Learning to Learn Sales Prediction with Social Media Sentiment

Zhaojiang Lin[1] , **Andrea Madotto**[1] , **Genta Indra Winata**[1] , **Zihan Liu**[1] , **Yan Xu**[1] ,
Cong Gao[1] and **Pascale Fung**[1,2]
[1]Center for Artificial Intelligence Research (CAiRE)
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
[2]EMOS Technologies Inc
{zlinao,amadotto,giwinata,zliucr,yxucb,eemiranda}@connect.ust.hk,
pascale@ece.ust.hk

Abstract

Social media sentiment has shown to be a useful resource for product sales forecast. However, research on modeling the correlation between sentiment index and sales is often limited by the scarceness of quarterly sales data. In this paper, we propose to learn how to learn sentiment-sales correlation from different source products and transfer to sales prediction of another, target product. We evaluated our approach on sales data of seven different smartphones and showed that the knowledge transfer from six source products significantly reduced the sales prediction error for the target product, in a 7-fold cross-validation experiment.

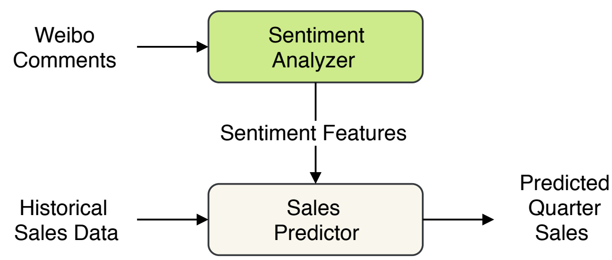

Figure 1: Overall architecture. In a big picture, it consists of sentiment analyzer to extract sentiment features from Weibo comments and they are fed into the sales prediction model along with historical sales data.

1 Introduction

The sales forecast is crucial in the financial domain since it indicates the future trend of a product and thus, it allows investors to make better decisions. During the years, time-series models have been widely applied in sales prediction using historical seasonal sales data. However, they are often unreliable since historical sales ignore the importance of customers opinions (e.g., Social Media, News), which are critical for sales prediction [Ahn and Spangler, 2014]. On the other hand, user-generated content in social media acts as word of mouth contains a large number of customer opinions. Sentiment analysis of social media provides a good summary of customers' feedback and allow companies to have a better intuition of how the market reacts to their products.

Several existing work use sentiment features to predict product sales, for instance for movie sales [Duan *et al.*, 2008; Gaikar and Marakarkandy, 2015; Ahn and Spangler, 2014; Marshall *et al.*, 2013; Asur and Huberman, 2010], e-commerce products [Davis and Khazanchi, 2008; Tuarob and Tucker, 2013] and car sales [Wijnhoven and Plant, 2017; Geva *et al.*, 2017; Barreira *et al.*, 2013]. These works show positive correlations between sentiments features and sales, and thus, the sentiment is a useful indicator to predict the outcome of future sales. However, most of them focus on correlation studies between features, but they have not explored the possibility to transfer information from different brands.

Moreover, the preferred models for the sales prediction task are usually linear (e.g., BIC, ARIMA) due to the particularly small datasets. Indeed, a well-known problem for deep learning models, and in general non-linear models, is that they require a large amount of data to work properly.

Differently from the previous work, in this paper, we study the sales of seven smartphones in China's market such as Samsung, Gionee, Huawei, Oppo, Vivo, Meizu, and iPhone. We show the importance of sentiment features by incorporating sentiment information – extracted from the biggest Chinese social media platform Weibo – for improving sales prediction. To extract reliable sentiment index from Weibo, we build an accurate sentiment analyzer by applying the state-of-the-art pre-trained model BERT: Bidirectional Encoder Representation from Transformer [Devlin *et al.*, 2018]. Moreover, we report the sales prediction results of several statistical models and show the usefulness of sentiment features. Most importantly, we propose a viable way to alleviate the scarceness of sales data by using meta-learning. This technique allows a non-parametric model such as neural networks to leverage historical sales of other brands, and use them as the prior knowledge. The intuition of applying meta-learning is that it optimizes the model for fast adaptability, allowing it to adapt to new prediction tasks.

The main contributions of this paper are: 1) Collecting and pre-processing a large scale dataset of user comments of seven different smartphone companies from a popular Chinese social media platform Weibo, and providing human la-

Proceedings of the First Workshop on Financial Technology and Natural Language Processing
(FinNLP@IJCAI 2019), pages 47-53, Macao, China, August 12, 2019.

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

Table 1: The number of Weibo comments for smartphone brands.

beled sentiment annotations of 25K comments; 2) Training a state-of-the-art sentiment classifier to produce reliable sentiment features for the sales prediction; 3) Reporting consistent improvements in the sales prediction by using the extracted sentiments features, which confirms existing related previous works; 4) Proposing a deep learning-based solution that can compete, and improve, with very strong statistical models. By using meta-learning, our model is able to leverage other brands sales history for making a more accurate sales prediction. To the best of our knowledge, we are the first to report positive results in this setting.

In the following sections, we introduce 1) Corpus collection and annotation, and the historical sales dataset used in our experiments; 2) Sentiment analyzer and sales prediction models; 3) Experiments and results; 4) Related work; and 5) Conclusion.

2 Dataset Collection

2.1 Weibo Sentiment Dataset

We crawl around 5 million Weibo comments for seven different smartphones: Samsung, Gionee, Huawei, Oppo, Vivo, Meizu, and iPhone from their company official accounts from 2013 to 2018. In the data cleaning process, we remove all emojis, user mentions such as "@user", hashtags, and hyperlinks using regular expressions. Then, we group them by quarter, a period of four months. The statistics of the dataset for each brand is showed in Table 1.

We randomly sample 25,000 Weibo comments and manually annotated them with *Positive*, *Negative*, and *Neutral* labels via crowd-sourcing. The agreement is taken by majority vote. The annotation result shows that the percentage of *Positive*, *Negative*, and *Neutral* labels are 20%, 16%, and 64% respectively. We further take around 5,000 comments as our test set.

2.2 Smartphone Sales Dataset

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 [1]. In each brand, we reserve the last five quarters for testing, and we use the rest for training our models.

[1]https://www.idc.com/

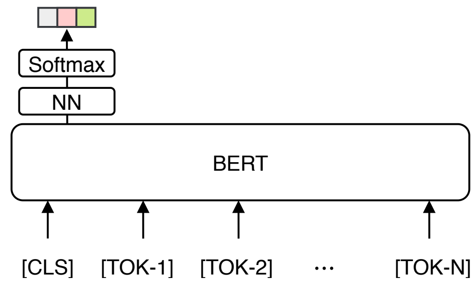

Figure 2: Sentiment Analyzer. The model accepts user comment tokens and generate a probability distribution over three classes. In the figure, green states for positive, red for negative, and gray for neutral.

3 Methodology

3.1 Sentiment Analysis

Building a reliable sentiment classifier is crucial to the final sales prediction. To alleviate the dependence of the human effort and build a robust sentiment classifier, we apply current state of the art pre-trained language model BERT: Bidirectional Encoder Representations from Transformers [Devlin *et al.*, 2018] to our task. It is a multi-layer bidirectional Transformer encoder pre-trained by using "masked language model" objective. In our proposed model, we adapt $BERT_{BASE}$ [2] [Devlin *et al.*, 2018] to generate the semantic representation of each comment to improve the sentiment prediction task. Following the same fine-tuning procedure of [Devlin *et al.*, 2018], a special token $[CLS]$ is added at the beginning of every input to obtain the fixed-dimensional representation of input sequence. As shown in Figure 1, we stack another linear layer with Softmax function on top of $BERT$ to compute the probabilities of three sentiment classes. We keep all parameters trainable, and they are fine-tuned with the sentiment training data. Our sentiment analyzer achieves around 80% accuracy in the final test set.

Sentiment Features 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 [Lassen *et al.*, 2014]:

$$x_t = \frac{p_t}{p_t + n_t} \tag{1}$$

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.

3.2 Sales Prediction

Let us define a vector $S = [s_0, \ldots, s_t, \ldots, s_N]$ as the sales at each quarter and vector $X = [x_0, \ldots, x_t, \ldots, x_N]$ as sentiment features at each quarter, where N is the total number of quarters, s_t is the sales value at quarter t, and x_t is the sentiment of comments posted in one month time before in each quarter. For example, in the second quarter of a year from

[2]We used a PyTorch implementation from https://github.com/huggingface/pytorch-pretrained-BERT

April to June, we use sentiment of the comments posted from March to May. The task of our model is to predict sales s_t by taking in input the sales history $S_{0:t-1} = [s_0, \ldots, s_{t-1}]$ and current sentiment value x_t. In this section, we introduce two different approaches: (1) a statistical-based model, Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (*SARIMAX*) (2) a gradient-based model, Multi-layer Perceptron (*MLP*). We also describe meta-learning procedure in our sales prediction task.

SARIMAX

The model is an extension of *SARIMA* model with external variables. We denote the model by $SARIMAX(p,d,q)(P,D,Q)_S(X)$, where p,d,q are orders of autoregressive, difference, and moving average and P,D,Q are orders of seasonal autoregressive, difference, and moving average. X is the external variable and S is the seasonal period (e.g., quarter). The quarterly sales series $S_{0:t}$ is computed given sentiment features x_t as follows:

$$S_{0:t} = \frac{\theta_q(B)\Theta_Q\left(B^S\right)}{\phi_p(B)\Phi_P\left(B^S\right)(1-B)^d\left(1-B^S\right)^D}\varepsilon_t + y_t, \quad (2)$$

$$y_t = w_0 + w_1 x_t \quad (3)$$

where w_0 and w_1 are regression coefficients, B and B^S are delay operators, $\phi_p(B)$ is a non-seasonal autoregressive operator with p-order, $\Phi_P\left(B^S\right)$ is a seasonal autoregressive operator with P-order, $\theta_q(B)$ is a non-seasonal moving average operator with q-order, $\Theta_Q\left(B^S\right)$ is a seasonal moving average operator with Q-order, and ε_t is a residual error.

MLP

MLP consists of multiple linear layers followed by a nonlinear activation function. Unlike autoregressive model *SARIMAX*, *MLP* requires a fixed-dimensional feature input. Therefore we take the sentiment feature alone with last four quarters historical sales number as our input feature:

$$s_t = f(S_{t-5:t-1}, x_t; \theta) \quad (4)$$

where f is *MLP* model parameterized by θ.

Meta-Learning

In this work, we apply Model-Agnostic Meta-Learning (MAML) [Finn *et al.*, 2017] to sales prediction task. The goal of MAML in our task is to find initial parameters θ_0 of sales predictor model f_θ (*MLP* in our case) such that the model can make an accurate prediction on a new product after training on few historical sales samples.

In our meta-learning scenario, every product is considered as a different task. As we showed in the Figure 3, datasets $\mathcal{D}_i$ are constructed separately for each task. We take one product out as meta-test set $\mathscr{D}_{meta-train}$, other datasets as meta-training set $\mathscr{D}_{meta-test}$. In meta-training setting of [Finn *et al.*, 2017], for each dataset $\mathcal{D}_i$, they random sample some data points $\mathcal{D}_{i\_train}$ for inner training and sample some other data points $\mathcal{D}_{i\_val}$ for meta-update. Instead, we always fix the split $\mathcal{D}_{i\_train}$ and $\mathcal{D}_{i\_val}$, because we are only interested in forecasting sales given historical sales. During the meta training,

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

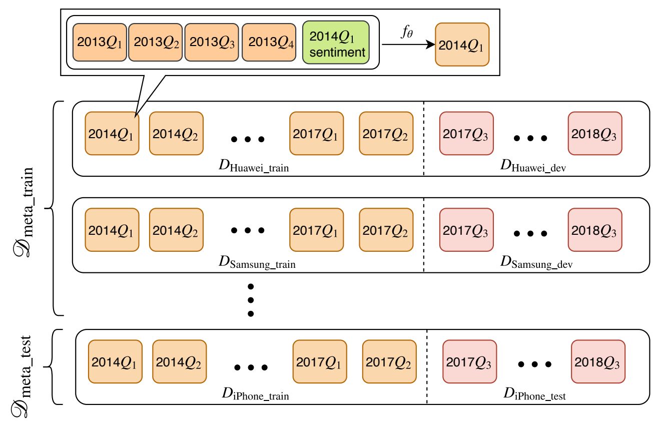

Figure 3: Example of meta-learning for sales prediction. The goal is to predict iPhone sales in next quarters. Meta-learning uses a series of historical data and sentiment from other smartphone brands to initialize the predictor model.

the model keeps simulating learning process that minimizes the prediction error by utilizing the historical training samples. The prediction error is measured by MSE (Mean Square Error) defined by equation (5). We describe the learning procedure in Algorithm 1. After meta-learning, we train our model on historical sales data $\mathcal{D}_{i\_train}$ from meta-training set $\mathscr{D}_{meta-test}$, and finally evaluate our model on $\mathcal{D}_{i\_test}$ from $\mathscr{D}_{meta-test}$.

$$\mathcal{L}_{\mathcal{D}_i}(f_\theta) = \sum_{\mathbf{x}^{(j)}, \mathbf{y}^{(j)} \sim \mathcal{D}_i} \|f_\theta(\mathbf{x}^{(j)} - \mathbf{y}^{(j)})\|_2^2 \quad (5)$$

4 Experimental and Results

4.1 Settings

In our experiments, we compare the sales prediction performance of our models with and without using sentiment information, with and without using the meta-learning method in our smartphone sales dataset. We also compare our model with two baselines: linear regression and SVR (Support Vector Regression). As mentioned in the dataset section we use

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**

Table 2: Results in Mean Squared Error (MSE).

Product	p	d	q	P	D	Q	S
iPhone	0	1	0	1	1	0	4
Gionee	0	1	0	1	0	0	4
Huawei	0	1	0	1	1	0	4
Meizu	0	1	0	1	1	0	4
Oppo	0	1	0	1	0	0	4
Samsung	0	1	0	1	0	0	4
Vivo	0	1	0	1	0	0	4

Table 3: *SARIMAX* hyper-parameters.

the last five quarters for testing and the previous for training. Hence, the model's performance to predict the next quarter sales is evaluated using Mean Squared Error (MSE) of the test set.

Hyper-parameters *SARIMAX* model is identified by following hyper-parameters: order of difference (d), the order of seasonal difference (D), non-seasonal autoregressive order (p), seasonal autoregressive order (P), non-seasonal moving average order (q), seasonal moving average order (Q). All of them are identified by Autocorrelations function (ACF) and partial autocorrelations function (PACF) as we showed in Table 3. For our gradient base model, we use two layer *MLP* with hidden size 5 and Rectified Linear Units (ReLU) as the activation function. For meta-learning, we use SGD optimizer with learning rate 0.01 for both inner and outer optimization. we run 9 iterations for each inner update, and 10 epochs of meta update.

4.2 Results

Table 2 shows the results for each model and each brand in the term of Mean Squared Error (MSE). Two results stand out: Sentiments Features consistently improves the MSE for all the models, and the *MLP* with sentiment features trained using Meta-Learning can improve the average MSE among different brands.

Sentiment Features The features help in all the evaluated models, this confirms the usefulness of such a feature in sales prediction. Indeed, this shows that Weibo comments hold essential information that can be used to predict future sales. However, from Table 2 we can notice that the only case where sentiment features hurt the performance is on Meizu data. One possible reason could be the price of Meizu is much lower than other brands; hence the sentiment might not affect the sales of low price products that target a different market.

SARIMAX vs MLP Moreover, in Table 2 we can see that both *SARIMAX* and *MLP* using sentiment features have a very similar average MSE and they performs consistently better than *SVR* and Linear Regression. Especially, *MLP* works the best for Huawei and Samsung where instead for iPhone, Oppo and Meizu *SARIMAX* works the best.

Meta-learning The best MSE average is achieved by the meta-learned model, *MLP+Sentiment+Meta* in Table 2. This is due to the ability to transfer knowledge between different brands. Indeed, meta-learning is trained to find a set of parameters that are able to quickly adapt to a given task. In our instance, this means to learn a set of parameters that can quickly adapt to the sales behavior of a certain brand.

Moreover, in Figure 4 we plot the Gionee, Vivo, Samsung and iPhone sales traces and the prediction made by *MLP* by using with and without sentiment feature including meta-learning to describe our findings. For Vivo, Samsung and iPhone, we can note that by just using *MLP* the sales predictions are not aligned with the real sales. Instead by adding sentiment features we can achieve a very good fit in the two quarters, but a more substantial error when a trend inversion appears (i.e., 2017Q4 in iPhone). This is mostly solved by meta-learning training, in which the model achieves almost a perfect fit (0.822 MSE).

We can also notice that in some brands predictions are easier than the others. For instance, the iPhone has seasonal patterns where there are peaks between the third and fourth quarters in the last two years. In this case, our autoregressive model *SARIMAX* can capture this pattern better than *MLP* with meta-learning as we showed in Table 2 . On the other hand, *SARIMAX* predicts very poorly on Gionee and Vivo which have less repeating sales patterns. Conversely to our meta-learning based model is more robust as it can accurately predict in sales trends with irregular changes.

5 Related work

5.1 Sales prediction with sentiment analysis

Sentiment and emotional analysis are important methods to quantify customers' emotional engagement [Winata *et al.*, 2019]. The importance and effectiveness of using social media opinion, a.k.a. Word-of-Mouth, for Sales Prediction is a well known topic [Hennig-Thurau *et al.*, 2003; Hennig-Thurau *et al.*, 2004; Ceron and d'Adda, 2016; Liu, 2012;

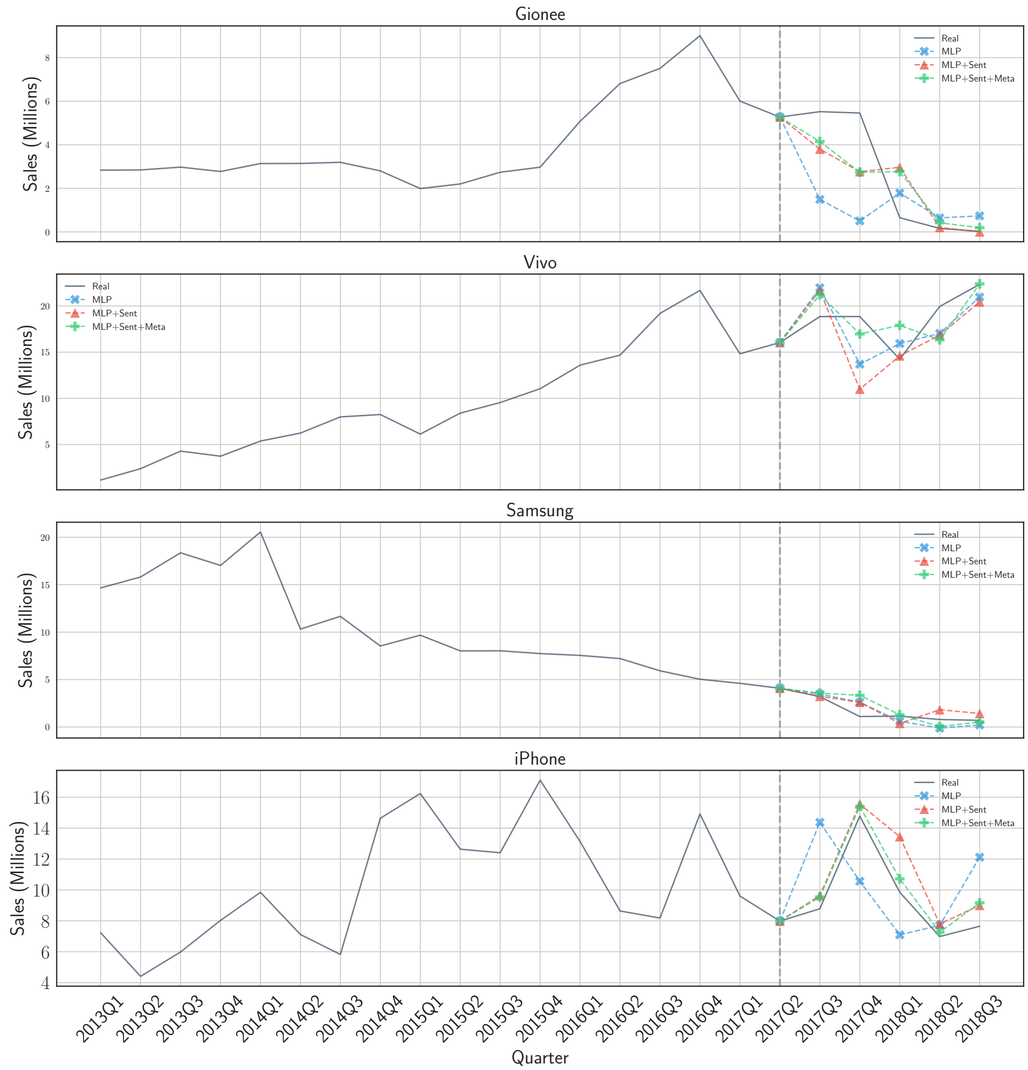

Figure 4: The sales prediction for Gionee, Vivo, Samsung and iPhone: Grey line represents the real sales, the blue line represents the prediction of MLP without sentiment information, the red line represents the prediction of MLP with sentiment information, and the green line represents the prediction of meta trained MLP with sentiment information.

Shi *et al.*, 2016; Asur and Huberman, 2010]. Among the years, using sentiment analysis as an additional features for sales forecasting has been widely used in different domains. For instance, it has been used for predicting: movies sales [Duan *et al.*, 2008; Gaikar and Marakarkandy, 2015; Ahn and Spangler, 2014; Marshall *et al.*, 2013; Asur and Huberman, 2010], e-commerce products [Davis and Khazanchi, 2008; Tuarob and Tucker, 2013], car sales [Wijnhoven and Plant, 2017; Geva *et al.*, 2017; Barreira *et al.*, 2013]. To the best of our knowledge we are the first to report positive correlation between sentiment feature and smartphones quarter sales.

5.2 Meta-learning

Meta-learning [Thrun and Pratt, 1998; Schmidhuber, 1987; Schmidhuber, 1992; Naik and Mammone, 1992; Bengio *et al.*, 1992] also known as Learning To Learn, is machine learning technique that tries to learn the algorithm itself. Recently, several meta-learning models has been proposed for solving few-shot image classification [Ravi and Larochelle, 2017; Vinyals *et al.*, 2016; Finn *et al.*, 2017; Mishra *et al.*, 2017; Santoro *et al.*, 2016], optimization [Andrychowicz *et al.*, 2016], dialogue system [Lin *et al.*, 2019] and reinforcement learning [Finn *et al.*, 2017]. In our setting, we are applying Meta-learning for learning a set of parameter that can adapt to certain products, and have good performance in sales prediction.

6 Conclusion

In this paper, we explore four different sales prediction models *SARIMAX*, *SVR*, *Linear Regression* and *MLP*. The results of our experiments show that sentiment information improves the performance of these models which confirms the effectiveness of the sentiment index. Moreover, the proposed meta-learning method help models transfer the knowledge of sentiment-sales correlation from different products, further reduce the sales prediction error.

Acknowledgments

This work has been partially funded by ITF/319/16FP and MRP/055/18 of the Innovation Technology Commission, the Hong Kong SAR Government.

References

[Ahn and Spangler, 2014] Hyung-Il Ahn and W Scott Spangler. Sales prediction with social media analysis. In *2014 Annual SRII Global Conference*, pages 213–222. IEEE, 2014.

[Andrychowicz *et al.*, 2016] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*, pages 3981–3989, 2016.

[Asur and Huberman, 2010] Sitaram Asur and Bernardo A Huberman. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 492–499. IEEE Computer Society, 2010.

[Barreira *et al.*, 2013] Nuno Barreira, Pedro Godinho, and Paulo Melo. Nowcasting unemployment rate and new car sales in south-western europe with google trends. *NETNOMICS: Economic Research and Electronic Networking*, 14(3):129–165, 2013.

[Bengio *et al.*, 1992] Samy Bengio, Yoshua Bengio, Jocelyn Cloutier, and Jan Gecsei. On the optimization of a synaptic learning rule. In *Preprints Conf. Optimality in Artificial and Biological Neural Networks*, pages 6–8. Univ. of Texas, 1992.

[Ceron and d'Adda, 2016] Andrea Ceron and Giovanna d'Adda. E-campaigning on twitter: The effectiveness of distributive promises and negative campaign in the 2013 italian election. *New media & society*, 18(9):1935–1955, 2016.

[Davis and Khazanchi, 2008] Alanah Davis and Deepak Khazanchi. An empirical study of online word of mouth as a predictor for multi-product category e-commerce sales. *Electronic markets*, 18(2):130–141, 2008.

[Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Duan *et al.*, 2008] Wenjing Duan, Bin Gu, and Andrew B Whinston. The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry. *Journal of retailing*, 84(2):233–242, 2008.

[Finn *et al.*, 2017] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.

[Gaikar and Marakarkandy, 2015] Dipak Gaikar and Bijith Marakarkandy. Product sales prediction based on sentiment analysis using twitter data. *Int. J. Comput. Sci. Inf. Technol.(IJCSIT)*, 6(3):2303–2313, 2015.

[Geva *et al.*, 2017] Tomer Geva, Gal Oestreicher-Singer, Niv Efron, and Yair Shimshoni. Using forum and search data for sales prediction of high-involvement projects. *MIS Quarterly*, 41(1):65–82, 2017.

[Hennig-Thurau *et al.*, 2003] Thorsten Hennig-Thurau, Gianfranco Walsh, and Gianfranco Walsh. Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the internet. *International journal of electronic commerce*, 8(2):51–74, 2003.

[Hennig-Thurau *et al.*, 2004] Thorsten Hennig-Thurau, Kevin P Gwinner, Gianfranco Walsh, and Dwayne D Gremler. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52, 2004.

[Lassen *et al.*, 2014] Niels Buus Lassen, Rene Madsen, and Ravi Vatrapu. Predicting iphone sales from iphone tweets. In *Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International*, pages 81–90. IEEE, 2014.

[Lin *et al.*, 2019] Zhaojiang Lin, Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. Personalizing dialogue agents via meta-learning. *ArXiv*, abs/1905.10033, 2019.

[Liu, 2012] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

[Marshall *et al.*, 2013] Pablo Marshall, Monika Dockendorff, and Soledad Ibáñez. A forecasting system for movie attendance. *Journal of Business Research*, 66(10):1800–1806, 2013.

[Mishra *et al.*, 2017] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. *ICLR*, 2017.

[Naik and Mammone, 1992] Devang K Naik and RJ Mammone. Meta-neural networks that learn by learning. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 1, pages 437–442. IEEE, 1992.

[Ravi and Larochelle, 2017] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[Santoro *et al.*, 2016] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850, 2016.

[Schmidhuber, 1987] Jurgen Schmidhuber. Evolutionary principles in self-referential learning. on learning now to learn: The meta-meta-meta...-hook. Diploma thesis, Technische Universitat Munchen, Germany, 14 May 1987.

[Schmidhuber, 1992] Jürgen Schmidhuber. Learning to control fast-weight memories: An alternative to dynamic recurrent networks. *Neural Computation*, 4(1):131–139, 1992.

[Shi *et al.*, 2016] Xiaohui Shi, Feng Li, and Ali Ziaee Bigdeli. An examination of npd models in the context of business models. *Journal of Business Research*, 69(7):2541–2550, 2016.

[Thrun and Pratt, 1998] Sebastian Thrun and Lorien Pratt, editors. *Learning to Learn*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[Tuarob and Tucker, 2013] Suppawong Tuarob and Conrad S Tucker. Fad or here to stay: Predicting product market adoption and longevity using large scale, social media data. In *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages V02BT02A012–V02BT02A012. Citeseer, 2013.

[Vinyals *et al.*, 2016] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pages 3630–3638, 2016.

[Wijnhoven and Plant, 2017] Alphonsus BJM Wijnhoven and Olivia Plant. Sentiment analysis and google trends data for predicting car sales. In *38th International Conference on Information Systems 2017*, 2017.

[Winata *et al.*, 2019] Genta Indra Winata, Andrea Madotto, Zhaojiang Lin, Jamin Shin, Yan Xu, Peng Xu, and Pascale Fung. CAiRE_HKUST at SemEval-2019 task 3: Hierarchical attention for dialogue emotion classification. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 142–147, Minneapolis, Minnesota, USA, June 2019. Association for Computational Linguistics.