg}}\left(\mathcal{F}'\left(\mathrm{e}_0^{t_{ij}}, \cdots, \mathrm{e}_\ell^{t_{ij}}\right), \mathcal{F}\left(\mathrm{e}_0^{d_i}, \cdots, \mathrm{e}_\ell^{d_i}\right)\right)$$

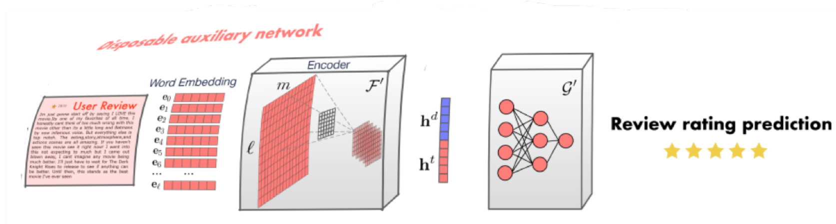
Review rating prediction ★★★★★

$$\mathcal{L}'_{\text{reg}}\left(\theta_{\mathcal{F}}, \theta_{\mathcal{F}'}, \theta_{\mathcal{G}'_{\text{reg}}}\right) = \sum_{i \in \mathcal{I}_{\text{old}}} \sum_{t_{ij} \in \mathcal{T}_i} \left(\hat{r}_{ij} - r_{ij}\right)^2$$

r_{ij}

Proposed Method

Rank-based Auxiliary Network



(movie1, movie2) ➡ **Text1 ➔ Word Embedding ➔ Document Embedding ➔ Rating1**
Text2 ➔ Word Embedding ➔ Document Embedding ➔ Rating2

$$S_{t_{ij}} = \mathcal{G}'_{\text{rank}}\left(\mathcal{F}'\left(\mathbf{e}_0^{t_{ij}}, \cdots, \mathbf{e}_\ell^{t_{ij}}\right), \mathcal{F}\left(\mathbf{e}_0^{d_i}, \cdots, \mathbf{e}_\ell^{d_i}\right)\right)$$
$$S_{t_{ik}} = \mathcal{G}'_{\text{rank}}\left(\mathcal{F}'\left(\mathbf{e}_0^{t_{ik}}, \cdots, \mathbf{e}_\ell^{t_{ik}}\right), \mathcal{F}\left(\mathbf{e}_0^{d_i}, \cdots, \mathbf{e}_\ell^{d_i}\right)\right)$$

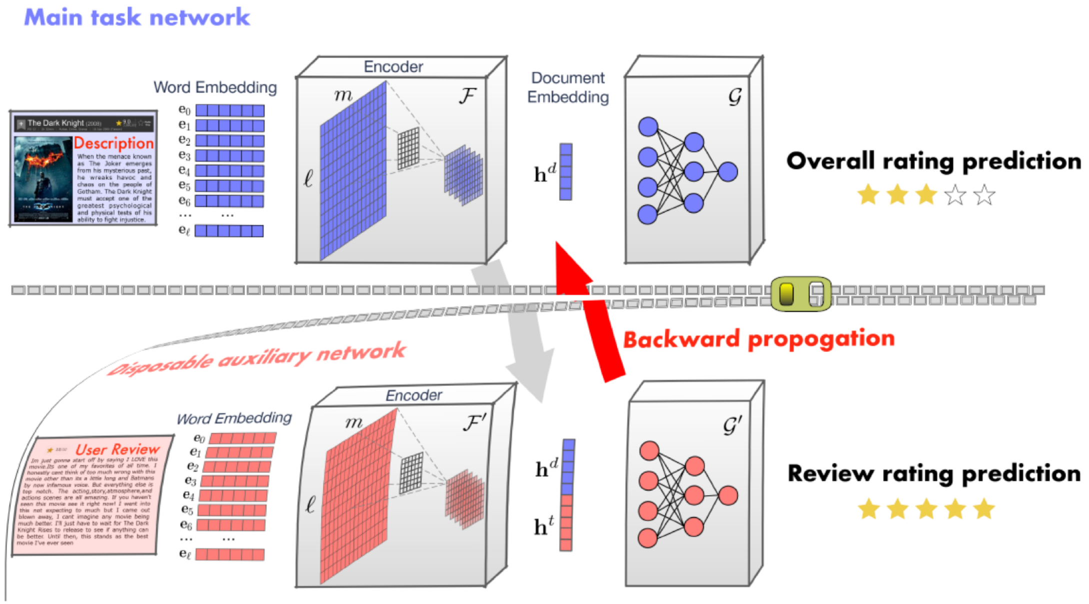
➡ $$\hat{P}(S_{t_{ij}}, S_{t_{ik}}) = \frac{1}{1 + \exp(-(S_{t_{ij}} - S_{t_{ik}}))}$$

➡ $$\mathcal{L}'_{\text{rank}}\left(\theta_{\mathcal{F}}, \theta_{\mathcal{F}'}, \theta_{\mathcal{G}'_{\text{rank}}}\right) = -\left(1 - \mathbb{1}_{\{r_{ij} > r_{ik}\}}\right) \log\left(1 - \hat{P}(S_{t_{ij}}, S_{t_{ik}})\right)$$

The rank-based network predicts the relative rank of two documents in a given document pair

Proposed Method

Full Network at training stage

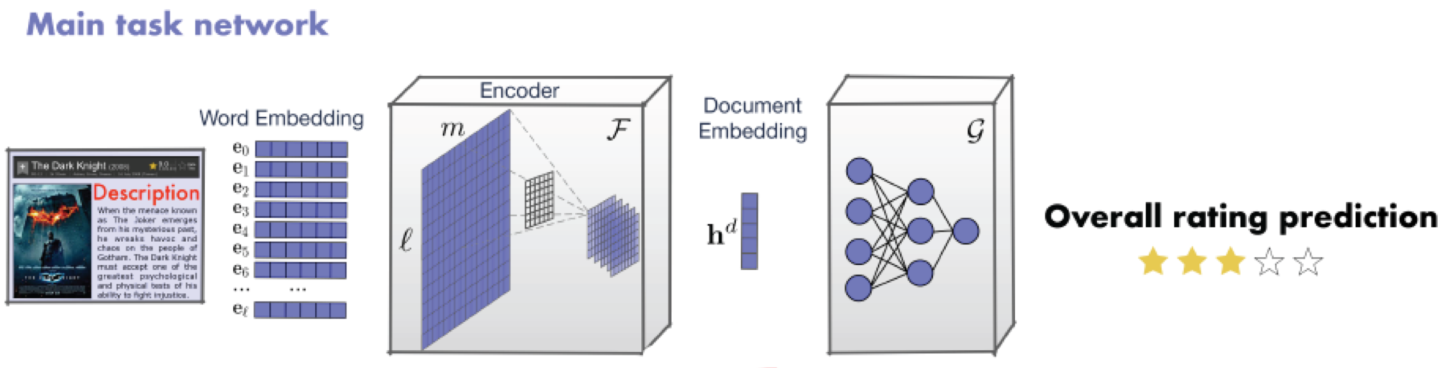


$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}}(\theta_{\mathcal{F}}, \theta_{\mathcal{G}}) + \gamma \mathcal{L}'_{\text{aux}}(\theta_{\mathcal{F}}, \theta_{\mathcal{F}'}, \theta_{\mathcal{G}'_{\text{aux}}})$$

γ [0,2] with a step size of 0.2

Proposed Method

At inference stage



$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}}(\theta_{\mathcal{F}}, \theta_{\mathcal{G}})$$

Experiments

Datasets - IMDB

Evaluate our framework on 2 real-world datasets

Dataset	IMDB	Filmarks
# movies	1,452	1,900
# user reviews	29,111	2,665,130
Rating range	0–10	0–5
Average rating	5.93	3.45
Average description length	26	75
Average review length	232	46
Average # reviews per movie	20	1,402

IMDB:

- We crawled the descriptions and the overall ratings of the movies via the OMDB API, and filtered out movies with fewer than ten reviews.
- The IMDB dataset only has at most 30 reviews.

Preprocessing

- We tokenized the movie descriptions and reviews using NLTK.
- The pre-trained word embeddings were obtained from Word2Vec.

Datasets - Filmarks

Evaluate our framework on 2 real-world datasets

Dataset	IMDB	Filmarks
# movies	1,452	1,900
# user reviews	29,111	2,665,130
Rating range	0–10	0–5
Average rating	5.93	3.45
Average description length	26	75
Average review length	232	46
Average # reviews per movie	20	1,402

Filmarks [3] :

- The biggest online review platform for films in Japan.
- We crawled the movies from 2001 to 2019 and filtered out those with fewer than 300 reviews.

Preprocessing

- We tokenized the texts using MeCab.
- The pre-trained Japanese word embeddings were obtained from Wikipedia2Vec.

[3] https://filmarks.com/

Main Results

Dataset	IMDB				Filmarks			
	ST approach	ST* approach	MT approach		ST approach	ST* approach	MT approach	
			Regression-based	Rank-based			Regression-based	Rank-based
Mean	0.764	0.764	-	-	0.354	0.354	-	-
KNN	0.718	0.760	-	-	0.336	0.315	-	-
MLP	0.743	2.067	0.680 $\pm 5.5e-03$	0.665 $\pm 1.7e-03$	0.356	2.157	0.281 $\pm 1.4e-02$	0.290 $\pm 9.6e-03$
CNN	0.738	1.424	0.474 $\pm 9.1e-03$	0.479 $\pm 8.0e-03$	0.360	0.936	0.214 $\pm 1.0e-03$	0.204 $\pm 7.0e-03$
Self-attention	0.902	2.567	0.765 $\pm 2.2e-02$	0.763 $\pm 1.1e-02$	0.373	1.628	0.353 $\pm 2.2e-03$	0.356 $\pm 8.8e-03$
BERT	0.347	1.026	0.227 $\pm 2.1e-02$	0.218 $\pm 2.7e-02$	0.233	0.458	0.168 $\pm 1.8e-02$	0.173 $\pm 7.1e-03$

- Simple Baseline Models:
  - Mean
  - KNN

Overall Score Prediction

Evaluation Metric: **RMSE**

ST/MT stands for single task/multi-task, respectively
ST∗ is a variation of the single task approach that concatenates the movie description and review texts together for training

Main Results

Dataset	IMDB				Filmarks			
	ST approach	ST* approach	MT approach		ST approach	ST* approach	MT approach	
			Regression-based	Rank-based			Regression-based	Rank-based
Mean	0.764	0.764	-	-	0.354	0.354	-	-
KNN	0.718	0.760	-	-	0.336	0.315	-	-
MLP	0.743	2.067	$0.680_{\pm 5.5e-03}$	$0.665_{\pm 1.7e-03}$	0.356	2.157	$0.281_{\pm 1.4e-02}$	$0.290_{\pm 9.6e-03}$
CNN	0.738	1.424	$0.474_{\pm 9.1e-03}$	$0.479_{\pm 8.0e-03}$	0.360	0.936	$0.214_{\pm 1.0e-03}$	$0.204_{\pm 7.0e-03}$
Self-attention	0.902	2.567	$0.765_{\pm 2.2e-02}$	$0.763_{\pm 1.1e-02}$	0.373	1.628	$0.353_{\pm 2.2e-03}$	$0.356_{\pm 8.8e-03}$
BERT	0.347	1.026	$0.227_{\pm 2.1e-02}$	$0.218_{\pm 2.7e-02}$	0.233	0.458	$0.168_{\pm 1.8e-02}$	$0.173_{\pm 7.1e-03}$

- Neural Models:
  - MLP
  - CNN
  - Self-attention
  - BERT

Overall Score Prediction

Evaluation Metric: **RMSE**

ST/MT stands for single task/multi-task, respectively

ST∗ is a variation of the single task approach that concatenates the movie description and review texts together for training

Main Results

Dataset	IMDB				Filmarks			
	ST approach	ST* approach	MT approach		ST approach	ST* approach	MT approach	
			Regression-based	Rank-based			Regression-based	Rank-based
Mean	0.764	0.764	-	-	0.354	0.354	-	-
KNN	0.718	0.760	-	-	0.336	0.315	-	-
MLP	0.743	2.067	0.680±5.5$e-03$	0.665±1.7$e-03$	0.356	2.157	0.281±1.4$e-02$	0.290±9.6$e-03$
CNN	0.738	1.424	0.474±9.1$e-03$	0.479±8.0$e-03$	0.360	0.936	0.214±1.0$e-03$	0.204±7.0$e-03$
Self-attention	0.902	2.567	0.765±2.2$e-02$	0.763±1.1$e-02$	0.373	1.628	0.353±2.2$e-03$	0.356±8.8$e-03$
BERT	0.347	1.026	0.227±2.1$e-02$	0.218±2.7$e-02$	0.233	0.458	0.168±1.8$e-02$	0.173±7.1$e-03$

- Effectiveness of auxiliary network:
  - MT approach consistently outperforms the ST approach.
  - BERT demonstrates the best performance improving ST approach by
    - 35%–37% and 26%–28% for IMDB and Filmarks

ST/MT stands for single task/multi-task, respectively
ST∗ is a variation of the single task approach that concatenates the movie description and review texts together for training

Effect of Review Size

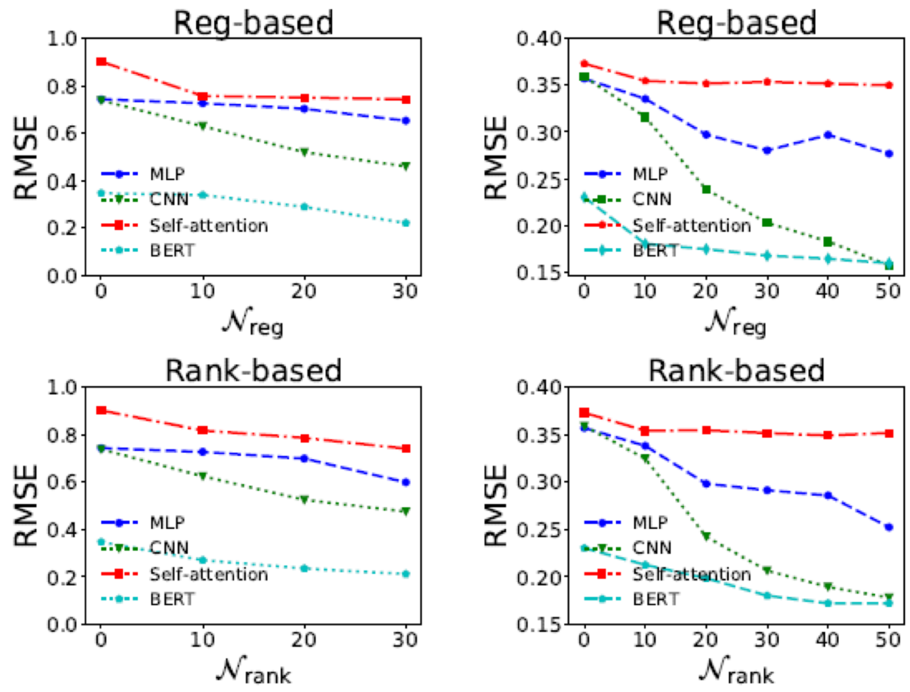
Larger user reviews pools generally improves performance

$\mathcal{N}$ <= 30

IMDB and Filmarks see a continuous improvement

$\mathcal{N}$ > 30

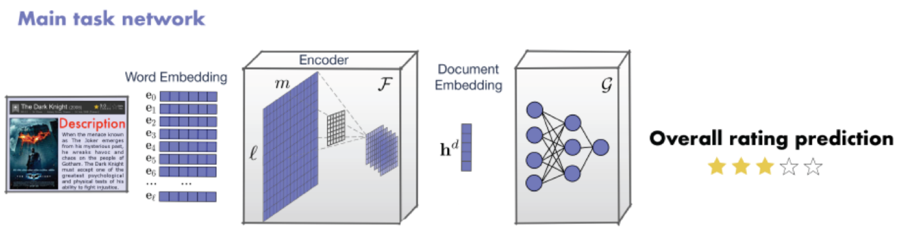
MLP with regression networks show a performance drop



(a) IMDB (b) Filmark

Effectiveness on Blockbuster Predictions

Blockbuster Predictions in testing sets



...

...

...

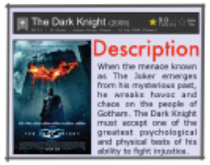
BERT as the encoder

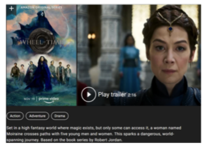
...

...

...

Predicted top X%-rated list

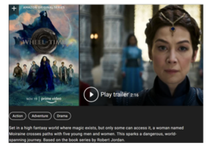


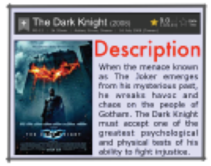


.

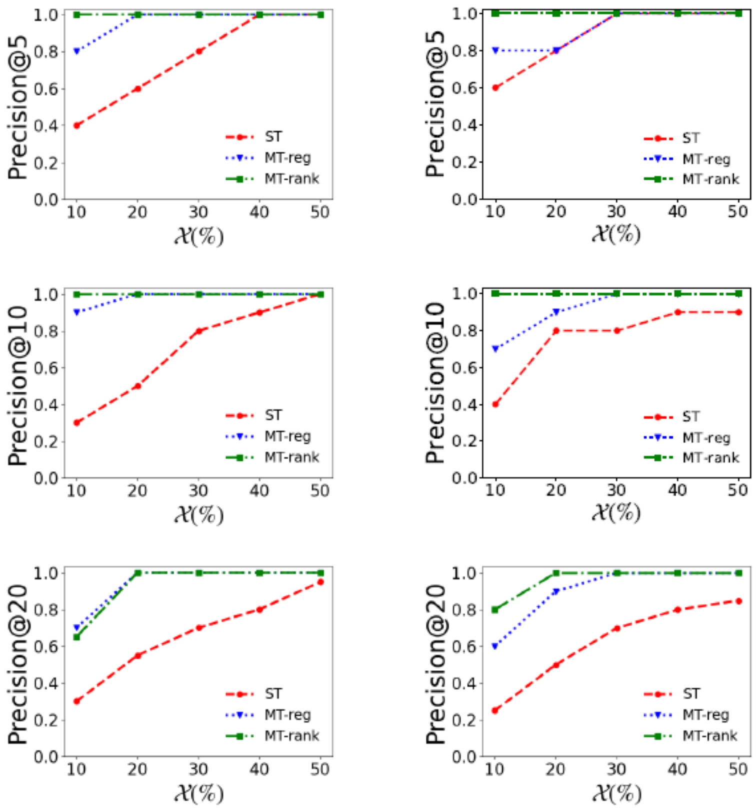
.

Top X%-rated movies





Effectiveness on Blockbuster Predictions



(a) IMDB (b) Filmark

Blockbuster Predictions in testing sets

- MT approaches consistently obtain much better results
- Reach 100% precision for top 30%- to 50%-rated movies predictions
- MT approaches identifies blockbusters at an early stage.

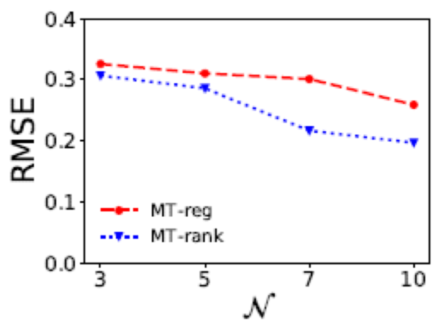
Advantage of Rank-based Auxiliary Network

The regression network vs. The rank network

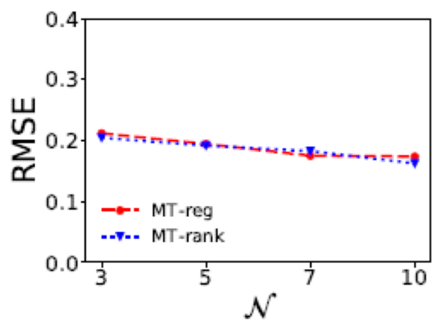
When only a limited number $(\mathcal{N})$ of user reviews are exposed:

$$\mathcal{N}_{\text{reg}} = \mathcal{N}$$

$$\mathcal{N}_{\text{rank}} = O(C_2^{\mathcal{N}})$$



(a) IMDB



(b) Filmark

- Both types show improvements in IMDB & Filmarks
- The improvement is more significant for the rank network in IMDB; but almost identical in Filmarks.
- **Rank-based auxiliary network has the potential to further exploits information from the limited user reviews.**

CASE STUDY

Movie: Gunga Din (1939)

Rating: 7.40/10	Predicted Rating: 5.92/10	Predicted Rating: 7.62/10
	In 19th century India three British soldiers and a native waterbearer must stop a secret mass revival of the murderous Thuggee cult before it can rampage across the land.	In 19th century India three British soldiers and a native waterbearer must stop a secret mass revival of the murderous Thuggee cult before it can rampage across the land.

Movie: 仮面ライダー平成ジェネレーションズ Dr.パックマン対エグゼイド＆ゴースト with レジェンドライダー

Rating: 3.70/5	Predicted Rating: 3.41/5	Predicted Rating: 3.61/5
	新た 仮面 ライダー 始まり 活躍 仮面ライダー 終わり 融合 刺激的 展開 始まり 続い 仮面ライダー×仮面ライダー ＭＯＶＩＥ 大戦 シリーズ 仮面ライダー 生誕 ４５ 周年 迎え ２０１６ 年 新た 衝撃 引っ提げ レベルアップ 果たす 名 仮面ライダー 平成 ジェネレーションズ ベース ２ 大 仮面ライダー 共闘 一気に ５ 人 仮面ライダー ドリームチーム 人類 危機 対峙 ビッグ スケール 闘い 進化	新た 仮面 ライダー 始まり 活躍 仮面ライダー 終わり 融合 刺激的 展開 始まり 続い 仮面ライダー×仮面ライダー ＭＯＶＩＥ 大戦 シリーズ 仮面ライダー 生誕 ４５ 周年 迎え ２０１６ 年 新た 衝撃 引っ提げ レベルアップ 果たす 名 仮面ライダー 平成 ジェネレーションズ ベース ２ 大 仮面ライダー 共闘 一気に ５ 人 仮面ライダー ドリームチーム 人類 危機 対峙 ビッグ スケール 闘い 進化

(a) Single-task approach (b) Multi-task approach

Conclusions

Take Home Messages

(1) We propose a learning framework to address a vital task for business—early prediction of product success when limited information is available during inference.

(2) The framework effectively combines a main task network and a disposable auxiliary network, the latter of which can be either a regression or a ranking model.

(3) The proposed framework yields an over-20% performance improvement on two real-world datasets in different languages.

Thank You!

Are there any questions you'd like to ask?

Chih-Ting Yeh justinyeh1995@citi.sinica.edu.tw

Ming-Feng Tsai mftsai@nccu.edu.tw

Chuan-Ju Wang cjwang@citi.sinica.edu.tw