

Rozdział 4

Moduł prognostyczny

Adam Gudyś, Michał Kozielski, Marek Sikora, Łukasz Wróbel

4.1 Wstęp

W ostatnim czasie obserwowany jest intensywny wzrost możliwości i popularności środowisk analitycznych zawierających narzędzia eksploracji danych. Środowiska tego typu najczęściej znajdują zastosowanie w marketingu, ubezpieczeniach, bankowości i finansach, handlu (zwłaszcza elektronicznym) oraz opiece zdrowotnej. Rzadziej metody eksploracji danych używane są do analizowania i nadzorowania procesów przemysłowych [4.3].

Istnieją przykłady metodyk pokazujących jak stosować metody zaawansowanej analizy do danych przemysłowych (np. Six Sigma [4.8], CRISP-DM [4.6]). Istnieją również środowiska dedykowane do analizy danych przemysłowych (np. AVL Concerto [4.1]) oraz moduły w uniwersalnych środowiskach analitycznych (np. moduł Monitoring and Alerting System w środowisku Statistica [4.2]). Jednakże rzadkością są systemy wspomagania decyzji bazujące na analizie danych pochodzących z procesów (dostarczanych w sposób ciągły). Jedną z przyczyn takiego stanu rzeczy może być fakt, że dane pochodzące z przemysłowych systemów monitorowania są przeważnie danymi temporalnymi i mają postać wielowymiarowych strumieni danych, a analiza takich danych, tworzenie modeli predykcyjnych i ich stosowanie w trybie ciągłym (ang. on-line) stanowi pewne wyzwanie.

Jednocześnie, zastosowanie metod zaawansowanej analizy danych do zadań predykcyjnych związanych z procesem przemysłowym może przynieść wymierne korzyści. Przykładowo, akwizycji danych w kopalni węgla kamiennego dokonuje się głównie w celach wizualizacji stanu procesu. Tymczasem zastosowanie wiedzy dziedzinowej w połączeniu z analizą danych historycznych może znacząco usprawnić pracę operatora na stanowisku dyspozytorskim. Na przykład, krótkoterminowe prognozy stężenia metanu połączone z informacją na temat położenia i intensywności prac kombajnu ścianowego umożliwiają zapobieżenie awaryjnym wyłączeniom energii i podtrzymanie ciągłości pracy [4.4, 4.5, 4.7].

W niniejszym rozdziale przedstawiony jest moduł prognostyczny systemu DISESOR umożliwiający realizację prognoz (np. takich jak wymieniona w poprzednim akapicie). Kolejne sekcje rozdziału opisują architekturę oraz wybrane aspekty funkcjonalności tego modułu.

4.2 Architektura

Moduł prognostyczny opiera swoje działanie na tzw. usługach prognostycznych. Usługa prognostyczna jest to usługa sieciowa (ang. webservice), której zadaniem jest zdalna predykcja wartości zmiennej (ciągłej bądź dyskretnej) w oparciu o wektor danych wejściowych. Z usługą prognostyczną nierozdzielnie związany jest model (regresyjny bądź klasyfikacyjny), w oparciu o który realizowana jest prognoza. Moduł prognostyczny bazuje na gotowych modelach predykcyjnych dostarczanych np. z modułu analitycznego. Podstawowy scenariusz użycia usługi prognostycznej jest następujący:

1. Klient wysyła do usługi żądanie wykonania prognozy wraz z wektorem atrybutów warunkowych oraz znacznikiem czasu.
2. Usługa dokonuje prognozy podając wektor atrybutów i ich wartości na wejście modelu.
3. Usługa zapisuje wynik prognozy do bazy danych.

Przedmiotem prognozy jest zazwyczaj przyszła wartość jednego z atrybutów. W związku z tym, po pewnym czasie (zależnym od ustalonego horyzontu prognozy) znana jest jego wartość rzeczywista. Dlatego usługa prognostyczna daje możliwość zapisania w bazie wartości faktycznej prognozowanego parametru (m.in. w celu monitorowania jakości prognoz). Scenariusz użycia modułu przedstawia się następująco:

1. Klient wysyła do usługi wartość rzeczywistą zmiennej wraz ze znacznikiem czasu.
2. Usługa uzupełnia na podstawie znacznika czasu odpowiedni rekord w bazie danych.

Każda usługa prognostyczna może mieć przyporządkowaną dowolną liczbę reguł monitorowania. Reguły te umożliwiają śledzenie jakości prognoz i podjęcia kroków zaradczych w momencie, gdy owa jakość spada poniżej akceptowalnego poziomu. Moduł prognostyczny zawiera trzy aplikacje:

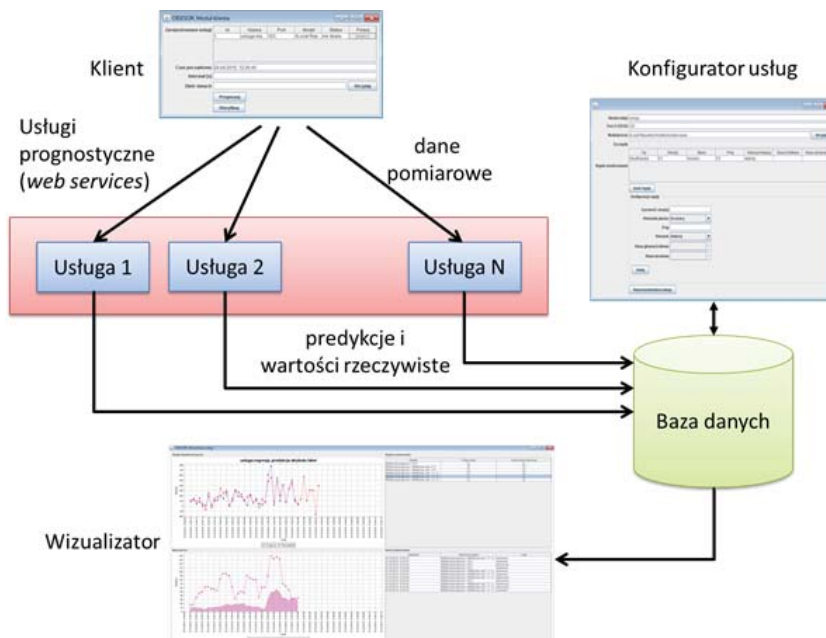
- Konfigurator usług,
- Wizualizator,
- Klient.

Architekturę modułu prognostycznego przedstawiono na rysunku 4.1.

4.3 Konfigurator usługi prognozowania

Po uruchomieniu konfiguratora usługi prognozowania wyświetlana jest lista istniejących usług prognostycznych wraz z podstawowymi informacjami (rysunek 4.2). Konfigurator umożliwia:

- definiowanie nowych usług prognostycznych,



Rys. 4.1. Architektura modułu prognostycznego

- edycję/usuwanie istniejących usług prognostycznych,
- zatrzymywanie/uruchamianie usług prognostycznych.

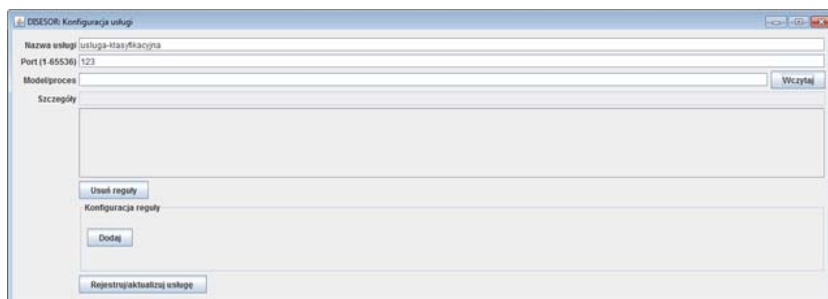
Zarejestrowane usługi	Id	Nazwa	Port	Model	Status	Start/Stop	Usun	Edytuj
48	usługa20	1234	!Local Repository!	nie dziala	Uruchom	Usun	Edytuj	
47	usługa1	1234	!Local Repository!	nie dziala	Uruchom	Usun	Edytuj	
52	usługa	123	!Local Repository!	nie dziala	Uruchom	Usun	Edytuj	

Rys. 4.2. Lista usług prognostycznych w podmodule konfiguratora

Usługi zapisywane są w bazie danych i identyfikowane jednoznacznie przez swoją nazwę (nie jest możliwe utworzenie dwóch usług o identycznej nazwie). Próba dodania/edycji usługi prognostycznej powoduje wyświetlenie okna umożliwiającego konfigurację usługi (rysunek 4.3). W trybie dodania nowej usługi okno to jest puste, w trybie edycji - wypełnione jest danymi wskazanej usługi. Możliwe jest ustawienie następujących danych:

- nazwa usługi,
- port, na którym usługa jest dostępna,

- model klasyfikacyjny/regresyjny (pobrane z repozytorium RapidMiner),
- reguły monitorowania jakości prognoz.



Rys. 4.3. Okno konfiguracji usługi prognostycznej

Monitorowanie jakości prognoz odbywa się w czasie rzeczywistym w przesuwanym oknie czasowym o szerokości podanej jako parametr usługi. Monitorowanie jakości rozpoczyna się po wypełnieniu okna czasowego danymi. Kryteria monitorowania różnią się w zależności od tego, czy usłudze prognostycznej przyporządkowany jest model klasyfikacyjny, czy regresyjny. W celu opisu owych kryteriów wprowadzono następującą notację pomocniczą:

- X - wektor wartości rzeczywistych prognozowanego parametru w analizowanym oknie czasowym,
- Y - wektor wartości prognozowanych w analizowanym oknie czasowym,
- $Q(X, Y)$ - miara wyznaczona w oparciu o wektory wartości rzeczywistych oraz prognoz.

4.3.1 Reguły monitorowania jakości klasyfikatora

Ogólna postać reguł monitorowania jakości klasyfikatora jest następująca:

$$Q(X, Y) < T$$

$$Q(X, Y) > T,$$

gdzie T to pewien minimalny próg jakości definiowany przez użytkownika. Relacja w regułach może być postaci $>$ lub $<$ w zależności czy monitorowany wskaźnik jest miarą korzyści i powinien być większy od progu (np. dokładność klasyfikacji), czy jest miarą kosztu i powinien być mniejszy od progu (np. błąd). W przypadku reguł klasyfikacyjnych Q może być jedną z poniższych miar:

- całkowita dokładność klasyfikacji,
- średnia dokładność klasyfikacji (tzw. balanced accuracy),
- specyficzność,
- czułość,
- błędna klasyfikacja.

W przypadku specyficzności oraz czułości wymagane jest dodatkowe wskazanie, która klasa jest traktowana jako pozytywna (pozostałe klasy będą traktowane

jako klasy negatywne). Wskaźnik błędnej klasyfikacji wymaga podania dwóch klas - źródłowej i docelowej. Wskaźnik jest zdefiniowany jako stosunek liczby błędnego sklasyfikowania przykładów z klasy źródłowej do klasy docelowej do ogólnej liczby przykładów z klasy źródłowej. Zasady zastosowania reguł monitorowania jakości są trzy i mogą działać oddzielnie lub wspólnie (łączy je wtedy spójnik lub):

- Zasada 1
 - Użytkownik wybiera wskaźniki oceny: całkowita dokładność klasyfikacji, średnia dokładność klasyfikacji, etc.
 - Użytkownik ustala wartości graniczne (minimalne) dla wybranych wskaźników.
 - Kryterium jakości nie jest spełnione jeżeli przekroczona została wartość graniczna dla dowolnego ze wskaźników oceny.
- Zasada 2
 - Użytkownik wskazuje klasę główną (domyślnie klasa główna to najmniej liczna klasa w zbiorze treningowym).
 - Użytkownik ustala progową wartość czułości i specyficzności; klasa pozytywna odpowiada klasie głównej, a suma pozostałych klas stanowi klasę negatywną.
 - Kryterium nie jest spełnione jeżeli wartość czułości lub specyficzności jest poniżej ustalonych progów.
- Zasada 3 (będąca najbardziej intuicyjną dla użytkownika)
 - Użytkownik dla wybranych par klas $k_i, k_j (i \neq j)$ definiuje maksymalną, dopuszczalną liczbę pomyłek pomiędzy tymi klasami, czyli liczbę przykładów w rzeczywistości należących np. do k_i ale przypisanych przez model do k_j . Użytkownik musi więc podać tylko wartości odpowiadające przypadkom false negative (FN) i false positive (FP), z których wynikają wartości innych miar.
 - Kryterium nie jest spełnione jeżeli zostanie przekroczona maksymalna liczba pomyłek dla dowolnej z wybranych par klas.

4.3.2 Reguły monitorowania jakości modelu regresyjnego

Reguły monitorowania jakości modeli regresyjnych obejmują jedno z kilku kryteriów. Pierwsze kryterium jest identyczne jak w przypadku reguł klasyfikacyjnych, tj.

$$Q(X, Y) < T$$

$$Q(X, Y) > T,$$

gdzie T to minimalny próg jakości. Pozostałe kryteria porównują wyznaczoną wartość wskaźnika jakości z hipotetyczną wartością wskaźnika wyznaczoną w oparciu o wektor wartości rzeczywistych X z uwzględnieniem pewnego odchylenia δ . Wynika to z faktu, że niewielkie błędy nie są tak istotne jak błędy o dużej wartości. Odchylenie może być bezwzględne, co daje poniższe kryteria:

$$Q(X, Y) < Q(X, X + \delta),$$

$$Q(X, Y) < Q(X, X - \delta),$$

$$Q(X, Y) > Q(X, X + \delta),$$

$$Q(X, Y) > Q(X, X - \delta).$$

Jeśli odchylenie jest względne, reguły zdefiniowane są następująco:

$$Q(X, Y) < Q(X, X * (1 + \delta)),$$

$$Q(X, Y) < Q(X, X * (1 - \delta)),$$

$$Q(X, Y) > Q(X, X * (1 + \delta)),$$

$$Q(X, Y) > Q(X, X * (1 - \delta)).$$

Miary jakości dostępne w przypadku reguł regresyjnych to:

- błąd średniokwadratowy,
- korelacja
- błąd bezwzględny.

Zasady zastosowania reguł monitorowania jakości są trzy i mogą działać oddzielnie lub wspólnie (łączy je wtedy spójnik lub):

- Zasada 1
 - Użytkownik podaje zakres (przedział) w jakim powinien mieścić się średni błąd prognozy. Zakres ten może być podany jako odchylenie bezwzględne lub względne.
 - Użytkownik podaje kryterium użyte do monitorowania prognoz - RMSE, MAE lub korelację.
 - Aktualna wartość RMSE/MAE/korelacji zestawiana jest z wartością pesymistyczną. Pesymistyczne RMSE/MAE/korelacja to takie, w którym wartość prognoz różni się od wartości rzeczywistych o "wartość1"/"procent1" w przypadku niedoszacowania oraz o "wartość2"/"procent2" w przypadku przeszacowania przez model.
 - Model nie spełnia kryterium jakości jeżeli aktualna wartość RMSE/MAE jest większa od pesymistycznego RMSE/MAE.
- Zasada 2
 - Użytkownik ustala wartość progową (podawaną w wartościach bezwzględnych).
 - Użytkownik określa ile maksymalnie razy może wystąpić taka sytuacja, że wartość prognozowana jest poniżej proggu a wartość rzeczywista (zmierzona) jest powyżej proggu (false negatives - FN)
 - Użytkownik określa ile maksymalnie razy może wystąpić taka sytuacja, że wartość prognozowana jest powyżej proggu a wartość zmierzona jest poniżej proggu (false positives - FP).
 - Model nie spełnia kryterium jakości jeżeli przekroczone są dopuszczalne wartości FN lub FP.
- Zasada 3
 - Zasada analogiczna do Zasady 2; progi są predefiniowane (np. wynikają z konfiguracji czujnika), może występować kilka wartości progowych (np. stan ostrzegawczy, stan alarmowy).
 - Użytkownik dla każdego proggu definiuje ile maksymalnie razy może wystąpić taka sytuacja, że wartość prognozowana jest poniżej/powyżej ustalonego proggu podczas gdy wartość zmierzona jest powyżej/poniżej proggu.
 - Model nie spełnia kryterium jakości jeżeli przekroczone są dopuszczalne wartości zdefiniowane dla dowolnego z progów.

Panel konfiguracji reguł monitorowania jakości prognoz zmienia się w zależności od tego, czy wybrano model klasyfikacyjny (rysunek 4.4), czy regresyjny (rysunek 4.5). Część parametrów jest wspólna dla obu typów reguł:

- szerokość okna - odcinek czasu, który uwzględniany jest przy ocenie jakości wyrażony w sekundach,
- wskaźnik jakości - miara Q brana pod uwagę przy ocenie,
- relacja (< lub >) - wskazuje, czy wskaźnik powinien znajdować się powyżej czy poniżej progu/wartości hipotetycznej.

Pozostałe parametry konfiguracyjne są specyficzne dla modeli. W przypadku reguł klasyfikacyjnych dostępne są następujące parametry:

- próg - próg T powyżej/poniżej, którego powinien znajdować się wskaźnik,
- klasa główna/źródłowa - klasa główna w przypadku wskaźników czułość i specyficzność, klasa źródłowa w przypadku wskaźnika błędna klasyfikacja,
- klasa docelowa - klasa docelowa dla wskaźnika błędna klasyfikacja.

The screenshot shows a configuration window titled "Konfiguracja reguły". It contains several input fields and dropdown menus:

- "Szerokość okna[s]": a text input field.
- "Wskaźnik jakości": a dropdown menu with "Dokladnosc" selected.
- "Próg": a text input field.
- "Relacja": a dropdown menu with "większy" selected.
- "Klasa główna/źródłowa": a dropdown menu.
- "Klasa docelowa": a dropdown menu.
- "Dodaj": a button at the bottom left.

Rys. 4.4. Panel konfiguracyjny reguł monitorowania dla modeli klasyfikacyjnych

The screenshot shows a configuration window titled "Konfiguracja reguły". It contains several input fields and dropdown menus:

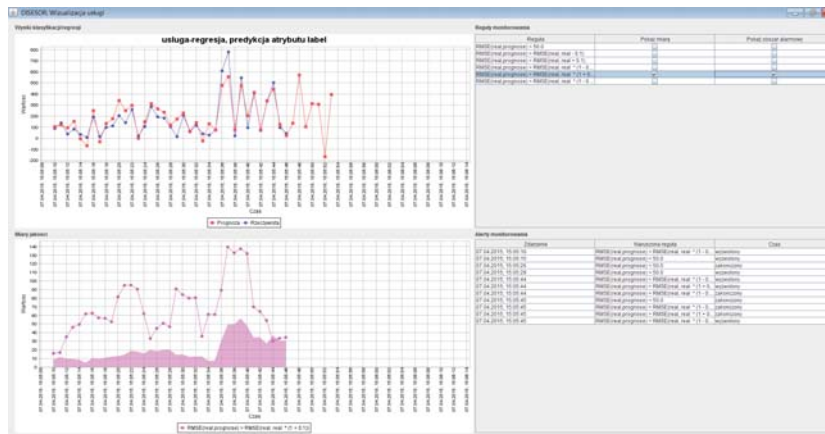
- "Szerokość okna[s]": a text input field.
- "Wskaźnik jakości": a dropdown menu with "RMSE" selected.
- "Relacja": a dropdown menu with "większy" selected.
- "Kryterium": a dropdown menu with "prog" selected.
- "Próg/odchylenie": a text input field.
- "Dodaj": a button at the bottom left.

Rys. 4.5. Panel konfiguracyjny reguł monitorowania dla modeli regresyjnych

4.4 Wizualizator

Podobnie jak w przypadku *Konfiguratora*, po uruchomieniu *Wizualizatora* wyświetlana jest lista istniejących usług prognostycznych wraz z podstawowymi informacjami. Za pomocą tej listy można dokonać wyboru, która usługa prognostyczna ma być wizualizowana. Okno wizualizacji otwarte dla wybranej usługi prognostycznej posiada wyodrębnione cztery części:

1. wykres wartości prognozowanej,
2. lista reguł monitorowania,
3. wykres reguł monitorowania,
4. lista alertów monitorowania.

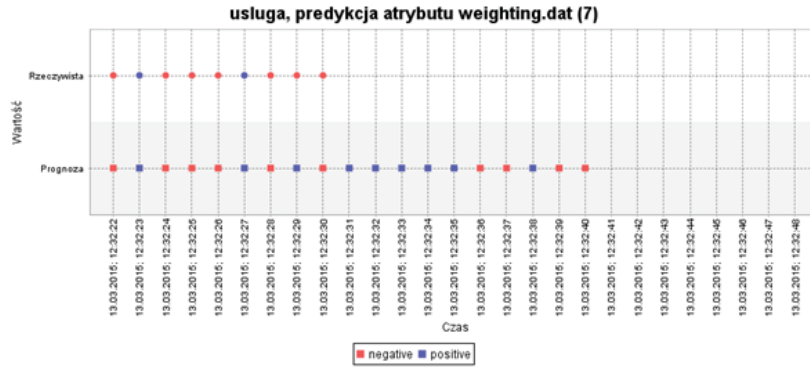


Rys. 4.6. Okno wizualizacji usługi prognostycznej

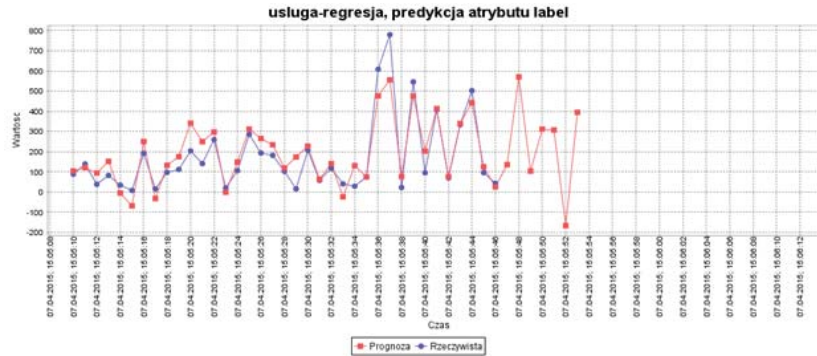
4.4.1 Wykres wartości prognozowanej

Wykres ten zawiera dwie serie danych - jedna reprezentuje wartość prognozowaną, a druga wartość rzeczywistą. W obu przypadkach oś odciętych reprezentuje czas. Znaczenie osi rzędnych różni się w zależności od tego, czy prognozowana wartość jest symboliczna (problem klasyfikacji), czy ciągła (problem regresji). W przypadku klasyfikacji na osi Y znajdują się dwie wartości symboliczne - jedna reprezentuje prognozę, a druga wartość rzeczywistą. Wartość zmiennej w konkretnym punkcie czasu jest z kolei reprezentowana przez znacznik - każdej klasie decyzyjnej przyporządkowany jest inny znacznik (rysunek 4.7). W przypadku regresji oś Y reprezentuje po prostu wartość prognozowanej zmiennej (rysunek 4.8).

Uruchomienie *Wizualizatora* powoduje ściągnięcie z bazy danych i wyświetlenie wszystkich zarchiwizowanych wartości (prognozowanych oraz rzeczywistych)



Rys. 4.7. Wykres reprezentujący wyniki klasyfikacji. Każdej etykietce klasy (w tym wypadku *negative* i *positive*) odpowiada pewien znacznik



Rys. 4.8. Wykres reprezentujący wyniki regresji. Serie odpowiadają wartości prognozowanej oraz rzeczywistej

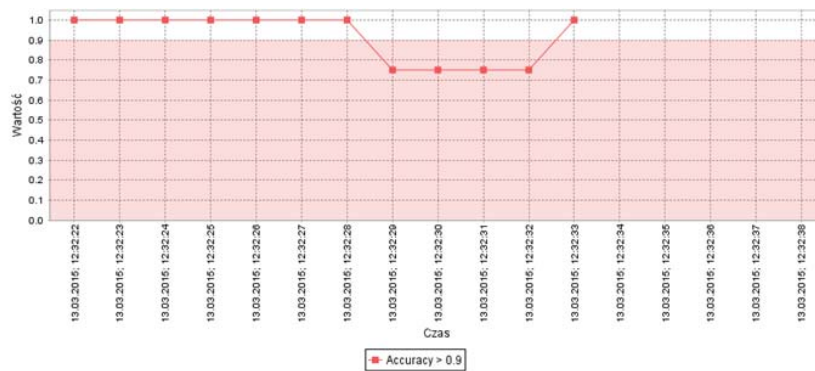
od zadanego momentu. Następnie uruchamiany jest wątek sprawdzający periodycznie stan bazy danych. W momencie gdy usługa zapisuje w bazie wartość monitorowaną bądź rzeczywistą, *Wizualizator* natychmiast uzupełnia wykres o nowy punkt. *Wizualizator* działa więc w czasie rzeczywistym.

4.4.2 Lista i wykres reguł monitorowania

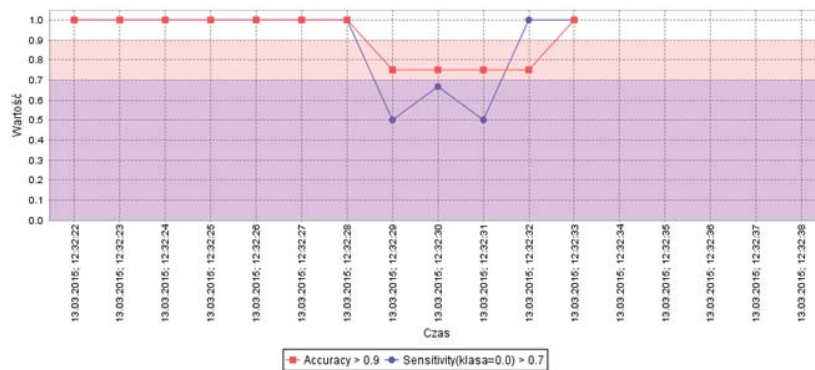
Wizualizator pozwala na śledzenie jakości dla stosowanego przez usługę modelu, na podstawie zdefiniowanych reguł monitorowania.

Lista reguł monitorowania zawiera wszystkie zdefiniowane dla wizualizowanej usługi reguły monitorowania. Dla każdej z reguł możliwa jest wizualizacja miary powiązanej z regułą oraz wizualizacja obszaru alarmowego. Obie wizualizacje pokazywane są na wykresie reguł monitorowania. Wykres ten służy do wizualizacji wartości miar jakości oraz obszarów alarmowych powiązanych z regułami monitorowania. Włączenie wizualizacji dla wybranej z listy reguły powoduje dodanie

do wykresu serii danych reprezentujących tę regułę (rysunek 4.9). Wykres może być uzupełniony o wizualizację obszaru alarmowego dla tej reguły. Możliwa jest równoczesna wizualizacja dowolnej ilości reguł (rysunek 4.10). Miara w danym punkcie czasu wyznaczana jest na podstawie wartości prognozowanej oraz rzeczywistej. Brak wartości rzeczywistej, dla pewnego punktu powoduje więc, że nie można wyliczyć dla niego miary jakości. Wykres miar jakości, podobnie jak wykres wyników prognozowania jest na bieżąco aktualizowany w czasie rzeczywistym.



Rys. 4.9. Wykres reprezentujący pewną regułę monitorowania - czerwona linia symbolizuje monitorowaną miarę jakości, a półprzezroczysty czerwony obszar wskazuje obszar alarmowy (jeśli miara znajdzie się w nim - wywoływany jest alarm)



Rys. 4.10. Wizualizacja wielu reguł monitorowania

4.5 Zasilanie danymi modułu prognostycznego

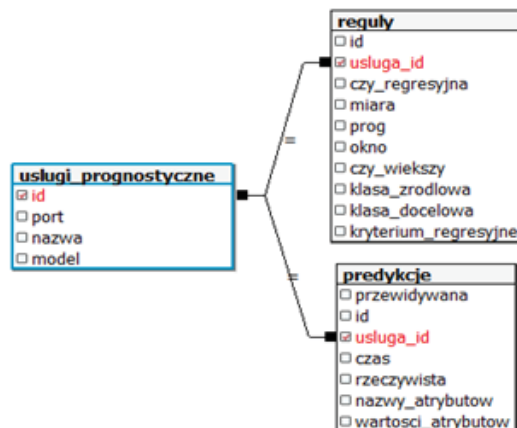
Modele prognostyczne przygotowywane są na podstawie danych historycznych pobranych z hurtowni danych, które następnie są analizowane w module analitycznym (patrz rozdział 5). Gdy tworzony model zostanie zaakceptowany (jego jakość będzie akceptowalna), zostaje on zapisany i może być wykorzystywany w praktycznym działaniu z poziomu modułu prognostycznego.

Dane dla modelu przekazanego do modułu prognostycznego nie mogą być jednak pobierane z hurtowni danych, gdyż nie jest ona uzupełniana na bieżąco ale co pewien czas. Z tego względu razem z modelem prognostycznym zapisywane są również wszystkie przekształcenia na "surowych" danych jakie zostały wykonane w celu uzyskania danych treningowych. Model prognostyczny wykorzystuje owe przekształcenia otrzymując jako dane wejściowe przychodzące na bieżąco (ang. on-line) dane pochodzące z systemu monitorowania. Przekształcenia, takie jak np. agregacja, utworzenie zmiennych wywiedzionych pozwalają wygenerować wektor, który może zostać przekazany modelowi w celu podjęcia decyzji o prognozie. W celu monitorowania jakości prognoz moduł prognostyczny otrzymuje również (z opóźnieniem) z systemu monitorowania rzeczywistą wartość, która może zostać porównana z prognozą.

4.6 Repozytoria modułu prognostycznego

Na potrzeby modułu prognostycznego utworzona została dodatkowa baza danych, w której zapisywane są np. wyniki prognoz oraz zdefiniowane reguły monitorowania. Na rysunku 4.11 zaprezentowano schemat tej bazy. Podstawowa tabela w bazie nosi nazwę *uslugi_prognostyczne* i przechowuje dane konfiguracyjne usług prognostycznych, takie jak: nazwa, port, model klasyfikacyjny/regresyjny. Reguły monitorowania zapisane są w tabeli *reguly*, która jest powiązana z tabelą *uslugi_prognostyczne* za pomocą klucza obcego *usluga_id*. Tabela *reguly* zawiera wszystkie dane reguł (np. szerokość okna, wykorzystana miara jakości, próg/odchylenie, relacja, kryterium). Wszystkie prognozy wykonane przez moduł przechowywane są w tabeli *predykcje*. Tabela ta również jest powiązana z usługami za pomocą klucza obcego *usluga_id*. Zapisane informacje to przede wszystkim wartość prognozowana i rzeczywista w postaci numerycznej oraz znacznik czasowy. Dodatkowo, dla każdej predykcji przechowywany jest wektor nazw atrybutów warunkowych oraz wektor ich wartości (oba wektory mają postać tekstową).

Informacje o modelach i ich historia przechowywane są natomiast w repozytorium zintegrowanym z programem RapidMiner, na bazie którego powstał system DISESOR. W tym samym repozytorium zapisywane są procesy przekształcające dane wejściowe dla modułu prognostycznego. RapidMiner zapisuje utworzone w nim procesy w postaci plików XML o odpowiedniej strukturze.



Rys. 4.11. Schemat tabel bazy danych utworzonej dla modułu prognostycznego

4.7 Przeuczanie modeli prognostycznych

W przypadku niespełnienia minimalnych wymagań jakości modelu, moduł prognostyczny próbuje w sposób automatyczny przetrenować model. Realizowane jest to w taki sposób, aby bazując na tej samej metodzie dopasować parametry do zmieniającego się strumienia danych. Ponowna nauka odbywa się na uaktualnionym zbiorze treningowym o takiej samej wielkości, ale obejmującym najbardziej aktualne dane. W przypadku problemu klasyfikacji oraz danych niezrównoważonych tj. takich, dla których wielkość klasy mniejszościowej nie przekracza 10% zbioru treningowego, wszystkie przykłady z klasy mniejszościowej z poprzedniego zbioru treningowego są zachowywane w nowym zbiorze treningowym. Przyjmując J_{old} jako jakość bieżącego modelu (wyrażoną np. dokładnością klasyfikacji) wyznaczoną na poprzednim zbiorze treningowym oraz J_{new} jako jakość nowego modelu wyznaczoną na nowym zbiorze treningowym, bieżący model jest zastępowany nowym jeżeli:

- J_{new} nie jest gorsza od J_{old} z ustaloną - parametrem - tolerancją (np. 5%),
- nowy model spełnia kryterium jakości dla aktualnego - najnowszego - okna czasowego.

Jeżeli nowy model nie spełnia wymagań odnośnie jego jakości moduł prognostyczny generuje komunikat o konieczności zmiany modelu i zaprzestaje prognozowania.

Literatura

- [4.1] AVL. AVL CONCERTO, 2015.
 [4.2] Dell. STATISTICA MAS (monitoring & alerting system), 2016.

-
- [4.3] P. Kadlec, B. Gabrys, S. Strandt. Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33(4):795–814, 2009.
 - [4.4] M. Kozielski. *A Meta-Learning Approach to Methane Concentration Value Prediction*, s. 716–726. Springer International Publishing, Cham, 2016.
 - [4.5] M. Kozielski, A. Skowron, Ł. Wróbel, M. Sikora. Regression rule learning for methane forecasting in coal mines. S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, D. Kostrzewa, redaktorzy, *Beyond Databases, Architectures and Structures*, vol. 521 serii *Communications in Computer and Information Science*, s. 495–504. Springer International Publishing, 2015.
 - [4.6] C. Shearer. The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5(4):13–22, 2000.
 - [4.7] M. Sikora, B. Sikora. Rough natural hazards monitoring. *Rough Sets: Selected Methods and Applications in Management and Engineering*, s. 163–179. Springer, 2012.
 - [4.8] G. Tennant. *Six Sigma: SPC and TQM in Manufacturing and Services*. Gower Publishing, Ltd., 2001.