Learning to Learn Sales Prediction with Social Media Sentiment

Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, Zihan Liu, Yan Xu, Cong Gao and Pascale Fung

FinNLP, IJCAI 2019

AI WITH HEART

CAiRE
Centre for Artificial Intelligence Research

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Sales prediction is a key element for business planning

- Demand Forecasting
- Supply Chain Management
- Finance Planning

Sales Prediction is crucial

Sentiment is an useful indicator to products' sales

1. Traditional time-series models of predicting future sales of a product or service only relied on the **past historical and seasonal sales**, and **often provided unreliable prediction outcomes** (Ahn et. al., 2014).
2. Sentiment about companies and products affect **buying behaviors** and **future sales** by words-of-mouth effects (Ahn et. al., 2014).

Existing Work

1. **Movie Sales** (Duan et al., 2008; Gaikar and Marakarkandy, 2015; Ahn and Spangler, 2014; Marshall et al., 2013; Asur and Huberman, 2010)

2. **E-commerce Products** (Davis and Khazanchi, 2008; Tuarob and Tucker, 2013)

3. **Car Sales** (Wijnhoven and Plant, 2017; Geva et al., 2017; Barreira et al., 2013)

Social Media Sentiment for Smartphone Sales Prediction

Tasks

1. Collecting user comments of different smartphone brands from social media (Weibo).
2. Training a sentiment classifier to produce reliable sentiment features.
3. Building a sales prediction model which takes sentiments into account and make prediction.

Collecting Weibo Data

We crawl around **5 million Weibo comments** from seven different smartphones official accounts from 2013 to 2018.

1

Crawl historical Weibo data over a period of last 5 years

2

Clean up the data to remove unwanted characters and hyperlinks.

3

Brands	# Comments
Samsung	288,081
Gionee	302,866
Huawei	524,468
Oppo	633,406
Vivo	670,408
Meizu	945,312
iPhone	994,155

Final smartphone comments

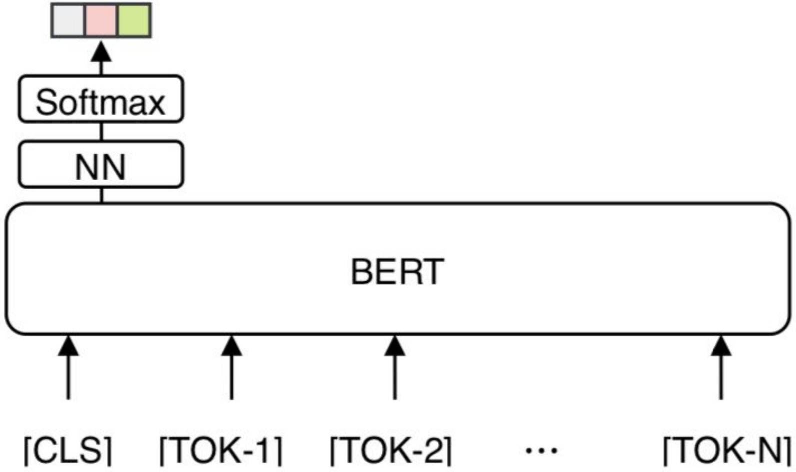
Crowd-sourcing

1. Randomly sample 25,000 Weibo comments
2. Annotate them with Positive, Negative, and Neutral labels.

The annotation result shows that the percentage of Positive, Negative, and Neutral labels are 20%, 16%, and 64% respectively.

Sentiment Analysis

1. From 25,000 comments, we take out 5000 comments as our test set.
2. We finetune pre-trained language model BERT (Devlin et.al., 2018) on the weibo sentiment data and achieve around 80% accuracy.



Sentiment Feature

To incorporate sentiment information into the sales predictor, we quantify the sentiment score of each brand in the quarter. We calculate the score x_t by the following:

$$x_t = \frac{p_t}{p_t + n_t}$$

where p_t is the number of comments with positive sentiment in the quarter t, and n_t is the number of comments with negative sentiment in the quarter t,. The score is normalized to 0-1 range

Smartphone Sales

We collect quarterly China sales data of seven smartphones: Samsung, Gionee, Huawei, Oppo, Vivo, Meizu, and iPhone from the first quarter of 2013 to the third quarter of 2018 released by IDC. In each brand, we reserve the last five quarters for testing, and we use the rest for training our models.

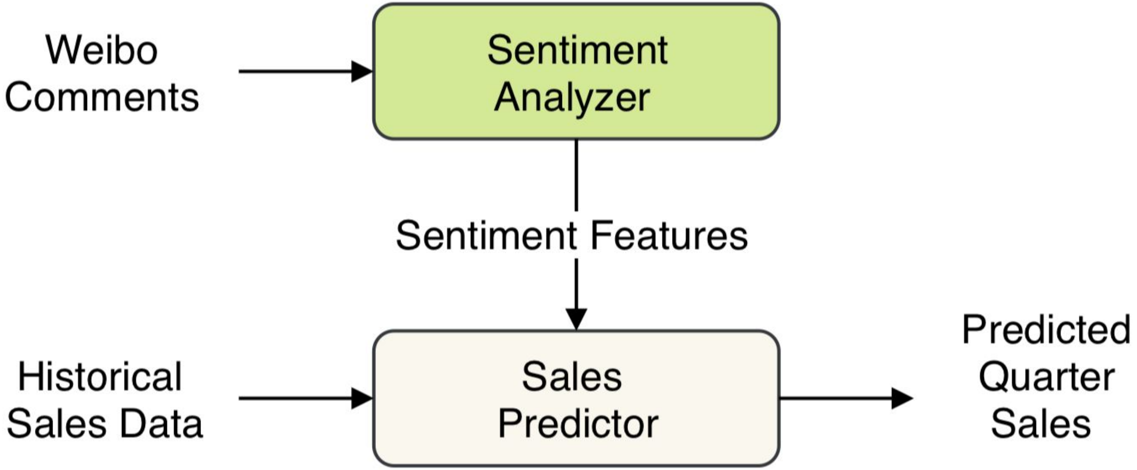
Number of Brands	Training datapoint for each brand	test datapoint for each brand
7	18	5

Sales Prediction

1. **DEFINE: Quarterly** sales $S = [s_0,...,s_t,...,s_N]$
2. **DEFINE: Sentiment features** at each quarter with one month gap $X = [x_0,...,x_t,...,x_N]$

The task of our model is to predict sales $\mathbf{s_t}$ by taking in input the sales history $S_{0:t-1} = [s_0 , . . . , s_{t-1}]$ and sentiment value x_t.

Overall Architecture



The architecture consists of sentiment analyzer that extract sentiment features from Weibo comments and sales prediction model that take sentiment features and historical sales as input and make the prediction.

Small Training Set for Each Brand!

Problem: With such small training data, it is hard for model to learn to condition on a complex historical sales pattern and sentiment feature, and make a accurate prediction.

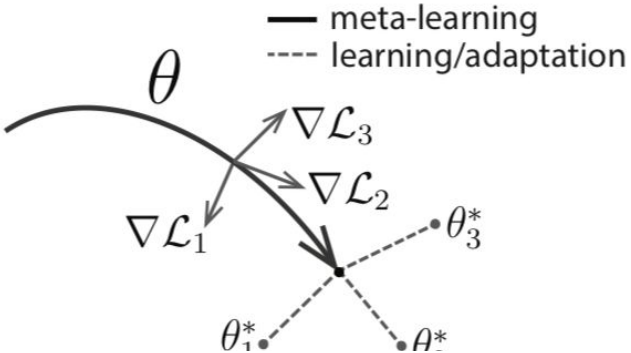
Solution: Apply meta-learning to leverage sales data from other brands as a prior knowledge. Learn a model which can make an accurate prediction on a new product after training on few historical sales samples.

Detail: We use Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) to simulate learning process that minimizes the prediction error by utilizing the historical training samples.

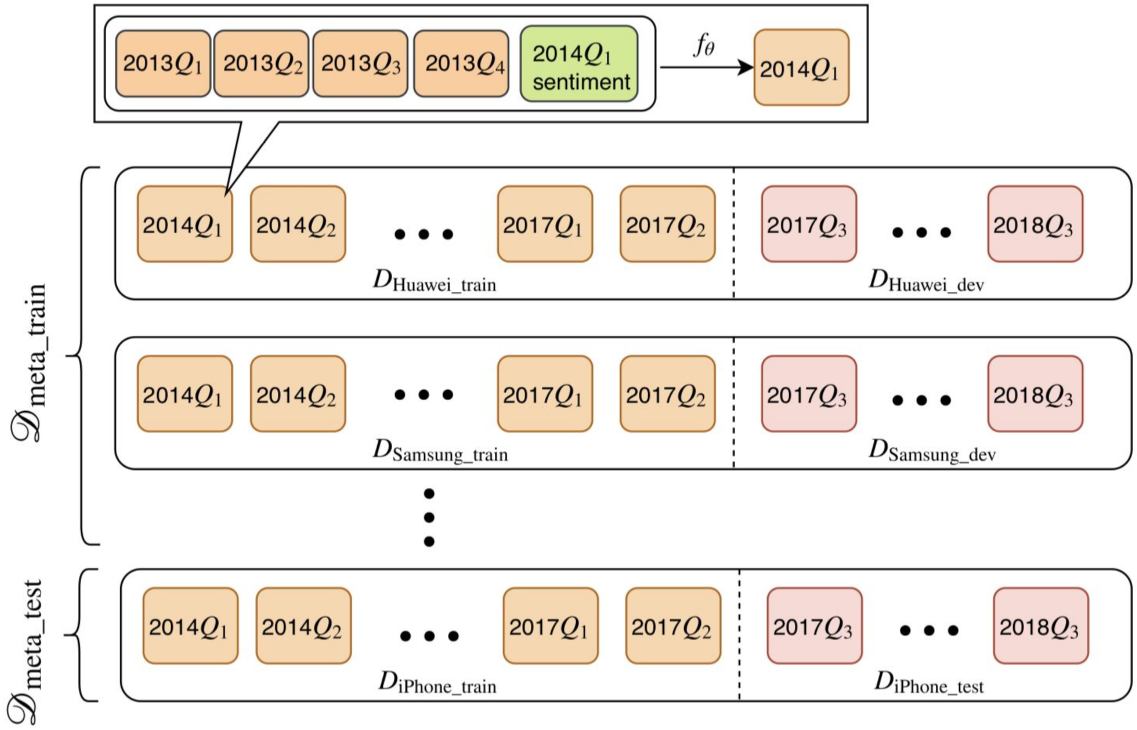
Brief about MAML (Finn et al., 2017)

The model or learner is trained during a meta-learning phase on a set of tasks, such that the trained model can quickly adapt to new tasks using only a small number of examples

The goal is to train the model's initial parameters such that the model has maximal performance on a new task after the parameters have been updated through one or more gradient steps computed with a small amount of data from that new task.



Learning Schema



MAML for Sales Prediction

Algorithm 1 MAML for sales prediction task

Require: $\mathscr{D}_{meta-train}$
Require: α, β learning rate

Randomly initialize θ
while not done **do**
    Sample batch of products $\mathcal{D}_i \sim \mathscr{D}_{meta-train}$
    for all $\mathcal{D}_i$ **do**
        $(\mathcal{D}_{i\_train}, \mathcal{D}_{i\_dev}) \leftarrow \mathcal{D}_i$
        Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{D}_i}(f_\theta)$ using $\mathcal{D}_{i\_train}$ and $\mathcal{L}_{\mathcal{D}_i}$ in Equation (5)
        Compute adapted parameters with gradient descent:$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{D}_i}(f_\theta)$
    end for
    Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{D}_i \sim \mathscr{D}} \mathcal{L}_{\mathcal{D}_i}\left(f_{\theta_i'}\right)$ using $\mathcal{D}_{i\_dev}$ and $\mathcal{L}_{\mathcal{D}_i}$ in Equation (5)
end while

Mean Squared Error loss (MSE)

$$\mathcal{L}_{\mathcal{D}_i}(f_\theta) = \sum_{\mathbf{x}^{(j)}, \mathbf{y}^{(j)} \sim \mathcal{D}_i} \| f_\theta(\mathbf{x}^{(j)} - \mathbf{y}^{(j)}) \|_2^2$$

Models

We compare four different models:

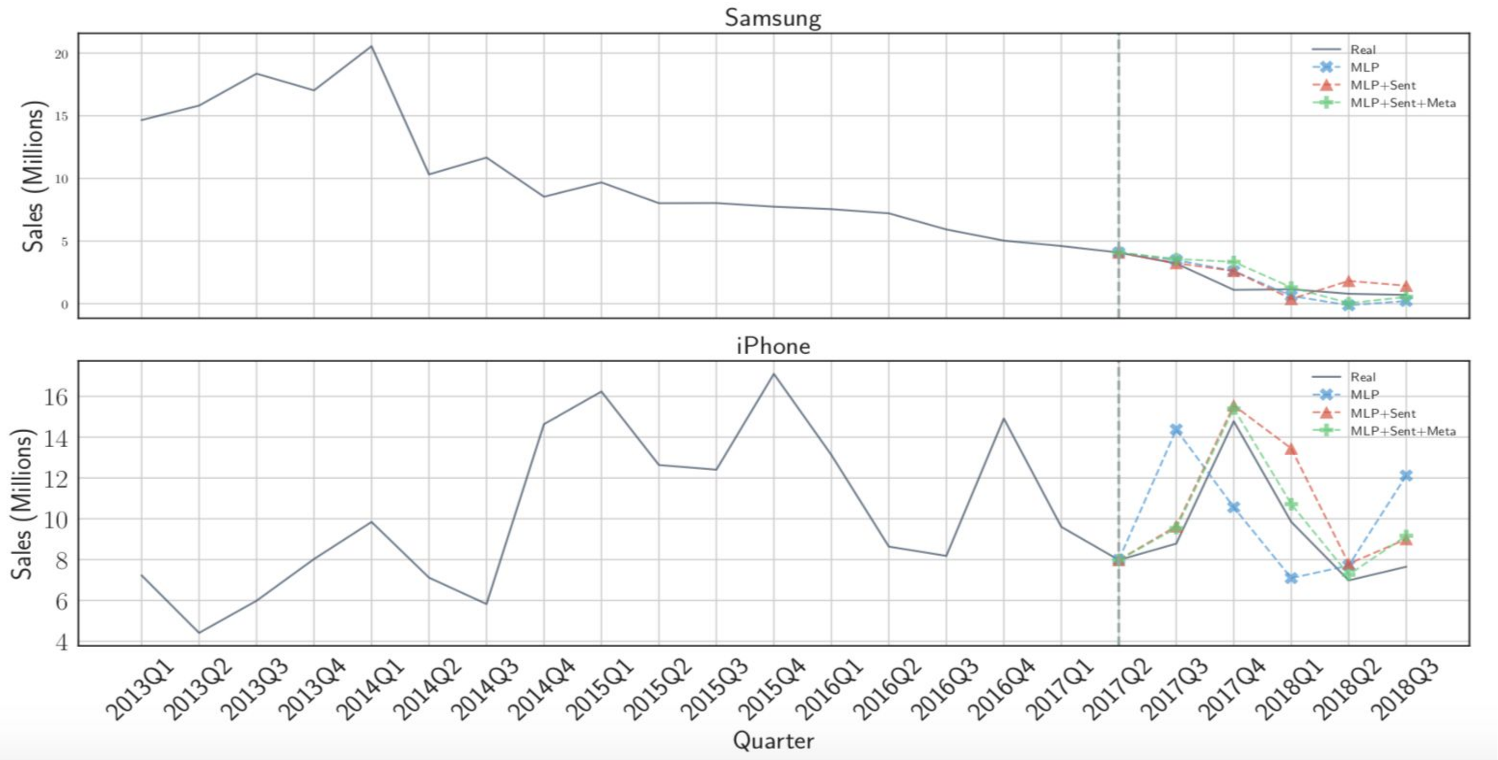
1. Linear regression
2. Support Vector Regression
3. Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX)
4. Multilayer Perceptron (MLP)

Results

MSE	iPhone	Gionee	Huawei	Meizu	Oppo	Samsung	Vivo	Average
Linear	10.956	23.733	4.139	5.745	12.639	2.702	7.779	9.670
Linear+Sentiment	5.501	4.082	7.420	6.749	13.215	2.251	7.394	6.659
SVR	7.733	7.533	4.37	6.044	4.764	9.203	8.672	6.903
SVR+Sentiment	4.106	4.444	4.714	11.836	6.869	4.532	9.107	6.515
SARIMAX	0.588	10.241	6.331	**2.783**	8.875	0.876	11.552	5.892
SARIMAX+Sentiment	**0.072**	8.232	6.667	5.742	**2.114**	1.073	10.869	4.967
MLP	15.429	8.565	3.684	6.55	11.03	**0.737**	9.931	7.990
MLP+Sentiment	3.625	3.128	**3.187**	6.199	2.782	0.891	16.648	5.209
MLP+Sentiment+Meta	0.822	**2.765**	4.906	9.114	3.525	1.145	**7.134**	**4.202**

Sentiment feature reduces the average sales prediction error. **Meta-learning** further improve the prediction performance. However, in some cases (i.e. Meizu) sentiment feature also **hurts** the performance. It might because the the **noisy** weibo data and other important factor which we haven't take into account (i.e. Price).

Results



Samsung
Sales (Millions)
20
15
10
5
0
Real
MLP
MLP+Sent
MLP+Sent+Meta
iPhone
Sales (Millions)
16
14
12
10
8
6
4
Real
MLP
MLP+Sent
MLP+Sent+Meta
2013Q1
2013Q2
2013Q3
2013Q4
2014Q1
2014Q2
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
2016Q2
2016Q3
2016Q4
2017Q1
2017Q2
2017Q3
2017Q4
2018Q1
2018Q2
2018Q3
Quarter

Conclusion

1. The results of our experiments show that sentiment information improves the average performance of sales prediction which confirms the effectiveness of the sentiment index.

2. The meta-learning method help models transfer the knowledge of sentiment-sales correlation from different products, further reduce the sales prediction error.

Thank you