Abstract

Valorization is one of the most heated discussions in the business community, and commodities valorization is one subset in this task. Features of a product is an essential characteristic in valorization and features are categorized into two classes: graphical and non-graphical. Nowadays, the value of products is measured by price. The goal of this research is to achieve an arrangement to predict the price of a product based on specifications of that. We propose five deep learning models to predict the price range of a product, one unimodal and four multimodal systems. The multimodal methods predict based on the image and non-graphical specification of product. As a platform to evaluate the methods, a cellphones dataset has been gathered from GSMArena. In proposed methods, convolutional neural network is an infrastructure. The experimental results show 88.3% F1-score in the best method.

Keywords: Price Prediction; Multimodal Learning; Convolutional Neural Network; Inception; Classification

1. Introduction

Prediction has always been attractive to human beings and also has significant effects on their future choices. Price prediction is one of the examples related to forecasting tasks. Furthermore, prediction and analysis of data price have played a crucial role in today's economy (Hiransha, Gopalakrishnan, Menon & Soman 2018). Examples of price prediction include stock prediction (Parmar, Agarwal, Saxena, Arora, Gupta, Dhiman & Chouhan 2018, Selvin, Vinayakumar, Gopalakrishnan, Menon & Soman 2017, Singh & Srivastava 2017), electricity price forecasting (Vilar, Aneiros & Raña 2018), ticket pricing (Abdella, Zaki, Shuaib & Khan 2019, Langseva, Mukhina, Nikishova, Ivanov & Knyazkov 2015), cryptocurrency forecasting (Azari 2019, Kadiroglu, Akilotu & Sengur 2019, Phaladisailoed & Numnonda 2018, Lahmiri & Bekiros 2019, Ji, Kim & Im 2019), product price prediction such as house (Park & Bae 2015) and smartphone (Chandrashekhar, Thungamani, Gireesh Babu & Manjunath 2019).

Predicting price can be seen as regression or classification task. For the regression task, the actual price is predicted and for the classification task, price range in form of classes is predicted. The accuracy of predicted price can be changed based on the number of classes. On the other hand, the price prediction task itself, regardless of being regression or classification, is made using the input features of the products. These features can be in any modality and form. For example, a product such as cellphone has features in both numerical and image form. Multimodal prediction and learning is popular in
other fields of AI too (Asgari-Chenaghlou, Feizi-Derakhshi, Farzinvash, Balafar & Motamed 2020). Even the text related to user comments can be investigated using different sentiment analysis and text classification methods (Minaee, Kalchbrenner, Cambria, Nikzad, Chenaghlou & Gao 2020).

Combining various and different modalities requires a more robust model that can generalize its learning for sample with different aspects. For case of cellphone price prediction, the RGB image data that is in form of a $M \times N \times 3$ matrix is combined with features such as 3G network support that is a boolean feature.

Although price prediction has been addressed using a variety of techniques such as machine learning and deep learning approaches, requiring a huge number of data for training prohibits researchers to use these approaches on small datasets. Hence, in this paper, a real cellphone dataset for applying learning-based approaches to cellphone price prediction has been provided. The collection of dataset itself, combined with the proposed multimodal learning approaches provides an insightful view to the price prediction problem.

However, some previous works proposed various methods for the problem of cellphone price prediction (Nasser & Al-Shawwa 2019, Asim & Khan 2018, Chandrashekhara et al. 2019), in current study we propose five novel models that aim to solve it.

The rest of this study is organized in the following manner: the second section contains a literature review of mobile price prediction. Section 3 presents the proposed methods. The experimental results are given and discussed in section 4. Some conclusions are drawn in the final section, and the areas for further researches are identified, as well.

2. Related study

Price prediction has long history in many fields. Nevertheless, in products price prediction, especially cellphone price prediction, there is a lack of works. Also, in the following, related works has been described. After an overview of price prediction, some related works about image feature extraction have been introduced.

Asim and Khan (Asim & Khan 2018) predicted a cellphone’s price in the form of classification. Prices were classified in four classes (very economical ($< 150$), economical (151 - 300), expensive (301 - 450), very expensive ($> 450$)). To obtain the classes Decision Tree (J48) classifier and Naive Bayes classifier were used. As a dataset, ten features of 134 cellphones were collected from GSMArena website. Two features brand, and memory card slot categorized features, and the rests were numerical.

Nasser and Al-Shawwa (Nasser & Al-Shawwa 2019) implemented the dataset of Sharma (Sharma n.d.) to their predictor. As (Asim & Khan 2018) the authors of (Nasser & Al-Shawwa 2019) solved prediction as a classification problem. Twenty features of mobile devices are considered as input and class of price as the output of the system. Prices were classified into four groups. The designed ANN is three layers neural network. It should be also noted that the results reported in (Nasser & Al-Shawwa 2019) are not accessible. We implemented the proposed ANN and evaluated it by the same dataset. The authors reported too far than the results that we have achieved, and the maximum accuracy of the implemented method was 49.3%.

Unlike the previous methods, regression models were utilized in (Chandrashekhara et al. 2019). These three regression methods were Support vector regression (SVR), Multiple linear regression, and Backpropagation neural network. 262 cellphones’ information was collected that consists of 13 features and prices.

In (Subhiksha, Thota & Sangeetha 2020), three regression-based price prediction models have been studied. Among SVM, Logistic regression, and Random forest authors reported that SVM and logistic regression has reached the best accuracy (81%). Proposed methods have been applied to Sharma (Sharma n.d.) dataset.

There are various methods to model an image or extract features. The most popular one is convolutional neural network (CNN) (LeCun, Boser, Denker, Henderson, Howard, Hubbard & Jackel 1989). This neural network is the foundation of many models, such as LeNet (LeCun, Bottou, Bengio & Haffner 1998), AlexNet (Krizhevsky, Sutskever & Hinton 2017), GoogleNet (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke & Rabinovich 2015), and Inception (Szegedy, Vanhoucke, Ioffe, Shlens & Wojna 2016, Szegedy, Ioffe, Vanhoucke & Alemi 2017). Inception series models have been categorized as depth-based CNNs (Khan, Sohail, Zahoora & Qureshi 2020) and InceptionV3 introduced in (Szegedy et al. 2016) are used in this paper. For more comprehensive understanding, Figure 2 illustrates the
Fig. 1. An overview of the idea expressed. Also features that GSMArena offered as essence and collected in this dataset. Example ”Xiaomi Redmi Note 8 Pro”
3. Method

In the present work, we propose two different approaches for predicting cellphones price. As described before, our goal is to classify the price. To this aim, five methods have been proposed. For the first method, a unimodal model and for the rest of methods, multimodal approaches are proposed. As the second method, Inception-V3 is used for images feature extraction. Subsections 3.2, 3.3, 3.4 equally have gotten a benefit from convolutional deep learning method as images feature extraction part and their differences are in textual features extraction and concatenating parts.

3.1 Model 1: Unimodal

The unimodal proposed method uses non-graphical features of cellphones. This method aims to be a platform for evaluating our dataset and methods with other state-of-the-art methods and multimodal approaches. Figure 3 shows the neural network architecture.

3.2 Model 2: Inception-based image feature extraction

Inception is an innovative model based on CNN that we used for the cellphone image’s feature extraction in this method. There are four versions of Inception, and in this paper, Inception-V3 (Szegedy et al. 2016) has been used. The output of Inception-V3 is a label, but we need embedding layer as feature extraction. In other words, the output of the second "Grid Size Reduction" (Figure 2) has been selected as the embed of images.

Images feature extraction is one side of work, and this work is multimodal. In the other side, dense neural network is applied for non-graphical part’s feature extraction. In the end, these two parts should be concatenated. Three dense layer networks do this duty and classify the price of the cellphone. Figure 4 shows the structure of this method.

3.3 Model 3: Convolutional image feature extraction and dense concatenating

Convolutional neural networks are one of the most popular and practical deep learning methods of feature extraction from images. So in the rest methods, CNNs are deployed as the underlying of architecture. In this method, three layers convolution with maxpooling in each layer are established as the graphical feature extraction. In the other side of feature extraction, three dense layers with a dropout have been used. Furthermore, in the concatenation segment, three dense layers with 300, 120, and 4 nodes have been planted. In order to comprehensive understanding, Figure 5 presents this method. Also, in Figure 6 outputs of each convolution layer have been shown.

3.4 Model 4: Convolutional image and textual features with dense concatenating

In this model, our approach is to extract features from non-graphical features by convolutional neural network. In such a way that non-graphical features are recorded in a matrix structured format and then as the Figure 7, two convolution layers with maxpooling after each layer have been applied. In concatenation, three dense layers by 300, 120, and 4 neurons have to be classified by the features.

3.5 Model 5: Fully convolutional network

The main goal of this method was to obtain a fully convolutional method that calculates the feature extraction part by only convolutional layers. As figure 7 shows, non-graphical features are calculated by two convolution layers and images features are extracted by three convolution layers with another convolution layer in size of $1 \times 1$, and in the concatenation part as the other models three dense layers of neural network classify the cellphone’s price range.

4. Experiments and results

As the aim of this section, experiments and evaluations of the proposed models will be discussed. To evaluate the proposed methods, a dataset that satisfies the problem is needed. For this purpose, at the first, the dataset will be described and then evaluations will be explained in detail.

4.1 Dataset

Due to lack of multimodal dataset in this topic, we describe how the dataset is gathered and it contains which features and information of cellphones in this part.

The dataset has been collected from GSMArena. GSMArena is a specialized and reputable website that provides cellphones’ specifications. In this website, each cellphone has a page that consists of a full specification of it.

Selection of important features is an essential part of data-dependent projects. Also in this

---

1 www.gsmarena.com
Fig. 2. Architecture of Inception-v3 (Review: *Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015* n.d.).

Fig. 3. Unimodal method’s neural network architecture.
Fig. 4. Architecture of Inception-based method (Model 2).

Fig. 5. Model 3’s neural network design.

Fig. 6. Visualization of different convolutional layers for model 3.
Max Pooling
2×2
Convolution
90@5×5
Max Pooling
2×2
Convolution
50@5×5
Convolution
Max Pooling
2×2
Convolution
20@5×5
Convolution
Max Pooling
2×2
Convolution
64@1×1
Max Pooling
2×2
Convolution
32@3×3
Max Pooling
2×2
FC
FC
Softmax
class
probabilities

Fig. 7. Neural network architecture of method 4. The only difference between method 4 and 5 is on "FC" layers of each feature selection parts. In method five, "FC" layers dismissed and Flattened arrays passed to concatenation part.

project, features selection should be discussed. A mobile phone device has many features. For example, "Xiaomi Redmi Note 8 Pro" consists of 49 features on 58 tuples that classified in 11 classes. In most of the time, this amount of features are not necessary. Indeed the GSMArena offered essence feature of a cellphone on top of each specific page. These features have been shown on fig[1]. The gathered dataset is made up of fig[1]'s features. The gathered dataset’s features have been listed as below:

1. Brand
2. Model
3. Release Date
4. Weight
5. Operation System
6. Storage
7. Hit
8. Hit Count
9. Display Size
10. Display Resolution
11. Camera
12. Video
13. Processor
14. Ram
15. Battery
16. Battery Type
17. Picture
18. Price (Euro)

The gathered data are raw and should be normalized before any use. In this regard, changes have been listed as below:

- **Weight**: Non-numerical characters such as "gr" have been deleted, and this column’s values are numerical.
- **OS**: There were too many customized models or names. To reduce overplus states, this feature has been configured that consist of 20 modes.
- **Display size**: Display size quotes as width × height. In this form, the feature should be processed as categorical. To access the numerical from of this feature, this column has been broken into two columns called "V_resolution" and "H_resolution" that vertical and horizontal pixels have been put in them.
- **Video**: In video’s resolution, format is an integer with "p" as the abbreviation of pixel. To make this feature a numerical feature, "p" is deleted.
- **Processor**: In GSMArena exact model of cellphones processor’s chipset will be written, if it is well-known. This type makes too many possibilities for the system as a category feature. Thus only the brand of processor and all of 26 types of processor are held.
- **Ram**: It is necessary that all values be on the same unit, so all values have been converted to Mb (MegaByte).
- **Hit & Hit Count**: These two features do not affect price prediction but only they maybe show the users perspective. So these were not used as proposed multimodal’s non-graphical feature.

For the last part, classification labels should be described. From our point of view, prices have been classified into four groups. These groups are created by thresholding prices as below:

- 0: < 250
- 1: 250 ≤ & < 500
- 2: 500 ≤ & < 750
3: \geq 750

At the end, it should be noticed that the dataset contains 3165 full information records. The learning rate has been set to 0.001, and the RMSProp optimizer is used for all of the five models.

4.2 Results

In this part of the paper, the results of the proposed system have been entered in 4.2.2 based on evaluation metrics of part 4.2.1. All tests were run via python on a Core i7-5600U 2.6 GHz processor with 12 GB of RAM and Nvidia 840M graphic processor.

4.2.1 Evaluation Measures

Precision, recall, f-measure, and accuracy are the most well-known evaluation measures in binary classification systems. Accuracy is primary metric to evaluate the quality of a classification model. Let TP, FP, TN, FN denote true positive, false positive, true negative, and false negative, respectively. The classification accuracy defined in Eq. 3. Precision and recall for binary classification are defined as Eq. 1, 2. The F1 measure is the harmonic mean of the precision and recall, as in Eq. 4.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (1)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (2)
\]

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]

\[
F1 - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} \times \text{recall}} \quad (4)
\]

4.2.2 System Outputs

In this part, results of proposed methods will be described and comparison of results with the state-of-the-art methods will be shown. The results will be on metrics described in part 4.2.1. As shown, method 3 has achieved the best results. Hence, for subsequent comparisons, method 3 has been used. Fig 8 is the chart illustrated by the information of the table 2.

As the final step, proposed methods have been tested by the state-of-the-art methods. To this aim, only two methods, (Asim & Khan 2018) & (Subhiksha et al. 2020), solved the problem based on classification and reported the results in the same units (accuracy). To summarize the comparison, the best method of our techniques (method 3) has been compared with (Asim & Khan 2018) and (Subhiksha et al. 2020) and the comparison results could be seen in table 3. Based on these results, the proposed method gets better results. It should be also noticed that these results have been reported by the authors and we did not implement them.

5. Conclusion

Proposing a multimodal price prediction system is the primary goal of this paper. The first step in this way is to gather an appropriate dataset that consists of both graphical and non-graphical features with the price. To this aim, cellphones information is seemed to be suitable, and specifications

| Dataset              | Precision | Recall | Accuracy | F-measure |
|----------------------|-----------|--------|----------|-----------|
| Sharma (Sharma n.d.) | 0.830     | 0.830  | 0.829    | 0.825     |
| Gathered dataset     | 0.880     | 0.855  | 0.855    | 0.861     |

| Method                | Precision | Recall | Accuracy | F-measure |
|-----------------------|-----------|--------|----------|-----------|
| Method 1              | 0.880     | 0.855  | 0.855    | 0.861     |
| Method 2              | 0.894     | 0.868  | 0.868    | 0.878     |
| Method 3              | 0.891     | 0.880  | 0.880    | 0.883     |
| Method 4              | 0.881     | 0.830  | 0.830    | 0.853     |
| Method 5              | 0.837     | 0.830  | 0.830    | 0.829     |

| Method                        | Accuracy |
|-------------------------------|----------|
| Asim and Khan (Asim & Khan 2018) | 0.742    |
| Subhiksha et al. (Subhiksha et al. 2020) | 0.81     |
| **Method 3** (best of the methods) | **0.880** |
and picture of cellphones were collected from GS-MArena. After that, five methods were proposed that one of them is unimodal and the rest are multimodal systems. The multimodal methods were created by Inception-V3 and convolutional neural networks and a combination of them with dense neural network. The evaluation of proposed methods was compared with not many available systems because this research presented a new strategy. All results show that presented methods give better performance in comparison with the other available researches.

References
Abdella, J. A., Zaki, N., Shuaib, K. & Khan, F. (2019), ‘Airline ticket price and demand prediction: A survey’.

Asgari-Chenaghlu, M., Feizi-Derakhshi, M. R., Farzinbash, L., Balafar, M. A. & Motamed, C. (2020), ‘A multimodal deep learning approach for named entity recognition from social media’.

Asim, M. & Khan, Z. (2018), ‘Mobile price class prediction using machine learning techniques’, International Journal of Computer Applications 179(29), 6–11.

Azari, A. (2019), ‘Bitcoin price prediction: An arima approach’. URL: http://arxiv.org/abs/1904.05315

Chandrashekara, K. T., Thungamani, M., Gireesh Babu, C. N. & Manjunath, T. N. (2019), Smartphone price prediction in retail industry using machine learning techniques, in ‘Lecture Notes in Electrical Engineering’.

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K. & Soman, K. P. (2018), NSE Stock Market Prediction Using Deep-Learning Models, in ‘Procedia Computer Science’.

Ji, S., Kim, J. & Im, H. (2019), ‘A comparative study of bitcoin price prediction using deep learning’, Mathematics.

Kadiroglu, Z., Akilotu, B. N. & Sengur, A. (2019), Mechanism of bitcoin and investigation of the studies in the literature related to bitcoin, in ‘1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings’, Institute of Electrical and Electronics Engineers Inc.

Khan, A., Sohail, A., Zahoora, U. & Qureshi, A. S. (2020), ‘A survey of the recent architectures of deep convolutional neural networks’, Artificial Intelligence Review pp. 1–62. URL: http://link.springer.com/10.1007/s10462-020-09825-6

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017), ‘Imagenet classification with deep
convolutional neural networks’, *Communications of the ACM* **60**(6), 84–90.  
**URL:** https://dl.acm.org/doi/10.1145/3065386

Lahmiri, S. & Bekiros, S. (2019), ‘Cryptocurrency forecasting with deep learning chaotic neural networks’, *Chaos, Solitons and Fractals* **118**, 35–40.

Lantseva, A., Mukhina, K., Nikishova, A., Ivanov, S. & Knyazkov, K. (2015), Data-driven modeling of airlines pricing, *in* ‘Proceedia Computer Science’, Vol. 66, Elsevier B.V., pp. 267–276.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989), ‘Backpropagation applied to handwritten zip code recognition’, *Neural Computation* **1**(4), 541–551.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998), ‘Gradient-based learning applied to document recognition’, *Proceedings of the IEEE* **86**(11), 2278–2323.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlou, M. & Gao, J. (2020), ‘Deep learning based text classification: A comprehensive review’, *arXiv preprint arXiv:2004.03705*.

Nasser, I. M. & Al-Shawwa, M. (2019), Ann for predicting mobile phone price range, Technical report.  
**URL:** www.GSMArena.com.

Park, B. & Bae, J. K. (2015), ‘Using machine learning algorithms for housing price prediction: The case of Fairfax county, Virginia housing data’, *Expert Systems with Applications* **42**(6), 2928–2934.

Parmar, L., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H. & Chouhan, L. (2018), Stock market prediction using machine learning, *in* ‘ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications’, pp. 574–576.

Phaladisaloed, T. & Numnonda, T. (2018), Machine learning models comparison for bitcoin price prediction, *in* ‘Proceedings of 2018 10th International Conference on Information Technology and Electrical Engineering: Smart Technology for Better Society, ICITEE 2018’, Institute of Electrical and Electronics Engineers Inc., pp. 506–511.

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K. & Soman, K. P. (2017), Stock price prediction using lstm, rnn and cnn-sliding window model, *in* ‘2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017’.

Sharma, A. (n.d.), ‘Mobile price classification — kaggle’.  
**URL:** https://www.kaggle.com/iabhishekofficial/mobile-price-classification

Singh, R. & Srivastava, S. (2017), ‘Stock prediction using deep learning’, *Multimedia Tools and Applications*.

Subhiksha, S., Thota, S. & Sangeetha, J. (2020), Prediction of phone prices using machine learning techniques, *in* ‘Advances in Intelligent Systems and Computing’, Vol. 1079, Springer, pp. 781–789.

Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. (2017), Inception-v4, inception-resnet and the impact of residual connections on learning, *in* ‘31st AAAI Conference on Artificial Intelligence, AAAI 2017’, AAAI press, pp. 4278–4284.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015), Going deeper with convolutions, *in* ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Vol. 07-12-June-2015, IEEE Computer Society, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016), Rethinking the inception architecture for computer vision, *in* ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Vol. 2016-December, IEEE Computer Society, pp. 2818–2826.
Vilar, J., Aneiros, G. & Raña, P. (2018), ‘Prediction intervals for electricity demand and price using functional data’, *International Journal of Electrical Power and Energy Systems*.