

Podstawy Przetwarzania Danych

Wykład 5: Selekcja cech

dr inż. Marcin Luckner
mluckner@mini.pw.edu.pl

Wydział Matematyki i Nauk Informatycznych

Wersja 1.1
5 marca 2021

Projekt „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca” współfinansowany jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Zadanie 10 pn. „Modyfikacja programów studiów na kierunkach prowadzonych przez Wydział Matematyki i Nauk Informatycznych”, realizowane w ramach projektu „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca”, współfinansowanego jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Selekcja w oparciu o wariancję

- Jeżeli zmienna ma zerową wariancję to nie niesie żadnej wartości informacyjnej.
- Przykładowo, nie może być użyta jako wyznacznik różnych klas w procesie klasyfikacji.
- Można rozważyć, czy nie należy także usuwać zmiennych o niezerowej, ale niskiej wariancji.

Klasyfikacja pająków



Mając za zadanie klasyfikację gatunków pająków możemy więc pominąć zmienną opisującą liczbę odnóży.

Rysunek 1: Przykłady pająków [Wikipedia, 2019]

Mała wariancja

- Usuwanie zmiennych o niezerowej wariancji wymaga dodatkowej analizy.
- W przypadku małej wariancji istotny jest jej rozkład między klasami.
- Jeżeli zmienna ma niską, najlepiej zerową, wariancję wewnątrz klasy, a niezerową wariancję ogólną, to będzie dobrym dyskryminatorem.
- Dla danej zmiennej i wyliczamy iloraz sumy wariancji wewnątrzklasowej i wariancji nad całym zbiorem danych.

$$\frac{\sum_{c \in C} \sigma_{i,c}^2}{\sigma_{i,B}^2}$$

- Mniejsza wartość wskazuje na duży rozdział między klasami i małą wariancję wewnętrzną.

Klasyfikacja rodzin zwierząt



Rysunek 2: Przedstawiciele trzech rodzin zwierząt
[Wydawnictwo Rebel, 2019]

- Dokonać klasyfikacji trzech rodzin zwierząt
 - ńasicowate, kotowate, psowate
- Zmienne objaśniające
 - ciężar, długość, długość ogona
 - Podano minimalną i maksymalną wartość dla każdego gatunku

Wpływ przetwarzania danych na wariancję

Tabela 1: Wariancja wyliczona dla cech bez normalizacji

ciężar		długość		ogon	
min	max	min	max	min	max
962.89	4294.63	1044.59	2547.30	375.63	547.05

- Wariancja wyliczona dla zmiennych bez normalizacji utrudnia porównanie cech.

Tabela 2: Wariancja wyliczona dla cech po normalizacji

ciężar		długość		ogon	
min	max	min	max	min	max
0.043	0.055	0.063	0.054	0.063	0.079

- Po normalizacji min-max widzimy, że ciężar minimalny charakteryzuje się najmniejszą wariancją.

Wyliczenie wariancji w podziale na klasy

Tabela 3: Wariancja w podziale na klasy

	ciężar		długość		ogon	
	min	max	min	max	min	max
całość	0.04	0.05	0.06	0.05	0.06	0.08
psowate	0.00	0.01	0.03	0.02	0.01	0.01
łasicowate	0.00	0.00	0.03	0.02	0.02	0.03
kotowate	0.10	0.12	0.07	0.08	0.09	0.11

- Widzimy, że dwie klasy charakteryzują się małą wariancją wewnętrzną w porównaniu do trzeciej klasy

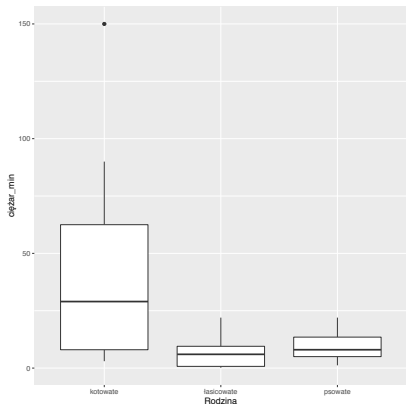
Wyliczenie współczynników

Tabela 4: Stosunek sumy wariancji wewnętrznej do wariancji

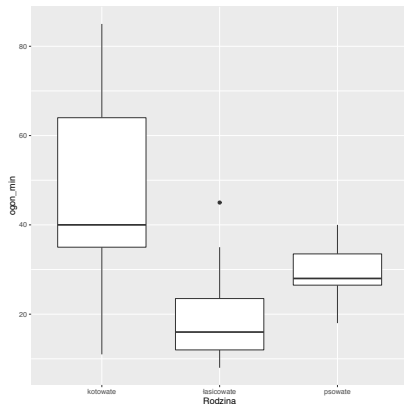
ciężar		długość		ogon	
min	max	min	max	min	max
2.32	2.34	1.96	2.23	1.85	1.9

- Wyliczenie stosunku sumy wariancji wewnątrz klasowej do ogólnej wariancji pokazuje, że minimalny ciężar nie odstaje od miar innych klas i nie ma powodu, aby usuwać tę zmienną.
- Dodatkowo widzimy, że najkorzystniej wypada zmienna minimalna długość ogona.

Klasyfikacja rodzin zwierząt



Rysunek 3: Rozkład zmiennej minimalny ciężar ($e = 2.32$) między klasami



Rysunek 4: Rozkład zmiennej minimalna długość ogona ($e = 1.85$) między klasami

Korelacja ze zmienną objaśnianą

- Naturalną miarą służącą do pozytywnej selekcji, jest korelacja zmiennych ze zmienną objaśnianą.
- W przypadku zadania regresji, test Fishera wskazuje jak silna jest liniowa regresja między zmiennymi, a celem.
- W przypadku zadania klasyfikacji, test ANOVA pokazuje korelację między zmiennymi, a celem.
- Jeżeli problem klasyfikacyjny jest opisany poprzez jakościowe dane nieujemne, można zastosować test χ^2

Test Fishera (F-test)

- W teście stawiamy hipotezę, że wariancje dwóch rozkładów są sobie równe $H_0 : \sigma_T^2 = \sigma_V^2$.
- Test działa poprawnie na zmiennych o rozkładzie normalnym.
- Statystyka F-testu może być wyliczona jako stosunek dwóch wariancji S_T^2 i S_V^2

$$F = \frac{S_T^2}{S_V^2}$$

ANOVA

- W teście stawiamy hipotezę, że średnie dla rozkładu w każdej z n klas są takie same $H_0 : \mu_1 = \mu_2 = \dots = \mu_n$
- Wyliczenie statystyki opiera się na następujących krokach
 - Wyliczenie wariancji wewnętrznej próbki S^2_{Within}
 - Wyliczenie średniej dla każdej z klas
 - Wyliczenie wariancji pomiędzy średnimi $S^2_{Between}$
- Statystyka może być wyliczona jako stosunek dwóch wariancji

$$F = \frac{S^2_{Between}}{S^2_{Within}}$$

Test HSD Tukeya

- Test HSD Tukeya (*Tukey's Honestly Significant Difference test*) pozwala nam określić, bazując na wynikach z analizy wariancji, czy zmienna różni się znacząco między klasami.
- Test wykorzystuje statystykę

$$q_s = \frac{\mu_A - \mu_B}{SE}$$

gdzie μ_A jest większą ze średnich μ_A, μ_B , a SE jest błędem standardowym sumy średnich.

- Uzyskana wartość q_s jest porównywana do krytycznej wartości q_α wynikającej z rozkładu t-Studenta.

χ^2

- Jeżeli zmienne przedstawiają np. zliczenia wystąpienia danego zdarzenia w klasie to możemy zastosować test χ^2 .
- Test bada hipotezę zerową, że kolumny i wiersze tabeli są niezależne.
- Dla każdej komórki tabeli wyliczamy wartość oczekiwaną jako sumę wartości dla wyznaczanego przez nią rzędu i kolumny w stosunku do wszystkich wartości

$$e = \frac{\sum_R \sum_C}{\Sigma}$$

- Statystyka testu jest wyliczana jako

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

gdzie o wartość obserwowana, e wartość oczekiwana

Stopnie swobody

- Liczba stopni swobody to liczba niezależnych zmiennych losowych, które wpływają na wynik.

Tabela 5: Stopnie swobody

	kobiety	mężczyźni	razem
wolę psy	X_1	X_2	100
wolę koty	X_3	X_4	100
razem	100	100	200

- Ile stopni swobody ma test χ^2 dla powyższej tabeli?
- Jeden, bo ustalenie dowolnej zmiennej ustala wszystkie pozostałe wartości.
- Liczba stopni swobody wpływa na poziom istotności testów, ale w statystyce komputerowej zazwyczaj używa się p wartości.

Prawdopodobieństwo testowe

- Wszystkie testy zwracają wartość p-value, czyli prawdopodobieństwo testowe.
- Można ją interpretować jako prawdopodobieństwo uzyskania obserwowanego wyniku jeśli hipoteza zerowa faktycznie jest prawdziwa.
- Wartość p niższa od krytycznego poziomu oznacza, że możemy odrzucić hipotezę zerową.

Przykład zastosowania F-testu

- Sprawdźmy, jaka zmienna najlepiej będzie wyjaśniać maksymalny ciężar zwierzęcia.

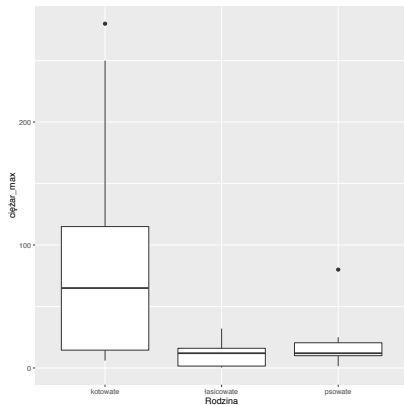
Tabela 6: Wyniki testu zależności z maksymalnym ciężarem

zmienna objaśniająca	F	p
ciężar_min	1.28	0.49
ciężar_max	1.00	1.00
długość_min	0.87	0.71
długość_max	1.02	0.97
ogon_min	0.87	0.69
ogon_max	0.69	0.30

- Wariancja długości maksymalnej najlepiej odpowiada ciężarowi maksymalnemu

Przykład zastosowania testu HSD Tukeya

- Przeprowadzony test HSD Tukeya na zmiennej ciężar maksymalny dał następujące wyniki
 - łasicowate-kotowate:
 $p = 0.014$
 - psowate-kotowate:
 $p = 0.030$
 - psowate-łasicowate:
 $p = 0.945$
- Które klasy są dobrze separowane przez tę zmienną?



Rysunek 5: Rozkład zmiennej ciężar maksymalny między klasami

Regresja linowa

- W zadaniu regresji liniowej dopasowujemy parametry (β_i, α) , aby przy pomocy zmiennych x_i wyjaśnić wartość y

$$y = \sum_{i=1}^n \beta_i x_i + \alpha$$

- Liczba niezerowych współczynników β_i oznacza ile zmiennych będzie użytych w budowanym modelu.

Wpływ liczby zmiennych na regresję

- Duża liczba zmiennych użytych do budowy modelu grozi nadmiernym dopasowaniem, modelu do zbioru treningowego.
- Większa liczba współczynników niezerowych utrudnia zrozumienie modelu i wnioskowanie na jego podstawie.
- Z powyższych powodów dążymy do budowy modeli rzadkich, z małą liczbą współczynników niezerowych.

Regularyzacja

- Regularyzacja jest techniką polegającą na dodawaniu do czynnika błędu kary, która jest zwiększana wraz ze wzrostem wartości beta.
- Następnie próbujemy zminimalizować sumę błędów i kar.
- Technika zabezpiecza przed powstawaniem dużej liczby niezerowych współczynników.

Metody regularyzacji

- Ograniczenie wartości współczynników modelu regresyjnego można przeprowadzać na różne sposoby.
 - Ograniczenie normy L_2 wektora współczynników - regresja grzbietowa.
 - Ograniczenie normy L_1 wektora współczynników - regresja LASSO.
 - Zastosowanie mieszaniny norm - sieci elastyczne.

Normy

Przypomnienie

- Norma L_1

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$

- Norma L_2

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$

Regresja grzbietowa

- Dla regresji grzbietowej wzór na estymator współczynników ma postać

$$\hat{\beta}_{ridge} = \operatorname{argmin}_{(\alpha, \beta)} (\alpha + \mathbf{X}\beta - y)^T (\alpha + \mathbf{X}\beta - y) + \lambda \beta^T \beta$$

- gdzie λ jest współczynnikiem funkcji kary regularyzacji

Regresja LASSO

- LASSO (least absolute shrinkage and selection operator).
- Dla regresji lasso wzór na estymator współczynników ma postać

$$\hat{\beta}_{lasso} = \operatorname{argmin}_{(\alpha, \beta)} \frac{1}{N} \|\alpha + \mathbf{X}\beta - y\|_2^2 + \lambda \|\beta\|_1$$

- gdzie λ jest współczynnikiem funkcji kary regularyzacji, a N wielkością próbki uczącej.
- Regresja LASSO pozwala uzyskać $\beta_i = 0$.

Sieci elastyczne

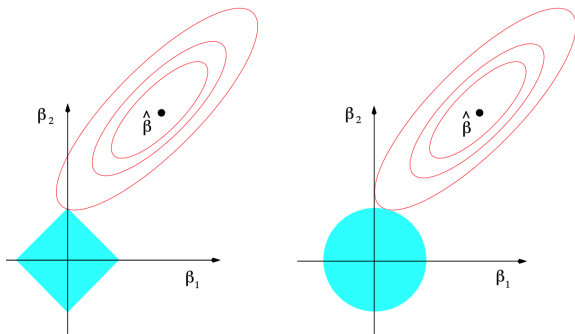
- Dla regresji siecią elastyczną wzór na estymator współczynników ma postać

$$\hat{\beta}_{net} = \operatorname{argmin}_{(\alpha, \beta)} \frac{1}{N} \|\alpha + \mathbf{X}\beta - y\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

- gdzie λ jest współczynnikiem funkcji kary regularyzacji, a N wielkością próbki uczącej.

Selekcja zmiennych

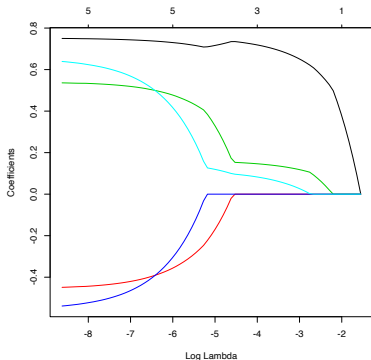
- Zwiększając współczynniki λ dowolnej metody regularyzacji będziemy zmniejszać współczynniki β .
- Jednakże tylko metoda Lasso może, dzięki zastosowaniu normy L_1 , zerować współczynniki.



Rysunek 6: Norma kulista regresji grzbietowej nie osiągnie $\beta_i = 0$ na co pozwala norma kwadratowa regresji Lasso [Hastie et al., 2001]

Estymacja ciężaru

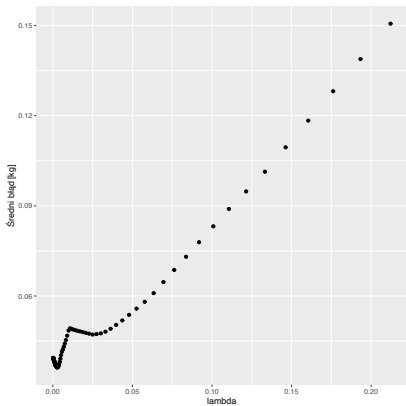
- Używając metody lasso, zbudujemy modele wyjaśniające maksymalny ciężar zwierzęcia.



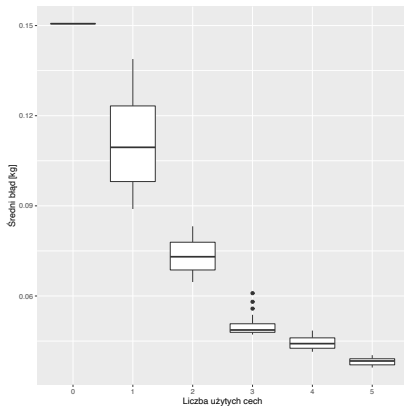
- Istotność zmiennych wynikająca z kolejności zerowania współczynników
 - ciężar_min
 - długość_max
 - ogon_max
 - długość_min
 - ogon_min
- Wykres ilustrujący zmianę współczynników β względem współczynnika λ nazywamy ścieżkami lasso.

Rysunek 7: Zmiana współczynników β ze wzrostem współczynnika λ

Wpływ liczby zmiennych na wyniki

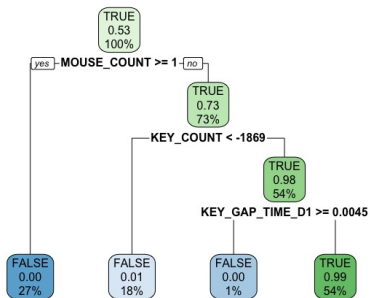


Rysunek 8: Wpływ współczynnika λ na wyniki



Rysunek 9: Wpływ liczby zmiennych na wyniki

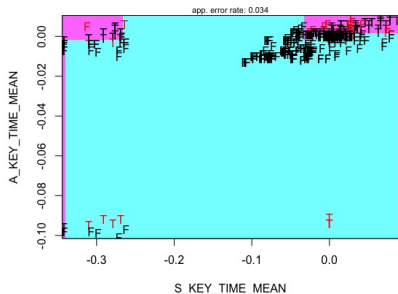
Drzewo decyzyjne



Rysunek 10: Drzewo decyzyjne

- Drzewo decyzyjne, w każdym węźle nie będącym liściem, dzieli analizowany zbiór na dwa podzbiory.
- Podział odbywa się na podstawie zmiennej i progu dla jej wartości.
- Ponieważ celem jest uzyskanie, w liściach, zbiorów jednorodnych, to kryterium podziału musi maksymalizować rozróżnienie zbiorów.

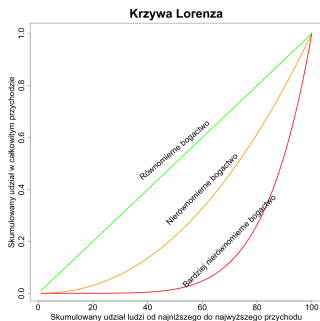
Podział przestrzeni



Rysunek 11: Podział przestrzeni średniego czasu wciśnięcia klawisza alfanumerycznego i specjalnego do rozpoznania czy mamy do czynienia z tym samym użytkownikiem

- Dla każdej zmiennej wybierana jest wartość dzieląca przestrzeń na dwa podzbiory.
- Podział dokonywany jest w taki sposób, aby minimalizować liczbę błędów dopasowania.
- W tym celu stosowane są specjalne miary dyskryminacji.

Krzywa Lorenza i wskaźnik Giniego



- Krzywa Lorenza opisuje nierównomierności podziału globalnego zasobu cechy jednowymiarowego rozkładu zmiennej losowej o wartościach nieujemnych.
- Wskaźnik Giniego jest polem pomiędzy równomiernym podziałem a krzywą Lorenza.
- W ekonometrii służą do liczbowego wyrażania koncentracji kapitału i nierównomierności zarobków.
- W uczeniu maszynowym pokazują jak dobrze zmienna różnicuje zbiór.

Rysunek 12: Przykłady różnych krzywych Lorenza [Economic Theory Blog, 2019]

Wyliczanie współczynnika w węźle

- W każdym węźle m drzewa T wyliczamy współczynnik Ginniego bazując na współczynniku f_i określającym udział klasy w zbiorze

$$I(m) = \sum_{i=1}^g f_i(1-f_i) = \sum_{i=1}^g (f_i - f_i^2) = \sum_{i=1}^g f_i - \sum_{i=1}^g f_i^2 = 1 - \sum_{i=1}^g f_i^2$$

- gdzie g jest liczbą klas w zadaniu klasyfikacyjnym.

Udział klasy w zbiorze

- Współczynnik f_i określa udział klasy i w zbiorze

$$f_i = \frac{n_{m_i}}{n_m}, i = 1, \dots, g$$

- n_m liczba obserwacji w węźle m

$$n_m = \text{card}(S_m)$$

- n_{m_i} liczba obserwacji należących do klasy i

$$n_{m_i} = \text{card}(\{x_k \in S_m : y_k = i\})$$

Wariancja w drzewach CART

- W drzewach CART (*Classification and Regression Tree*) [Maimon and Rokach, 2005] współczynnik Ginniego stosowany jest jako miara wariancji definiowanej dla węzła m i drzewa T jako

$$Q_m(T) = I_G(m).$$

- W przypadku podziału drzewo tworzy dwóch potomków m_L and m_R dla których można wyliczyć łączną wariancję

$$\frac{n_{m_L}}{n_m} Q_{m_L}(T) + \frac{n_{m_R}}{n_m} Q_{m_R}(T).$$

Ocena jakości podziału

- Zmiana różnorodności wynikająca z podziału wężła m na wężły potomne m_L oraz m_R , wynosi:

$$\Delta Q_{m,m_L,m_R} = Q_m(T) - \left(\frac{n_{m_L}}{n_m} Q_{m_L}(T) + \frac{n_{m_R}}{n_m} Q_{m_R}(T) \right)$$

- W celu uzyskania najlepszego podziału chcemy maksymalizować różnicę $\Delta Q_{m,m_L,m_R}$.
- Wtedy, w efekcie wprowadzenia podziału uzyskujemy możliwie niewielką różnorodność w wężłach m_L oraz m_R .
- W idealnym przypadku, w każdym z tych wężłów znajdą się przedstawiciele tylko jednej klasy.

Problem doboru podziału

- Rozpatrzmy jedną cechę X_i przyjmującą wartości X_{ij} . Liczba możliwych podziałów wynosi $\frac{1}{2}2^D - 1 = 2^{D-1} - 1$, gdzie $D = \text{card}(\{X_{ij} \in S_m\})$.
- Liczba podziałów wynika z liczby podzbiorów, jakie możemy wydzielić ze zbioru wartości atrybutu X_i w ramach rekordów w węźle m , czyli zbioru S_m .
- Redukując zadanie przyjęto, iż analizowane są wyłącznie podziały monotoniczne.
- Nowym celem jest znalezienie wartości c takiej, że do węzła m_L trafią rekordy spełniające $X_{ij} < c$, a do węzła m_R pozostałe rekordy spełniające $X_{ij} \geq c$.
- Ogranicza to liczbę analizowanych podziałów względem atrybutu X_i do $D = \text{card}(\{X_{ij} \in S_m\})$, czyli liczby unikalnych wartości atrybutu X_i w S_m .

Ograniczenia metody

- Algorytm ma zastosowanie tylko dla zmiennych liczbowych i uporządkowanych.
- W przypadku zmiennych nominalnych konieczne jest dokonanie analizy wszystkich podzbiorów.
- Z powyższego powodu, algorytmy budowy drzew decyzyjnych nakładają ograniczenie na liczbę wartości cech nominalnych.
- Metoda nie ma też zastosowania dla zadań klasyfikacyjnych z większą liczbą klas.

Wybór cech

- Algorytm wyboru podziału należy powtórzyć dla poszczególnych atrybutów X_i w celu ustalenia najlepszego podziału, czyli wyboru atrybutu X_i i jego wartości c , względem której zostanie wykonany podział.
- Pozwala nam to na uzyskanie informacji na ile dobrze każda z cech dyskryminuje dane.
- Może być to podstawą do selekcji zmiennych.

Określenie istotności cech

- Dla każdej zmiennej f wyliczamy zmianę będącą sumą zmian dokonanych przez tę cechę we wszystkich węzłach gdzie stanowi ona podstawę podziału. $\Delta Q_f = \sum_{m \in M_f} \Delta Q_{m, m_L, m_R}$
 - gdzie M_f jest zbiorem węzłów z podziałem opartym o f
- Następnie ustalamy wpływ cechy na całkowitą zmienność w drzewie

$$I(f_i) = \frac{\Delta Q_{f_i}}{\sum_{k=1}^M \Delta Q_{f_k}}$$

- Wyliczona wartość jest istotnością cechy.

Uwagi do istotności cech

- Istotność cech nie jest wartością znormalizowaną. Można dokonać normalizacji wyliczając

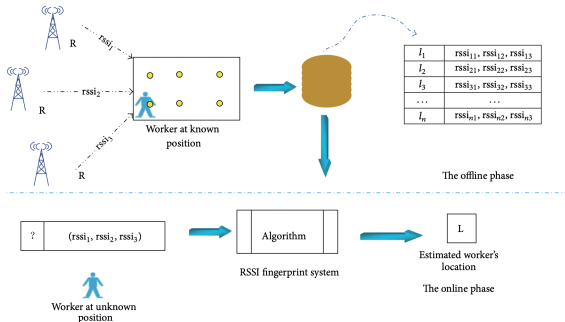
$$\hat{I}(f_i) = \frac{I(f_i)}{\max_k I(f_k)}$$

- Cechy nie występujące w strukturze drzewa mają istotność równą 0.
- Można stosować metodę to wyliczania istotności cech w zadaniach regresji zamieniając indeks Ginniego na błąd średnio kwadratowy.
- Metodę wyliczania istotności cech można uogólnić na lasy losowe wyliczając wartość sumaryczną dla wszystkich drzew [Breiman, 2001].

Selekcja źródeł sygnałów

- Na wydziale MiNI zbudowano system lokalizacji oparty o siłę sygnału Wi-Fi rejestrowaną na telefonie komórkowym [Górak and Luckner, 2018].
- Lokalizacja (x, y) jest wyznaczana na podstawie wektora odczytanych sygnałów.
- Wymiar wektora istotnie wpływa na czas uczenia modelu lokalizacyjnego i powinien być zredukowany.
- Dane użyte do lokalizacji pozyskane zostały metodą *fingerprinting*.

Fingerprinting



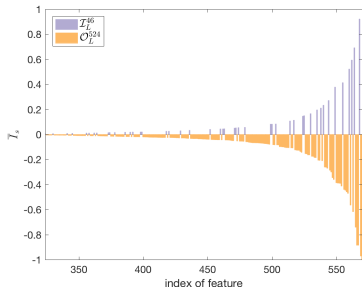
Rysunek 13: Schemat procedury fingerprinting [Lin et al., 2013]

- W metodzie *fingerprinting* zbieramy w ustalonych punktach informacje o sile rejestrowanych sygnałów.
- Następnie na podstawie odczytanych sygnałów staramy się określić lokalizację.

Dane

- Podczas zbierania danych odnotowano 570 unikalnych źródeł sygnałów \mathcal{S} .
- Są nimi punkty dostępowe (*Access Points*) sieci wydziałowej i innych okolicznych sieci.
- Sieć wydziałowa uwzględnia 46 APs \mathcal{I}
- Inne sieci obejmują 524 APs \mathcal{O}

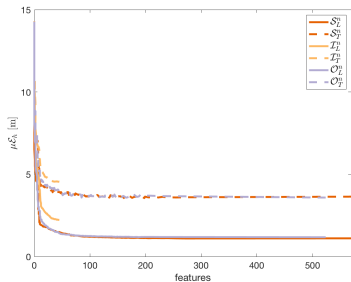
Istotność źródeł sygnałów



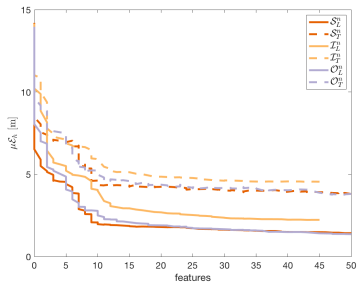
Rysunek 14: Istotność źródeł sygnałów. Dla czytelności zmieniono znak istotności źródeł zewnętrznych.

- Korzystając z lasów losowych estymujących położenie (x, y) wyliczono znormalizowaną istotność cech.
- Wyliczenia pokazały, że zarówno sygnały sieci wydzielowej jak i sygnały z innych sieci mają istotny wpływ na lokalizację.

Wpływ cech na jakość lokalizacji



Rysunek 15: Błąd lokalizacyjny w zależności od liczby użytych cech



Rysunek 16: Przybliżenie najistotniejszych 50 cech

Wybór liczby cech

- Wybór liczby cech wynika z obserwacji redukcji błędu na zbiorze uczącym.
- Pierwsze podejście wybiera liczbę cech oferującą najmniejszy błąd $\mu\mathcal{E}_h$

$$m_h = \arg \min_{n \leq N} (\mu\mathcal{E}_h(\hat{\mathcal{L}}, \mathcal{S}_L^n)).$$

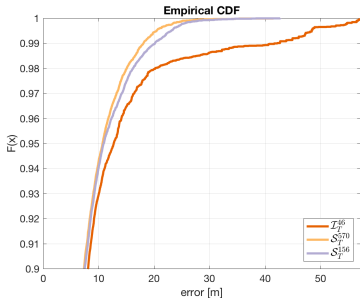
- Jednakże ze względu na malejący charakter funkcji $\mu\mathcal{E}_h(n)$ należy przypuszczać, że wartość m_h uplasuje się blisko wartości maksymalnej N .

Alternatywny wybór liczby cech

- Alternatywne podejście określa rozdzielczość błędu, czyli wartość θ_h , którą uznajemy za nieistotny przyrost błędu.
- Następnie liczbę cech określamy jako minimalną spośród wartości dla których uzyskany błąd nie jest większy o θ_h od błędu minimalnego

$$n_h = \min\{n : \mu\mathcal{E}_h(\hat{\mathcal{L}}, \mathcal{S}_L^n) - \mu\mathcal{E}_h(\hat{\mathcal{L}}, \mathcal{S}_L^{m_h}) < \theta_h\}.$$

Porównanie modeli



Rysunek 17: Błąd lokalizacji dla poszczególnych modeli

- Porównano modele lokalizacyjne działające w oparciu o
 - Pełny zbiór 570 AP.
 - Zbiór oparty o 46 AP sieci wydziału.
 - Zbiór $n_h = 156$ najlepszych cech.
- Stworzony podzbiór daje w 95 procentach pomiarów takie same wyniki jak pełny zbiór.
- Daje też mniejsze błędy grube niż model oparty na sieci wydziałowej.

Bibliografia I

[Breiman, 2001] Breiman, L. (2001).

Random forests.

Machine Learning, 45(1):5–32.

[Economic Theory Blog, 2019] Economic Theory Blog (2019).

Kod wyliczający krzywą lorenza.

[Górak and Luckner, 2018] Górak, R. and Luckner, M. (2018).

Automatic detection of missing access points in indoor positioning system.

Sensors, 18(11):3595.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001).

The Elements of Statistical Learning.

Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Bibliografia II

[Lin et al., 2013] Lin, P., Li, Q., Fan, Q., and Gao, X. (2013).

Real-time monitoring system for workers' behaviour analysis on a large-dam construction site.

International Journal of Distributed Sensor Networks, 9(10):509423.

[Maimon and Rokach, 2005] Maimon, O. and Rokach, L., editors (2005).

Data Mining and Knowledge Discovery Handbook. A Complete Guide for Practitioners and Researchers.

Springer, San Francisco, CA, USA.

[Wikipedia, 2019] Wikipedia (2019).

Pająki.

[Wydawnictwo Rebel, 2019] Wydawnictwo Rebel (2019).

Fauna.