

## Case study 23 z przedmiotu „Zaawnsowana Eksploracja Danych” – TPD (listopad 2010)

### Analiza diagnostycznej bazy danych.

(Copyright Jerzy Stefanowski - Instytut Informatyki Politechnika Poznańska; zastrzeżenia dotyczą opisu problemu i ograniczonej dostępności do danych)

#### Cel :

„Case study” powinien prowadzić do odkrycia użytecznych i potencjalnie interesujących regularności z rzeczywistych danych. Należy także dokonaniu interpretacji i oceny znalezionych regularności. Możliwie jest interpretowanie znalezionych regularności jako form reprezentacji wiedzy odkrytych w bazie danych. Problem dotyczy analizy stanu technicznego autobusów używanych przez jedno z przedsiębiorstw w Polsce. Z metodologicznego punktu widzenia sugeruje się wykorzystywanie poznanych metod eksploracji danych i odkrywania wiedzy - zarówno statystycznych jak i wywodzących się ze sztucznej inteligencji.

Podsumowaniem analizy powinien być krótki raport zawierający listę najbardziej interesujących regularności oraz komentarz lub ich interpretację - raport ten powinien być tworzony na bieżąco podczas zajęć.

#### Wprowadzenie :

Pewne przedsiębiorstwo wydobywcze użytkuje systemy przenośników taśmowych. Jednym z kluczowych elementów takiego urządzenia są rolki, po których przesuwa się taśma z urobkiem. Rolki osadzone są na łożyskach, które mogą stopniowo ulegać uszkodzeniu oraz zatarciu. Stan niesprawności łożyska, które w praktyce zablokuje ruch obrotowy pewnej liczby rolek prowadzi do zniszczenia taśmy a w skrajnym przypadku także doprowadzić do zbyt wysokiego wzrostu temperatury taśmy, co może spowodować jej zapalenie. Dlatego ocena stanu technicznego łożysk ma znaczenie praktyczne. W ogólności obiektywna ocena jest związana z wymontowaniem łożyska z urządzenia i zbadaniu stanu jego powierzchni, uszkodzeń i ew. wad materiału w warunkach specjalizowanego laboratorium diagnostycznego.

Z uwagi na to, że takie postępowanie diagnostyczne jest uciążliwe i kosztowne często stosowaną praktyką jest stosowanie metod wibro-akustycznej oceny. Taka procedura diagnostyczna wykorzystuje czujniki drgań (najczęściej montowane na obudowach urządzeń) a także specjalizowane mikrofony i mierniki do pomiaru poziomu hałasu. Urządzenie obrotowe z uszkodzonymi elementami tocznymi łożyska mają podwyższony poziom drgań i emitują odpowiednie dźwięki.

Takie pomiary są podstawą zdefiniowania tzw. symptomów stanu technicznego. To znaczy, jeśli symptomy są dobrze wybrane dla określonego zastosowania, to ich wartości będą podstawą do decyzji nt. całościowej oceny stanu technicznego urządzenia. Na ogół przekroczenie pewnych poziomów progowych pojedynczych symptomów lub kombinacji zakresu wartości kilku symptomów równocześnie może świadczyć, że urządzenie jest już w złym stanie technicznym i nie powinno być dalej użytkowane.

W niniejszym studium przypadku dokonywano pomiarów wibroakustycznych dla rzeczywistych przenośników taśmowych.

Zarejestrowane sygnały są przetwarzane, np. z wykorzystaniem specjalnych metod filtrów pasmowych. W niniejszym studium przypadku rozważa się np. przetwarzanie niskiego i wysokiego zakresu częstotliwości – dla hałasu od 500 do 2000 Hz i od 4 do 16 kHz; analogicznie od 100 do 1000 Hz i od 1 do 11 kHz dla drgań. Oprócz różnych sposobów przetwarzania pomiarów dysponuje się różnymi źródłami – tzn. czujniki i mierniki ustawione są na różnych kierunkach obrotu łożysk (w osi i na promieniu kołowym).

Z uwagi na to, że powyższe badania mają charakter eksperymentu naukowego w celu zbudowanie nowych metod oceny stanu technicznego tego rodzaju łożysk. Dlatego oprócz samych pomiarów „zewnętrznych” pracującego przenośnika wykonano także zbadanie eksperckie samego łożyska i można było podjąć decyzje o rzeczywistym stanie technicznym „obsłuchiwanego” łożyska (sprawne, uszkodzone).

Zebrany zbiór danych obejmuje ponad 110 pomiarów scharakteryzowanych poprzez 11 następujących symptomów:

s\_1 -- poziom hałasu zmierzony dla zakresu częstotliwości 500 -- 2000 Hz [dB],

s\_2 -- poziom hałasu dla zakresu częstotliwości 4 -- 16 kHz in [dB],

Pozostałe symptomy dotyczą poziomu drgań (pomiar tzw przyspieszenia m/s)

s\_3 -- drgania na kierunku osi obrotu i zakresu częstotliwości 0.1 -- 1 kHz,

s\_4 – drgania na kierunku osi zmierzony z tył urządzenia dla zakresu 0.1 – 1 kHz

s\_5 -- drgania na kierunku osi obrotu i zakresu częstotliwości 1 -- 11 kHz,

s\_6 -- drgania na kierunku radialnym pionowym i zakresu częstotliwości 0.1 -- 1 kHz,

s\_7 -- drgania na kierunku radialnym pionowym i zakresu częstotliwości 1 -- 11 kHz,

s\_8 -- drgania na kierunku radialnym poziomym i zakresu częstotliwości 0.1 -- 1 kHz,

s\_9 -- drgania na kierunku radialnym pionowym i zakresu częstotliwości 1 -- 11 kHz,

s\_10 – zaagregowane poziomy drgań w zakresie 0.1 -- 1 kHz

s\_11 -- zaagregowane poziomy drgań w zakresie 1 --11 kHz

Na podstawie innych badań celowo uszkodzonych podobnych łożysk zamontowanych w stanowiskach laboratoryjnych wykonano w przeszłości zdefiniowanie tzw. wartości graniczne symptomów stanu technicznego – mają one na celu identyfikację przedziałów „normalnych” wartości stanu technicznego oraz przedziałów świadczących o pogorszeniu się stanu technicznego [informacja o ich wartościach może być udostępniona do wglądu].

Należy zwrócić uwagę, że symptomy stanu technicznego mogą mieć różną przydatność dla całościowej oceny stanu urządzenia. Niektóre z nich niosą wartościową informację. Inne mają mniejszy wpływ dla ostatecznej oceny. W przypadku prowadzenia badań niektórych urządzeń rozważa się większą liczbę symptomów niż jest to konieczne.

Celem badań diagnostycznych jest:

(a) ocena przydatności diagnostycznych pojedynczych symptomów oraz wartości granicznych (lub samodzielnego określenia takich wartości),

(b) redukcja zbioru symptomów do podzbioru najważniejszych zapewniających satysfakcjonującą ocenę stanu technicznego (najbardziej pożądaną sytuacją jest identyfikacja pojedynczych symptomów, dla których wartości graniczne pozwalają na skuteczną ocenę stanu technicznego),

(c) poszukiwanie zależności między wartościami wybranych symptomów a całościową oceną stanu technicznego,

(d) konstrukcja tzw. klasyfikatora stanu technicznego, czyli wskazań dla podejmowania końcowej decyzji diagnostycznej na podstawie bieżących obserwacji.

W przypadku rozważanego case study dostarczono Tobie gotowy zbiór danych i poproszono o przebadanie w celu odpowiedzenia na powyższe pytania i problemy związane z zagadnieniami diagnozowania stanu technicznego pojazdów. Nie jesteś w stanie zażądać już dokonania dodatkowych pomiarów – musisz starać się odkryć jak najwięcej interesujących elementów w danych, które otrzymałeś.

Twoje wnioski powinny być interpretowane w kategoriach przydatności odkrytej wiedzy z danych (czyli znalezionych regularności, klasyfikatorów itp.) dla bieżącej oceny diagnostycznej podczas badań stanu technicznego pojazdów i podejmowania decyzji o ich dalszej eksploatacji na podstawie wartości wybranych symptomów stanów technicznego.

Ponadto można badać współzależności pomiędzy samymi symptomami stany technicznego, wyboru najważniejszych symptomów, czy dokonywać oceny klasyfikacyjnej tworzonych klasyfikatorów. Uwaga stworzenie klasyfikatora nie jest jedynym i najważniejszym celem tego studium. Lecz jeśli będziesz to wykonywał pamiętaj, że rozpoznawanie klasy uszkodzonej jest ważniejsze niż sprawnej, ale w przypadku obu oczekuje się dokładności powyżej 90%.

Dane są dostępne jako plik ASCII o formacie tekstowym w zapisie jeden wiersz zawierający opis jednego pojazdu za pomocą powyższych symptomów i atrybutów decyzyjnych. Plik może zawierać błędy wykonane przez osobę wprowadzającą dane.

## **Uwagi metodyczne :**

Powinieneś pamiętać, iż nie masz wpływu na rozmiar dostępnych danych, nie możesz oczekiwać dostarczenia dodatkowych opisów przypadków; wszystko zostało to wykonane przed Twoim udziałem w studium badawczym - nie możesz żądać dodatkowych czynności pozyskiwania informacji.

Jest to typowa eksploracja dostępnych danych.

Jeśli potrafisz ocenić jakość otrzymanych danych możesz dokonywać przeskalowań lub przededefiniować atrybutów (w oparciu o określone metody).

Konieczne jest badanie jakości dostarczonych danych (mogą być zbierane przez osoby, które nie znają własności Twoich metod); Ponadto podczas przygotowywania danych mogły wystąpić pomyłki wprowadzania pomiarów

- warto tutaj skontrolować, czy nie występują pomiary o wartościach mocno odległych od innych typowych wartości (tak zwane obserwacje odstające),
- sprawdź także, czy nie ma błędnych lub nieznanymi wartości niektórych atrybutów = w wyniku badań diagnostycznych starano się dokonać wszystkich pomiarów.

Metodycznie potraktuj problem jako odpowiednie zadania tzw. uczenia nadzorowanego (dany jest opis przykładów za pomocą atrybutów jak i klasyfikacja). Rozważaj także właściwe metody statystyczne

Inne uwagi metodyczne:

- Interesujące jest badanie wzajemnych współzależności tkwiących w danych;
- Analizuj każdy atrybut decyzyjny oddzielnie a później ich właściwe podzbiory.
- Warto stosować więcej niż jedną metodę eksploracji danych (ukierunkowanych na różne formy wiedzy i różne ich reprezentacje)
- Uwaga z powodu silnego eksploatowania na dotychczasowych zajęciach technik budowy klasyfikatorów, zwracam uwagę, że automatyczna klasyfikacja jest tylko jedną z miar oceny; Jeśli chcesz rozważyć budowę klasyfikatorów to pamiętaj także, że użytkownicy nie skupiają się wyłącznie na globalnej trafności klasyfikacji, lecz ważniejsza jest dla nich trafność w poszczególnych klasach, w szczególności urządzeń w gorszym stanie (analiza „confusion matrix” jest bardzo pożądana).

Dostępne oprogramowanie:

Oprogramowanie Statystyczne Statistica, R, itp.

Środowiska jak WEKA, RapidMiner, i inne.

System indukcji drzew decyzyjnych C4.5; System indukcji reguł decyzyjnych CN2; System oparty na teorii zbiorów przybliżonych ROSE

Inne wg. uznania

### Proponowany przebieg zajęć:

1. W pierwszej części "Case Study" prowadzący omawia problemy eksploracji danych w diagnostyce technicznej; następnie krótko charakteryzuje poniższy problem.
2. Uczestnicy zapoznają się z niniejszym tekstem i danymi, starając się określić cel i zakres swojej analizy oraz zidentyfikować podstawowe właściwości danych; Ponadto starają się przygotować przed zajęciami plan swoich zamierzeń. Powinien on obejmować zakładany cel analizy; listę problemów diagnostycznych, które zamierza się rozwiązać i powiązanych z tym interesujących typów regularności, których zamierza się poszukiwać w danych. Należy także określić podstawowe właściwości danych i listę metod, które się zamierza użyć. Uczestnicy są podzieleni na zespoły i wewnątrz zespołów dyskutują na temat problemu oraz przygotowują propozycje rozwiązania. Zespoły uczestników powinny **realizować samodzielnie analizę i starają się na bieżąco prowadzić raport z wykonywanych czynności i uzyskanych wyników**.
3. Zespoły w dyskusji (**druga część spotkania**) prezentują propozycje rozwiązania problemu, jak i także omawiają wyniki samodzielnej analizy – zalecanie wykorzystanie tworzone **raportu - sprawozdania** z dotychczasowej analizy i otrzymanych rezultatów. Na końcu zajęć prowadzący omawia się wspólnie otrzymane wyniki, prowadzący prezentuje także inne znane wyniki dla tego przypadku. Prowadzi się dyskusję na ich temat i podsumowuje całość zajęć.
4. Zespoły dostarczają sprawozdanie końcowe. Sprawozdanie, aktywność na zajęciach oraz obecność na tych zajęciach jest podstawą otrzymania oceny za przebieg tego case study.

---

### Dodatek B: Problemy diagnostyki technicznej – podstawowe informacje

Przedstawimy kilka uwag na temat celów diagnostyki technicznej i zadań, w których rozwiązaniu mogą być przydatne metody odkrywania wiedzy oraz eksploracji danych. Efektywna eksploatacja obiektów mechanicznych czy maszyn w procesach przemysłowych wymaga wiarygodnej informacji o

ich stanie technicznym. Informacja ta jest często rozszerzona o predykcję zmiany ich stanu technicznego. Ma to szczególne znaczenie w przypadku maszyn o krytycznym znaczeniu w procesie, wymagających specjalnych skomputeryzowanych systemów nadzoru (monitorowania) – np. turbiny w elektrowniach. W przypadku maszyn o mniejszym znaczeniu dokonuje się okresowych pomiarów za pomocą urządzeń przenośnych. Ocena stanu technicznego wykonywana jest na podstawie tzw. symptomów, czyli wielkości, które zmieniają się wraz z pogarszaniem się stanu technicznego maszyny. Przykładami symptomów są poziom drgań reprezentatywnych punktów obiektu, poziom hałasu w otoczeniu pracującej maszyny, temperatura, ciśnienie czy wzajemna pozycja części składowych maszyny. Z diagnostycznego punktu widzenia poszczególne symptomy mają różną przydatność dla konkretnego problemu. Dla pojedynczych symptomów mierzonych na skalach liczbowych definiuje się czasami tzw. wartości graniczne. Dzielą one dziedzinę symptomu na pewne podprzedziały, które mogą być interpretowane w kategoriach normalnych warunków pracy i ich stopniowego pogarszania się. Są one definiowane na podstawie zaleceń producentów i wskazań literaturowych, lecz praktyczne badania nie zawsze potwierdzają ich użyteczność. Ponadto w postępowaniu diagnostycznym mamy zwykle dostęp do dużych zbiorów danych. Dane gromadzone podczas obserwacji obiektów diagnozowania mogą być także niekompletne, sprzeczne, niedokładne czy obarczone niepewnością. Także sama wiedza diagnostyczna ma często charakter przybliżony. Dyskusja tych zagadnień przedstawiona jest w [Moczulski 97]. Z punktu widzenia zastosowań rozważanych w tym rozdziale istotnymi zadaniami badań diagnostycznych są :

1. ocena zdolności diagnostycznej poszczególnych symptomów,
2. ocena różnych metod definiowania wartości granicznych dla tych symptomów,
3. selekcja podzbioru symptomów zapewniających satysfakcjonującą ocenę stanu technicznego,
4. stworzenie klasyfikatora stanu technicznego.

Moczulski W., Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn. Zeszyty Naukowe Politechniki Śląskiej. Monografie, Mechanika, z. 130, Gliwice 1997.