UDC 004.8 (045)
DOI: 10.18372/1990-5548.55.12778

[1]B. M. Sineglazov,
[2]M. O. Omelchenko,
[3]V. P. Hotsyanivskyy

INTELLIGENT WEATHER FORECAST SYSTEM

[1,2,3]Aviation Computer-Integrated Complexes Department, Educational & Scientific Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine
E-mail: [1]svm@nau.edu.ua, [2]max.omel4enko@yandex.ua,[3]sttt912@yahoo.com

Abstract*—It is considered the creation of intelligent weather forecast system in traditional statement as the prediction of time series For the solution of this problem it is chosen a random forest algorithm. Examples of program implementation of the algorithm are given using libraries for the python programming language, namely Pandas and Skicit-learn. It is considered a problem solution for a number of real examples and corresponding results are given.*

Index Terms— Machine learning; python; data science.

I. INTRODUCTION

The weather studying has great practical meaning, because of its prognosis is needed in many areas. At present, forecasts are made for the general use, orientated on all media, as well as on special ones - intended for the satisfaction of certain economy branches requests, taking into account the specifics of different professions people activities (aviators, sailors, builders, farmers, etc.).

Weather forecasting from a scientific point of view is a complex physical problem, for the solution of which there are several methods, but in full measure for all meteorological values and phenomena characterizing the state of weather, no method provides an exact solution.

Weather can be predicted by local features therefore the observation of local weather patterns is of great practical importance. Such features of weather (temperature change, character of clouds, pressure, wind force, etc.) can be observed on the phenomena of inanimate nature, plants, animals.

Another important indicator for predicting weather is the climate. The climate is named long-term weather conditions, due to solar radiation, the nature of the underlying surface and the associated circulation of the atmosphere in the area for a long time. The climate is formed as a result of a natural sequence of meteorological processes, defined by the complex of physical and geographical conditions and expressed in long-term weather mode.

Using information provided climate parameters and local features, it is formed the history of meteorological researches. Using the methods of machine learning to analyze the previous information, it is possible to predict the weather for a long period of time.

Machine learning researches the learning and construction of algorithms that can be adjusted from data, and perform a predictive analysis based on it. [1]. Such algorithms operate by way of model constructing and its adjusting with help of training set incoming observations for forecasts creation and making decisions that are expressed as outputs of the system [2].

II. PROBLEM STATEMENT

Within the field of data analysis a methods of machine learning used for a finding of algorithms that serve prediction. In commercial applications, this phenomenon is known as forecast for analytics. These analytical models allow researchers, engineers and analysts "to produce reliable solutions and results" and to reveal "hidden understanding" by learning from the historical relationships and trends in the data.

On this basis, it is possible to create a model of machine learning, with which it is possible to predict the weather based on information from previous years. The main task of the program will be to predict the maximum temperature for the desired days based on meteorological data over the past years.

The solution of given problem includes the execution of follows steps

– data receiving;
– correction of missing data;
– data converting into a format that is accessible to the machine;
– data preparation for a machine learning model;
– creating of basic model;
– model learning ;
– forecast receiving;
– results presentation in visual or numerical form.

III. PROBLEM SOLUTION

As the sourse of information of meteorological it is chosen NOAA Climate Date Online. Servis NOAA easy to use, and temperature data can be downloaded in the form of ready files csv, which can be processed on the following language Python.

The following Python code loads data and displays their structure (Fig. 1.)

```
import pandas as pd
features = pd.read_csv('temps.csv')
features.head(5)
```

	year	month	day	week	temp_2	temp_1	average	actual	forecast_noaa	forecast_acc	forecast_under	friend
0	2016	1	1	Fri	45	45	45.6	45	43	50	44	29
1	2016	1	2	Sat	44	45	45.7	44	41	50	44	61
2	2016	1	3	Sun	45	44	45.8	41	43	46	47	56
3	2016	1	4	Mon	44	41	45.9	40	44	48	46	53
4	2016	1	5	Tues	41	40	46.0	44	46	46	46	41

Fig. 1. Structure of weather data from the NOAA Climate site Date Online

If look at the amount of data, can see that there is information for 348 days. In our case, the missing data is not critical because taken from a reliable source. To determine whether these abnormalities choice in value, conduct statistical calculations.

features.describe()

	year	month	day	temp_2	temp_1	average	actual	forecast_noaa	forecast_acc	forecast_under	friend
count	348.0	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000	348.000000
mean	2016.0	6.477011	15.514368	62.652299	62.701149	59.760632	62.543103	57.238506	62.373563	59.772989	60.034483
std	0.0	3.498380	8.772982	12.165398	12.120542	10.527306	11.794146	10.605746	10.549381	10.705256	15.626179
min	2016.0	1.000000	1.000000	35.000000	35.000000	45.100000	35.000000	41.000000	46.000000	44.000000	28.000000
25%	2016.0	3.000000	8.000000	54.000000	54.000000	49.975000	54.000000	48.000000	53.000000	50.000000	47.750000
50%	2016.0	6.000000	15.000000	62.500000	62.500000	58.200000	62.500000	56.000000	61.000000	58.000000	60.000000
75%	2016.0	10.000000	23.000000	71.000000	71.000000	69.025000	71.000000	66.000000	72.000000	69.000000	71.000000
max	2016.0	12.000000	31.000000	117.000000	117.000000	77.400000	92.000000	77.000000	82.000000	79.000000	95.000000

Fig. 2. Summary data that proves that there are no abnormal values in the sample

As can be seen from Fig. 2, did not find any points that showed themselves to be abnormal, nor do we have columns with zero values of the measurement results.

It is necessary to make a small modification so that the model of machine learning can read the information correctly. For this we will use the Python-library Pandas, based on the known structure dataframe.

Need simple data encoding, which will convert verbal names of the days of the week into their numerical counterparts. This step is needed, because the short days of the week are intuitive for us, but the machines do not have this knowledge.

It is known the numbers it is necessary provide for machine learning. Thus, the information must be brought to the required form by the following team.

```
features = pd.get_dummies(features)
features.iloc[:,5:].head(5)
```

Got information in the following form (Fig. 3).

	average	actual	forecast_noaa	forecast_acc	forecast_under	friend	week_Fri	week_Mon	week_Sat	week_Sun	week_Thurs	week_Tues	week_Wed
0	45.6	45	43	50	44	29	1	0	0	0	0	0	0
1	45.7	44	41	50	44	61	0	0	1	0	0	0	0
2	45.8	41	43	46	47	56	0	0	0	1	0	0	0
3	45.9	40	44	48	46	53	0	1	0	0	0	0	0
4	46.0	44	46	46	46	41	0	0	0	0	0	1	0

Fig. 3. Data after direct coding

After this, transform information in the Numpy arrays, this is required in order for the machine learning library to read this data, for this we use the following code

```
import numpy as np
labels = np.array(features['actual'])
features= features.drop('actual', axis = 1)
feature_list = list(features.columns)
features = np.array(features)
```

Before can get and evaluate the forecasts, need to conduct analytical work to establish a limit for model.

The base forecast for our case will be the average temperature. In other words, our base line is an error we would have if simply predicted the average maximum temperature for all days.

The base forecast for our case will be the average temperature. Base line is an error we would have if simply predicted the average maximum temperature for all days.

```
from sklearn.model_selection import train_test_split
train_features, test_features, train_labels, test_labels = train_test_split(features, labels, test_size = 0.25, random_state = 42)
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor(n_estimators = 1000, random_state = 42)
rf.fit(train_features, train_labels);
predictions = rf.predict(test_features)
errors = abs(predictions - test_labels)
mape = 100 * (errors / test_labels)
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')
```

Make predictions on the test information. Then can compare predictions with known answers. Having compared the data for 2016 we have the result in 93.94% of the correct answers.

V. CONCLUSIONS

The basic problem of weather forecasting is the complexity and cost of the research. Therefore, it is possible to use machine learning algorithms to process optimization.

Although here are their nuances, namely the fact that the data needs to be worked out in detail, remove artifacts that take a long time, but this procedure needs to be done only once.

In this paper consider the methods of teaching the model of machine learning to predict the maximum temperature for the required days. For the solution of this problem it is chosen a random forest algorithm. Examples of program implementation of the algorithm are given using libraries for the python programming language, namely Pandas and Skicit-learn.

REFERENCES

[1] Ron Kohavi; Foster Provost, “Glossary of terms,” *Machine Learning*, Kluwer Academic Publishers, Boston, Manufactured in The Netherlands, vol. 30, pp. 271–274, 1998.

[2] Machine learning and pattern recognition "can be viewed as two facets of the same field."

Received December 19, 2017

Sineglazov Victor. Doctor of Engineering Science. Professor.
Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant.
Publications: more than 600 papers.
E-mail: svm@nau.edu.ua

Omelchenko Maxym. Student.
Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine.
Research interests: Artificial Neural Network
Publications: 2.
E-mail: max.omel4enko@yandex.ua

Hotsyanivskyy Vladimir. Student.
Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine.

Research interests:
Publications: 2.
E-mail: sttt912@yahoo.com

В. М. Синєглазов, М. О. Омельченко, В. П. Хоцянівський. Інтелектуальна система прогнозування погоди
Розглянуто створення інтелектуальної системи прогнозування погоди в традиційному викладі прогнозування часових рядів. Для вирішення цієї проблеми обраний випадковий алгоритм лісу. Приклади реалізації програми алгоритму представлено з використанням бібліотек для мови програмування Python, а саме Pandas і Skicit-learn. Показано вирішення проблеми для низки реальних прикладів.
Ключові слова: машинне навчання; пітон; аналіз даних.

Синєглазов Віктор Михайлович. Доктор технічних наук. Професор.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки.
Кількість публікацій: більше 600 наукових робіт.
E-mail: svm@nau.edu.ua

Омельченко Максим Олегович. Студент.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна.
Напрям наукової діяльності: штучний інтелект.
Кількість публікацій: 2.
E-mail: max.omel4enko@yandex.ua

Хоцянівський Володимир Петрович. Студент.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна.
Напрям наукової діяльності: штучний інтелект.
Кількість публікацій: 2.
E-mail: sttt912@yahoo.com

В. М. Синеглазов, М. О. Омельченко, В. П. Хоцяновский. Интеллектуальная система прогноза погоды
Рассмотрено создание интеллектуальной системы прогнозирования погоды в традиционном изложении как прогнозирование временных рядов. Для решения этой проблемы выбран случайный алгоритм леса. Примеры реализации алгоритма программно представлены с использованием библиотек для языка программирования Python, а именно Pandas и Skicit-learn. Показано решение проблемы для ряда реальных примеров.
Ключевые слова: машинное обучение; питон; анализ данных.

Синеглазов Виктор Михайлович. Доктор технических наук. Профессор.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Киевский политехнический институт, Киев, Украина (1973).
Направление научной деятельности: аэронавигация, управление воздушным движением, идентификация сложных систем, ветроэнергетические установки.
Количество публикаций: более 600 научных работ.
E-mail: svm@nau.edu.ua

Омельченко Максим Олегович. Студент.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина.
Направление научной деятельности: искусственный интеллект.
Количество публикаций: 2.
E-mail: max.omel4enko@yandex.ua

Хоцяновский Владимир Петрович. Студент.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина.
Направление научной деятельности: искусственный интеллект.
Количество публикаций: 2.
E-mail: sttt912@yahoo.com