

Prognoza czasu wykonywania zadań przez pracowników floty pojazdów dla planowania pracy floty pojazdów elektrycznych pod kątem ich ładowania

1. Marcin KOPYT, 2. Dariusz BACZYŃSKI , 3. Tomasz GULCZYŃSKI

Plan prezentacji

- Motywacja
- Opis zagadnienia
- Analiza statystyczna zmiennych wejściowych do prognoz czasu wykonania zadania
- Wybór kombinacji zmiennych wejściowych i pytania badawcze
- Prognozy czasu wykonania zadania
- Ocena przewidywanej użyteczności prognoz w potencjalnej optymalizacji pracy floty pojazdów elektrycznych, w tym czasów ładowania
- Odpowiedź na pytania badawcze

1

Motywacja

2

- Usługodawcy używający w swoich działalnościach floty pojazdów coraz częściej rozważają ich elektryfikację
- Operatorzy flot starają się maksymalizować użycie poszczególnych pojazdów by zyskać najwyższą stopę zwrotu z zainwestowanego kapitału
- Zasięg pojazdu i czas ładowania akumulatorów stanowią istotną różnicę w sposobie zarządzania flotą
- Wyznaczanie rozwiązania optymalnego może wymagać wielokrotnego rozwiązywania skomplikowanych problemów optymalizacyjnych z rodziny Vehicle Routing Problem
- Określenie rozwiązania optymalnego tego rodzaju problemów nie jest możliwe bez dokładnych informacji definiujących to zadanie. Jedną z tych informacji jest przewidywany czas postoju pojazdów w poszczególnych punktach trasy, związany z wykonywaniem przez techników floty zadań u klienta.

Opis zagadnienia

3

- W badaniach rozpatruje się pojazdy floty, której właścicielem jest jedna z dużych firm europejskich zajmująca się wykonywaniem zadań (usług) u klientów.
- Czas wykonywania zadania oznacza w badaniach czas od momentu opuszczenia przez pracownika pojazdu floty, przez realizację zadania, do powrotu do pojazdu.
- W badaniach każde parkowanie jest tożsame z jednym rekordem.
Brak więc możliwości stosowania zmiennych cofniętych jak w klasycznej autoregresji.
- Zmienne powiązane z procesem pochodzą z systemu zarządzania flotą. Dodatkowe zmienne pochodzą z systemu prognoz pogody oraz danych gospodarczych.
- Dla anonimizacji danych wszystkie zmienne ciągłe zostały znormalizowane metodą min-max.

Zmienne stosowane w badaniach

4

Zmienna	OPIS	Typ zmiennej
BR	Numer brygady	1
TC	Numer technika	1
TP	Typ zadania	1
RD	Rodzaj zadania	1
PU	Wymagany poziom umiejętności do wykonania zadania	1
NS	Liczba umiejętności potrzebna do wykonania danego zadania	1
DG	Długość geograficzna	2
SG	Szerokość geograficzna	2
IWD	Dzień tygodnia	3
H	Godzina zegarowa	3
WID1	Numer węzła siatki meteo 1	4
WID2	Numer węzła siatki meteo 2	4
WR1	Rząd siatki meteo 1	4
WC1	Kolumna siatki meteo 1	4
WR2	Rząd siatki meteo 2	4
WC2	Kolumna siatki meteo 2	4
MN1	Numer oczka siatki 1	4
MN2	Numer oczka siatki 2	4

Zmienna	OPIS	Typ zmiennej
T	Temperatura powietrza	5
a10	Kierunek wiatru na 10 m	5
a30	Kierunek wiatru na 30 m	5
lc	Ilość niskich chmur	5
pr	Ilość opadów deszczu	5
sn	Ilość opadów śniegu	5
tc	Zachmurzenie całkowite	5
v10	Prędkość wiatru na 10 m	5
v30	Prędkość wiatru na 30 m	5
pd	Gęstość zaludnienia	6
ETD	zgrubny szacunek	7
ave	Średni czas danego typu i rodzaju zadania	8
RTD	Rzeczywisty czas wykonania zadania	Zmienna wyjścia

Charakterystyka zmiennej prognozowanej

Miara	ETD [j.w.]	RTD [j.w.]
średnia	0,396	0,268
odch. Std.	0,058	0,180
min	0,000	0,000
p25	0,286	0,131
p50	0,375	0,240
p75	0,500	0,370
max	1,000	1,000

5



Analiza statystyczna powiązania zmiennych wejściowych ze zmienną prognozowaną

6

Zmienne rozdzielono przed badaniami na zmienne ciągłe i kategoryczne.

Do zmiennych zakodowanych dalej binarnie zakwalifikowano zmienne typu 1.
W badaniach przyjęto próg istotności statystycznej $\alpha = 0.05$

Dla zmiennych ciągłych zastosowano jako miary powiązania:

- korelację R-Pearsona (R),
- korelację tau-Kendalla (τ),
- informację wzajemną (mi)

Dla zmiennych kategorycznych zastosowano jako miary powiązania:

- korelację punktowo-dwuseryjną (pbs),
- informację wzajemną (mi)



Analiza statystyczna powiązania zmiennych wejściowych ze zmienną prognozowaną

7

Dla zmiennych ciągłych zaznaczono w tabelach wyników:

- Wartości $p > \alpha$
- 6 zmiennych o największej wartości modułu miary powiązania



Powiązania zmiennych wejściowych ze zmienną prognozowaną – wejścia ciągłe

8

Zmienna	DG	SG	PU	NS	IWD	H	WID1	WID2	WR1	WC1	WR2	WC2	MN1
R	-0,02539	0,07260	-0,05378	0,13975	-0,02295	-0,03310	0,06887	0,06863	0,06896	-0,03143	0,06871	-0,03154	0,06139
R_p	0,02243	6,3E-11	1,3E-06	1,5E-36	0,039056	0,002908	5,6E-10	6,5E-10	5,3E-10	0,00471	6,1E-10	0,00456	3,3E-08
τ	-0,00744	0,05120	-0,04625	0,05588	-0,01162	-0,02688	0,05333	0,05261	0,05484	-0,01436	0,05493	-0,01514	0,03883
τ_p	0,31541	5,0E-12	9,8E-08	8,5E-12	0,14910	0,00056	7,8E-13	1,6E-12	4,7E-13	0,05538	5,3E-13	0,04342	3,4E-07
mi	0,04038	0,02826	0,02157	0,03680	0,00614	0,00511	0,03058	0,03047	0,01955	0,03132	0,02750	0,02418	0,03097
Zmienna	MN2	T	a10	a30	lc	pr	sn	tc	v10	v30	pd	ave	
R	0,05987	-0,02702	0,01704	0,01588	0,00142	0,00270	-0,00244	0,01630	-0,01791	-0,00453	-0,08573	0,28730	
R_p	7,1E-08	0,01509	0,12556	0,15328	0,89822	0,80809	0,82658	0,14273	0,10728	0,68387	1,1E-14	5,5E-152	
τ	0,03741	-0,01644	0,00473	0,00589	0,00208	-0,00318	-0,00071	0,01154	-0,01460	-0,00685	-0,05081	0,19748	
τ_p	9,1E-07	0,02667	0,52356	0,42686	0,78040	0,68419	0,93577	0,12084	0,04904	0,35536	7,5E-12	1,8E-125	
mi	0,03122	0	0	0,01064	0,00198	0	0,00222	0	0	0	0,00801	0,06975	

Analiza statystyczna powiązania zmiennych wejściowych ze zmienną prognozowaną

9

Dla zmiennych kategorycznych pokazano w tabelach wyników:

- 5 zmiennych o najwyższym module siły powiązania ze zmienną wyjścia według danej miary
- Pogrubieniem wartości $p > \alpha$
- Tabele zawartości danego typu zmiennej kategorycznej w górnym 10 % wyników o największej sile powiązania wg. danej miary



Powiązania zmiennych wejściowych ze zmienną prognozowaną – wejścia kategoryczne

zmienna	<i>pbs</i>	<i>pbs_p</i>	<i>mi</i>	miara
TP2	0,16824	1,7E-52	0,00065	<i>pbs</i>
RD5	0,16785	3,1E-52	0,00368	<i>pbs</i>
RD8	0,15539	6,2E-45	0,00012	<i>pbs</i>
RD3	-0,14572	1,1E-39	0,00200	<i>pbs</i>
TP3	-0,13555	1,6E-34	0,00034	<i>pbs</i>
BR125	0,06205	2,3E-08	0,04126	<i>mi</i>
TC22	0,06559	3,4E-09	0,03947	<i>mi</i>
TC234	0,06205	2,3E-08	0,03329	<i>mi</i>
TC113	0,08227	1,2E-13	0,02160	<i>mi</i>
TC54	0,07046	2,1E-10	0,02008	<i>mi</i>

10

Powiązania zmiennych wejściowych ze zmienną prognozowaną – wejścia kategoryczne

11

- W danych kategorycznych można było zaobserwować zmienne bardziej korelujące z analizowanym procesem niż zmienne ciągłe, zarówno pozytywnie jak i negatywnie.
- Za najbardziej powiązane z procesem informacja wzajemna wskazała zwykle numer technika (TC), a *pbs* rodzaj wykonywanego zadania (RD).
- Celem zróżnicowania występowania zmiennych w większej skali obliczono występowanie danej kategorii danych w górnych 10% wyników o największej wartości modułu *pbs* i *mi*.

Powiązania zmiennych wejściowych ze zmienną prognozowaną – wejścia kategoryczne

12

Kryterium	BR	TC	TP	RD
moduł <i>pbs</i>	25	26	4	9
<i>mi</i>	39	24	0	1

- Informacja wzajemna wyraźnie ukazała dominujący charakter danych o zespole i w dalszej kolejności techniku wykonującym zadanie.
- Dla pbs wpływ ludzki (BR, TC) w porównaniu do wpływu cech zadania (TP, RD) rozkładał się w proporcji 80 do 20 %

Wnioski z analizy statystycznej

13

- Całościowo, jako zmienne najbardziej powiązane z procesem można przestrankowo określić średni czasy wykonywania danego rodzaju i typu zadania (ave).
- W dalszej kolejności na czas wykonania zadania wpływ mają zmienne katagoryczne, a głównie technik i brygada wykonująca zadanie
- Oprócz średniego czasu najwyższą co do modułu korelacją R charakteryzowała się liczba umiejętności wymagana do wykonania zadania (NS). Drugą co do wielkości modułu tej korelacji była gęstość zaludnienia. W dalszej kolejności korelowały zmienne związane z położeniem lokalizacji zadania.
- Ponieważ wspomniane zmienne mają charakter przestrankowy, w dalszych badaniach rozważono również kilka parametrów które nie zostały wprost wykazane jako powiązane z procesem, ale mają fizyczne uzasadnienie co do wpływu na analizowany proces parkowania.



Pytania badawcze dot. danych prognostycznych

14

- Jak bardzo można zredukować liczbę wejść bez widocznej utraty jakości?
- Czy zgrubne dane położenia pochodzące z siatek prognoz pogody mogą poprawiać dokładność prognoz?
- Który z wariantów zapisu informacji o położeniu jest najkorzystniejszy?

Prognozy czasu wykonania zadania

- Jako metodę naiwną zastosowano zmienną ave - średnią czasową po dniu tygodnia, godzinie i kwadransie
- Jako metody prognostyczne zastosowano drzewa decyzyjne wzmacniane gradientowo (XGBT), lasy losowe (RF) oraz sztuczną sieć neuronową typu MLP
- Całościowo pozyskano ok. 8000 rekordów
- Jako miarę błędu prognoz zastosowano NRMSE [%]
- Dane podzielono następująco:
 - Losowe 20% próbek oznaczono jako dane testowe
 - Pozostałe dane podzielono losowo 5+1 razy na dane treningowe i walidacyjne w proporcji 80% - 20%. Pierwsze 5 podziałów służyło walidacji krzyżowej, a ostatni podział ocenie zdolności generalizacji modeli.

15



Kombinacje danych wejściowych

16

Zmienna	Nr kombinacji zmiennych wejścia				
	1	2	3	4	5
DG	1	1	1	1	1
SG	1	1	0	0	1
PU	1	1	0	0	1
NS	1	1	1	1	1
IWD	1	1	0	0	0
H	1	1	0	0	0
WID1	1	0	0	1	1
WID2	1	0	0	1	1
WR1	1	0	0	0	0
WC1	1	0	1	1	1
WR2	1	0	0	0	0
WC2	1	0	0	0	0
MN1	1	0	1	1	1
MN2	1	0	1	1	1
pd	1	1	0	0	1
BR	1	1	1	1	1
TC	1	1	1	1	1
TP	1	1	0	1	1
RD	1	1	0	1	1
ave	1	1	1	1	1



Modele prognostyczne: hiperparametry

17

- XGBT – 100 kombinacji hiperparametrów na kombinację zmiennych wejścia
 - Liczba drzew 50-100
 - Maksymalna głębokość drzew 2-6
 - Współczynnik uczenia 0,1 – 0,25
- RF – 300 kombinacji hiperparametrów na kombinację zmiennych wejścia
 - Liczba drzew 50-400
 - Maksymalna głębokość drzew 3-30
 - Liczba zmiennych brana pod uwagę podczas poszukiwania rozdziałów w drzewie 1-10
- MLP – 240 kombinacji hiperparametrów na kombinację zmiennych wejścia
 - Liczba neuronów w pierwszej warstwie ukrytej: 3 do 15
 - Liczba neuronów w drugiej warstwie ukrytej: 0 do 6
 - Funkcja aktywacji warstwy ukrytej: tanh/relu
 - Liczba iteracji: 200-500
 - Cierpliwość iteracyjna 10-50 iteracji



Wyniki prognoz

XGBT		NRMSE [%]	
Kombinacja	trening	walid.	test
1	14,913	16,931	16,643
2	14,968	16,849	16,485
3	15,747	16,936	16,615
4	16,261	16,848	16,709
5	15,472	16,916	16,561
metoda naiwna	17,614	17,485	17,403

RF		NRMSE [%]	
Kombinacja	trening	walid.	test
1	12,410	16,960	16,647
2	14,326	16,883	16,741
3	14,935	17,274	16,829
4	13,850	17,044	16,784
5	13,210	16,949	16,679

MLP		NRMSE [%]	
Kombinacja	trening	walid.	test
1	15,899	16,686	16,450
2	15,723	16,970	16,591
3	15,832	17,268	16,877
4	15,907	16,926	16,639
5	15,856	16,761	16,420

18



Wnioski do prognoz – odpowiedź na pytania badawcze

19

- Wynik prognoz zależy od sparowania kombinacji z metodą prognoz. Ogólnie, najlepsze wyniki uzyskało MLP na najbardziej złożonym zbiorze wejść (k1). Na drugim miejscu były lasy losowe uczone na n kombinacji z 12 wejściami (k2).
- Wariant najprostszy uzyskiwał najgorsze wyniki niezależnie od metody.
- Dodani więcej zgrubnej informacji o położeniu poprawiało wynik (k4 lepsze niż k3, k5 zwykle lepsze niż k4). Najbardziej efekt można było zaobserwować dla MLP.
- Ponieważ najlepszy RF ma wyniki zbliżone do najlepszego MLP a wymaga znacznie mniej wejść, może stanowić model backupowy.

Wnioski do prognoz – wnioski ogólne

20

- Zarówno MLP jak i XGBT uzyskiwały najlepsze wyniki na prostych strukturach informacyjnych i szybkim uczeniu (płytkie sieci z szybkim przerywaniem uczenia przy braku poprawy, płytkie drzewa, wysoki współczynnik uczenia). Dla odmiany lasy losowe wyraźnie preferowały głębokie, złożone struktury, z dużą liczbą głębokich drzew i cech rozpatrywanych przy podziale. Można to interpretować jako próbę rozwiązania przez metodę silnego czynnika losowego w danych.
- Przedstawione średnie błędy względne prognoz przekładają się na ok. 5-minutowy średni błąd oszacowania czasu wykonania zadania przez pracowników. Taki wynik w środowisku, w którym istnieje wiele interakcji klient-pracownik wydaje się akceptowalnym. W całym procesie biznesowym istnieją również czasy związane z czasem dojazdu do klienta które powinny zostać dodatkowo uwzględnione w procesie optymalizacji

Podziękowania

21

Prezentacja bazowała na wynikach projektu „System automatycznego przydzielania zadań w warunkach dużej zmienności wykorzystujący mechanizmy sztucznej inteligencji i optymalizacji geoprzestrzennej”, finansowanego ze środków NCBiR w ramach programu „Inteligentny Rozwój 2014-2020”.

Dziękuję za uwagę