

Zastosowanie fuzji danych z czujników i eksploracji danych w prognozowaniu stężenia metanu w kopalniach węgla

W ostatnich latach można było zaobserwować niespotykany wzrost użycia czujników w wielu zastosowaniach przemysłowych. Nowoczesne czujniki są w stanie nie tylko generować duże ilości danych, lecz również przysyłać te dane za pomocą sieci i przechowywać je do późniejszej analizy. Umożliwia to opracowanie systemów do łączenia danych w czasie rzeczywistym w celu prognozowania określonych zdarzeń. Celem niniejszej pracy jest prognozowanie poziomów stężenia metanu w kopalniach węgla za pomocą technik fuzji danych i eksploracji danych. Artykuł przedstawia zastosowanie generycznej metody, która może być użyta do dowolnego zbioru danych wielowymiarowych szeregów czasowych w celu przeprowadzenia zadań klasyfikacji lub regresji. Zaprezentowane rozwiązanie zostało opracowane w ramach konkursu eksploracji danych IJCRS'15 i – pokonując inne rozwiązania – zostało jego zwycięzcą.

słowa kluczowe: eksploracja danych, badanie danych, prognozowanie, wybuchy metanu.

1. WPROWADZENIE

Wybuchy metanu i tąpnięcia górotworu są nieodłącznie związane z procesem wydobywania węgla w kopalniach podziemnych. Dlatego bezpieczeństwo to jeden z głównych priorytetów przemysłu górniczego. Nowe technologie, takie jak czujniki sieciowe i analizy prognostyczne, wspomagają działania zmierzające do zapewnienia bezpiecznego środowiska pracy dla personelu podziemnego. Potencjał tych technologii to nie tylko niezawodne monitorowanie w czasie rzeczywistym, lecz także zaawansowane sposoby alarmowania oparte na prognostycznych modelach numerycznych [8, 9]. W szczególności techniki eksploracji danych umożliwiają automatyczne tworzenie modeli na podstawie danych historycznych, a nie wiedzy eksperckiej. Teoretycznie są one w stanie identyfikować wzorce potencjalnie nieznanymi ekspertom, mogą prowadzić do lepszego prognozowania i często pozwalają uzyskać lepszy wgląd w daną dziedzinę.

Artykuł przedstawia generyczne podejście do klasyfikacji danych wielowymiarowych szeregów czasowych w prognozowaniu wybuchów metanu. Zostało ono opracowane i ocenione w ramach konkursu 2015 AAIA Data Mining Competition, gdzie otrzymało drugą najwyższą ocenę [10]. Publikacja dotyczy zastosowania tego samego, podstawowego podejścia do danych zbieranych z czujników kopalnianych w ramach innego konkursu z dziedziny eksploracji danych: IJCRS'15 Data Challenge: Mining Data from Coal Mines. Przedstawione w tym konkursie rozwiązanie zdobyło pierwszą nagrodę. Chociaż założenia podejścia pozostają takie same, w procesie opracowania modeli wymagany jest etap dostosowania rozwiązania do potrzeb użytkownika, co jest specyficzną cechą problemu i charakterystyki zbioru danych. Etap dostosowania budzi szczególne zainteresowanie ekspertów w tej dziedzinie, ponieważ rzuca nowe światło na charakterystyki odczytów czujnika, które mogą wcześniej, w horyzoncie czasowym kilku minut, dostarczyć danych świadczących o wzroście poziomemu stężenia metanu.

Dalsze części artykułu uporządkowano w następujący sposób: w kolejnej sekcji omówiono krótko technologie dotyczące tematyki artykułu oraz koncepcję konkursów z dziedziny eksploracji danych, następnie przedstawiono zadanie konkursu wraz ze szczegółami dotyczącymi czujników, dostępnymi danymi i oceną, w ostatnim rozdziale omówiono zaproponowane podejście do klasyfikacji danych wielowymiarowych szeregów czasowych. Bardziej szczegółowo omówiono każdy z etapów proponowanego podejścia: ekstrakcję cech i rzeczywistą klasyfikację.

2. KONTEKST BADAŃ

W ciągu ostatniej dekady niezawodne i przystępne cenowo czujniki, które są w stanie gromadzić duże ilości danych, stały się popularne w wielu zastosowaniach: przemysłowych, komercyjnych i związanych z codziennym życiem społeczeństwa. Jednym z najczęściej gromadzonych przez te czujniki typów danych są szeregi czasowe. Składają się one zazwyczaj z sekwencji pomiarów dokonywanych w określonych przedziałach czasowych. Dzięki zastosowaniu niedrogich czujników, zdolnych do przesyłania danych w sieci, wielowymiarowe szeregi czasowe stały się powszechnym przedmiotem analizy. Przykłady ich użycia to monitorowanie pojazdów lub maszyn, czujniki dla smartfonów, czy też zestawy czujników instalowane na ludzkim ciele. Z powodu właściwości szeregów czasowych (zebrane pomiary nie są zwykle bezpośrednio gotowe do użycia, ponieważ każdy pomiar składa się z wielu punktów danych) są one często narażone na szумы i wymagają dalszej analizy w celu identyfikacji interesujących wzorców, które mogą być wykorzystane przez użytkowników. Przetwarzanie surowych pomiarów i przekształcanie danych w wiedzę użyteczną dla użytkowników to zadanie wymagające i kosztowne. Jest to szczególnie widoczne w przypadku wielowymiarowych szeregów czasowych, ponieważ szeregi czasowe charakteryzują się dużą ilością danych i często muszą być poddawane transformacjom (np. transformacji Fouriera, różnym rodzajom filtrowania itp.) w celu uwidocznienia potencjalnie użytecznych wzorców. Z drugiej strony, jeżeli opracowane zostaną metody generyczne do przekształcania wielowymiarowych szeregów czasowych, mogą one prowadzić do szybkiego postępu w użyciu danych z czujników w wielu obszarach. W niniejszym artykule zaprezentowano zastosowanie metody do klasyfikacji wielowymiarowych szeregów czasowych. Metoda ta została opracowana w ramach konkursu z dziedziny

eksploracji danych, w tym czujników ruchu umieszczonych na ciele ludzkim. Następnie metoda została użyta do rozwiązywania różnych problemów dotyczących czujników zamontowanych w kopalniach.

W zastosowaniach przedstawionych w niniejszym artykule dane składały się z pomiarów zebranych przez czujniki zainstalowane na maszynach działających w różnych miejscach kopalni. Czujniki te odpowiedzialne były za różne rodzaje pomiarów, głównie pomiary środowiskowe, czyli poziomu wilgotności, temperatury, ciśnienia powietrza, stężenia metanu itp., oraz niektóre pomiary związane ze stanem działających urządzeń mechanicznych, takie jak prędkość i kierunek pracy kombajnu ścianowego oraz prądy w różnych częściach urządzeń. Ich zadaniem było prognozowanie, czy poziom metanu przekraczający określone progi mógłby wystąpić w ciągu najbliższych 3-6 minut po dokonaniu pomiarów. Ta wiedza pozwoliłaby uzyskać dodatkowy czas na ostrzeżenie, zanim zostanie przekroczony poziom ostrzegawczy, oraz na przedsięwzięcie działań prewencyjnych.

Aby zaproponować model prognostyczny, przygotowano konkurs z dziedziny eksploracji danych w ramach konferencji naukowej. Konkursy takie są często organizowane celem zachęcenia naukowców, studentów i hobbystów do zajęcia się danym problemem i zidentyfikowania najlepszej metody. Mają one wiele wspólnego z zawodami sportowymi, gdzie uczestnicy przedstawiają swoje rozwiązania, a działanie przedstawionych modeli zależy od niezależnej i uczciwej oceny organizatorów. Rezultaty publikowane są w formie publicznie ogłaszanej tablicy wyników. Uczestnicy mogą obserwować przebieg konkursu w czasie i zgłaszać dowolną liczbę pomysłów. Typowa długość konkursu to kilka miesięcy, co daje wystarczająco dużo czasu na sprawdzenie różnych rozwiązań i opracowanie bardziej wyrafinowanych i dojrzałych propozycji. Zwycięzcy otrzymują nagrody, zazwyczaj w formie pieniężnej (kilka tysięcy dolarów), i wyróżnienia, takie jak darmowa rejestracja na konferencji albo sprzęt komputerowy. Konkursy z dziedziny eksploracji danych są wysoko oceniane przez społeczność akademicką ze względu na ich obiektywizm oraz realistyczne środki oceny algorytmów i pomysłów.

3. PROGNOZOWANIE STĘŻENIA METANU – KONKURS IJCRS'15

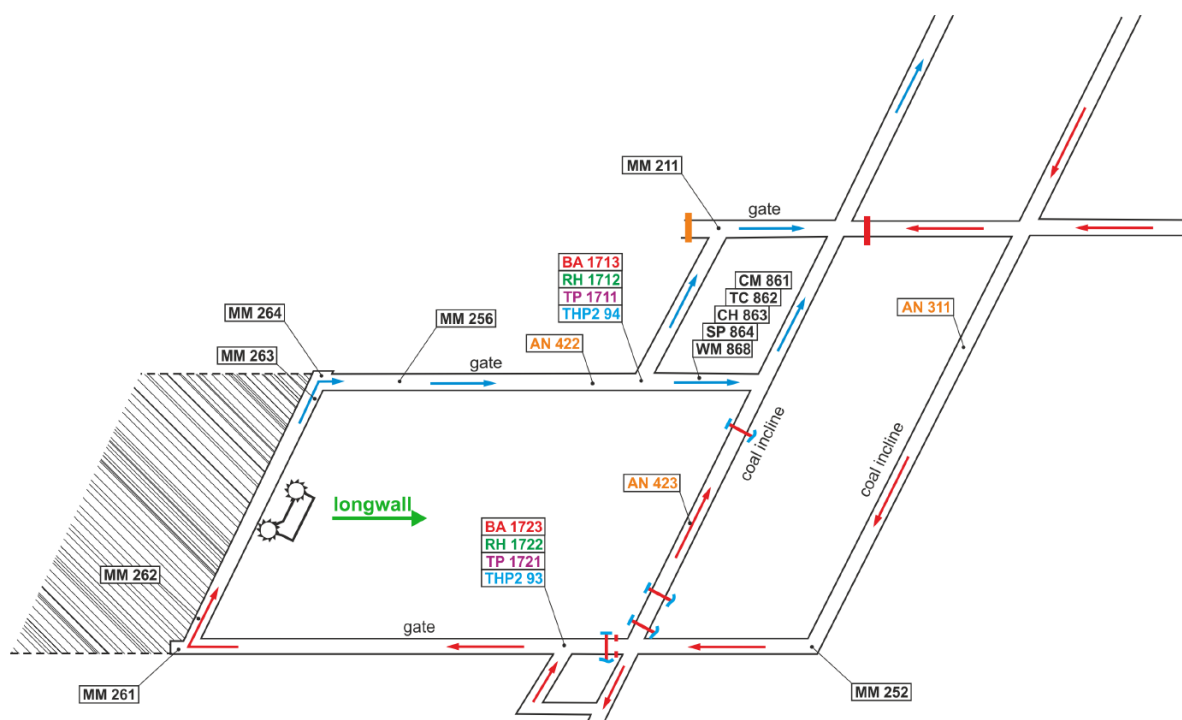
W artykule przedstawiono rozwiązanie zgłoszone do konkursu IJCRS'15, który był zorganizowany przy użyciu platformy konkursowej Knowledge Pit [6, 7]. Celem konkursu było uzyskanie wiedzy na

temat zależności pomiędzy działaniem kombajnu ścianowego a poziomem stężenia metanu, mierzonym przez czujniki rozmieszczone w różnych miejscach kopalni.

Głównym zadaniem konkursu było utworzenie modelu numerycznego do prognozowania przekroczonych poziomów progowych na trzech czujnikach metanu w odniesieniu do najbliższej przyszłości (3 do 6 minut) i w oparciu o odczyty z wielu czujników.

Dane wykorzystane w konkursie zebrano w jednej z polskich kopalń w okresie od 2 marca do 16 czerw-

ca 2014 r. Podstawowy zbiór danych do konkursu składał się z wielowymiarowych szeregów czasowych właściwych dla odczytów czujników używanych do monitorowania środowiska i stanu maszyn górniczych w rejonie ściany wydobywczej. Schemat przedstawiający rozmieszczenie czujników środowiskowych został udostępniony uczestnikom konkursu (rys. 1). Dodatkowo organizatorzy dostarczyli opisy czujników, tak aby umożliwić uczestnikom zrozumienie zależności pomiędzy odczytami różnych czujników.



Rys. 1. Schemat procesu urabiania, na podstawie którego wygenerowano dane konkursowe

Kombajn poruszał się wzdłuż ściany wydobywczej pomiędzy czujnikami MM262 i MM264. Jego pracę można oszacować poprzez pomiar prądów elektrycznych i takie pomiary stanowiły dane konkursowe. Uważa się, że im bardziej intensywna jest praca kombajnu, tym więcej metanu wydziela się ze ściany do powietrza kopalni. Kierunki przepływu powietrza w korytarzach dostarczono uczestnikom – są przedstawione na schemacie. Jeżeli stężenie metanu zmierzone przez jakikolwiek z czujników osiąga poziom alarmowy, kombajn wyłącza się automatycznie, powodując przestój i straty finansowe.

Na tej podstawie zidentyfikować można praktyczne korzyści prognozowania wybuchów metanu. Dzięki skutecznej prognozie wartości stężeń metanu można dostosować prędkość kombajnu w taki sposób, żeby nadmiar metanu mógł opuścić dany obszar, unikając przekroczenia progów bezpieczeństwa i tym samym pozwalając na nieprzerwaną pracę urządzenia.

Wszystkie dane wykorzystane w konkursie zostały dostarczone przez Instytut Technik Innowacyjnych EMAG¹, który był również głównym sponsorem konkursu.

3.1. Dane

Dane udostępnione do konkursu składały się wyłącznie z szeregów czasowych i obejmowały w sumie 51 700 zapisów, co tworzyło zbiór szeregów czasowych. Każdy zapis złożony był z 28 szeregów czasowych, a każdy szereg dokładnie z 600 wartości i odnosił się do 10-minutowych pomiarów. Pomiary przeprowadzane były w anemometrach, czujnikach temperatury, czujnikach prądów (dot. maszyn) oraz czujnikach kierunku posuwu i prędkości kombajnu.

¹ www.ibemag.pl

W sumie każdy zapis w zbiorze danych składał się z 16 800 atrybutów liczbowych.

Zadanie polegało na prognozowaniu wartości stężeń metanu przekraczających określony poziom i mierzonych przez trzy czujniki metanu. Zmiennymi zależnymi były trzy atrybuty binarne (przekroczony lub nieprzekroczony próg) dla trzech wybranych czujników metanu. Trzy zależne atrybuty wskazywały, czy próg ostrzegawczy dla trzech czujników metanu MM263, MM264 i MM256 osiągnięty został w czasie od 3 do 6 minut po zakończeniu okresu uczącego. Jeżeli dany rząd odnosi się do okresu pomiędzy $t_{.599}$ i t_0 , to etykieta czujnika metanu MM w tym rządzie ma status „ostrzegawczy” wtedy i tylko wtedy, gdy $\max(\text{MM}(t_{181}), \dots, \text{MM}(t_{360})) \geq 1,0$, co w praktyce oznacza, że poziom ostrzegawczy stężeń metanu został przekroczony w czasie od 3 do 6 minut po zakończeniu odpowiadającego mu szeregu czasowego.

Oryginalne dane podzielono na dwa zbiory: zbiór uczący i zbiór testowy. Dla zbioru uczącego dostarczono zmienne zależne. Celem zadania była prognoza dotycząca prawdopodobieństwa przekroczenia progu ostrzegawczego dla każdej z trzech zmiennych zależnych w zbiorze testowym. Wartości zmiennych zależnych dla zbioru testowego nie były udostępnione użytkownikom, a jedynie organizatorom konkursu.

Warto zauważyć, że szeregi czasowe w danych testowych nie pokrywały się i były podane w porządku losowym. To czasowe rozłączenie pomiędzy zbiorem uczącym i testowym powoduje, że powszechne w tej dziedzinie założenie dotyczące danych o takim samym niezależnym rozkładzie (ang. *i.i.d.* – *identically independently distributed*) pozostaje niespełnione [2] i stwarza największe trudności w rozpatrywanym tu zadaniu.

3.2. Kryterium oceny wyników

Ocenę wyników przeprowadzono za pomocą koncepcji pomiaru pola pod krzywą ROC (ang. *Area Under the ROC Curve* – AUC) [5]. Było to możliwe, ponieważ zmienne obowiązkowe zostały zdefiniowane w formie prawdopodobieństwa przekroczenia progu dla każdej z trzech zmiennych. Dla każdej z nich najpierw wyliczono osobny wynik AUC. Przyjmijmy, że wynik AUC dla n -tej (ang. i^{th}) zmiennej obowiązkowej to AUC_i . Wynik końcowy to średnia z trzech indywidualnych wyników:

$$AUC_{total} = \frac{1}{3} \sum_{i=1}^3 AUC_i \quad (1)$$

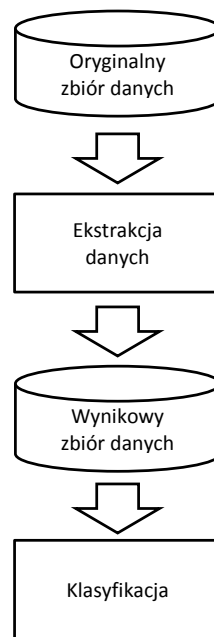
Podczas konkursu użytkownicy mieli dostęp tylko do wstępnych wyników. Były one oparte na podzbiórze końcowego zbioru testowego i odnosiły się do

około 20% danych testowych. Ocena końcowa została przeprowadzona po zakończeniu konkursu.

4. OPIS ROZWIĄZANIA

W niniejszej sekcji przedstawiono opis rozwiązania zadania postawionego w konkursie. Wykorzystano metodę opartą na tej opracowanej do konkursu AAIA'15, opisaną w [10]. Kluczową różnicą jest to, że na potrzeby konkursu IJCRS'15 nie użyto etapu ekstrakcji cech. Decyzja ta wynikała z faktu, że w tym konkretnym konkursie wybór atrybutów powodował gorsze wyniki i okazał się niepotrzebny. Na rys. 2. przedstawiono podstawowe kroki metody.

Pierwszym i prawdopodobnie najbardziej decydującym krokiem była ekstrakcja cech. Na tym etapie zbiór danych oryginalnych został przekształcony na zbiór danych wtórnych składający się z cech generowanych z danych szeregów czasowych. Krok ten został szczegółowo opisany w sekcji 5.



Rys. 2. Schemat podstawowych kroków zrealizowanych podczas konkursu

Dla każdej z trzech zmiennych obowiązkowych zdecydowano się utworzyć oddzielny klasyfikator, który podejmował decyzję binarną. W rezultacie nastąpiła potrzeba nauczenia trzech oddzielnych klasyfikatorów (jednego dla każdej zmiennej decyzji). Należy podkreślić, że pomiędzy trzema klasyfikatorami nie było żadnej wymiany informacji i wszystkie trzy były uczone za pomocą tego samego zbioru danych cech. Zadanie wymagało przypisania prawdopodobieństwa (w miejsce

twardej decyzji) przekroczenia progu. Jako głównego klasyfikatora użyto metody lasów losowych (ang. *Random Forest*). Korzyścią płynącą z zastosowania tej metody jest to, że pozwala ona obliczyć prawdopodobieństwo przypisania klasy.

Zdecydowano się rozszerzyć oryginalny zbiór szeregów czasowych używanych do wyboru cech poprzez utworzenie dodatkowych szeregów czasowych, pochodzących z oryginalnych szeregów czasowych. Pochodne szeregi czasowe zostały wygenerowane z pary szeregów oryginalnych. Załóżmy, że $x(t)$ oraz $y(t)$ to dwa oryginalne szeregi czasowe. W takiej sytuacji pochodne szeregi czasowe są generowane, jeżeli:

- zarówno $x(t)$, jak i $y(t)$ są czujnikami metanu (ich nazwy rozpoczynają się od M),
- zarówno $x(t)$, jak i $y(t)$ rozpoczynają się od BA, RH, TP i AN – wszystkie odnoszą się do konkretnego typu czujników środowiskowych: ciśnienia, wilgotności, temperatury i prędkości powietrza.

Dla każdej pary sygnałów $x(t)$ i $y(t)$, które spełniły powyższe warunki, utworzono dwa pochodne szeregi czasowe $d1(t)$ i $d2(t)$. Zostały one wyprowadzone w następujący sposób:

- $d_1(t) = x(t) - y(t)$ – prosta różnica pomiędzy odpowiednimi pomiarami,
- $d_2(t) = (x(t) - y(t)) / x(t)$ – względna różnica pomiędzy odpowiednimi pomiarami.

Wynikiem tej operacji są 52 pochodne szeregi czasowe.

5. EKSTRAKcja CECH

Następnym krokiem było przekształcenie danych w formie szeregów czasowych na zbiór wartości liczbowych, które podsumowują różne aspekty danych szeregów czasowych.

Najbardziej podstawowymi cechami, jakie można wyprowadzić z pojedynczych szeregów czasowych, są proste statystyki (np. średnia, odchylenie standardowe). Bardziej złożone cechy natomiast można wyprowadzić z więcej niż jednego szeregu czasowego (np. współczynnik korelacji pomiędzy dwoma szeregami). W trakcie konkursu przeprowadzono wiele eksperymentów z użyciem różnych cech. Użyto algorytmów wyboru cech z oprogramowania Weka [3], aby zidentyfikować cechy najbogatsze w informacje. W wyniku tej analizy szczególny nacisk położono na cechy związane z wartościami maksymalnymi lub minimalnymi, ponieważ wydawały się najbogatsze w informacje, przynajmniej według algorytmów wyboru cech. Warto

zauważyć, że po wybór cech sięgnięto wyłącznie w celu przekazania informacji do ekstrakcji danych. Nie wykorzystano wyboru danych, aby określić cechy do klasyfikacji – użyto wszystkich wygenerowanych cech do zadania klasyfikacji.

5.1. Wygenerowane cechy

Dla każdego z szeregów czasowych (czy to oryginalnych, czy pochodnych) dokonano ekstrakcji takich cech, jak:

- wartość średnia,
- odchylenie standardowe,
- wartość minimalna,
- wartość maksymalna,
- średnia pięciu najmniejszych wartości minimalnych,
- średnia pięciu największych wartości maksymalnych,
- wartość minimalna wyrażona standardowymi odchyleniami od średniej,
- wartość maksymalna wyrażona standardowymi odchyleniami od średniej,
- średnia pięciu najmniejszych wartości minimalnych wyrażona standardowymi odchyleniami od średniej,
- średnia pięciu największych wartości maksymalnych wyrażona standardowymi odchyleniami od średniej,
- maksymalna różnica pomiędzy wartościami minimalnymi i maksymalnymi zebranymi z niemalejących sekwencji pomiarów,
- maksymalna różnica pomiędzy wartościami minimalnymi i maksymalnymi zebranymi z nierosnących sekwencji pomiarów,
- wartości maksymalne (częstotliwości i mocy) dla szybkiej transformacji Fouriera z pominięciem pierwszych trzech częstotliwości,
- parametry regresji liniowej: nachylenie, wyraz wolny, błąd średniokwadratowy oraz bezwzględna wartość nachylenia,
- parametry dopasowania wielomianu (wykonywane tylko dla dopasowania parabolicznego): a_0 , a_1 i a_2 ,
- parametry dopasowania wielomianu zebrane z pierwszej połowy sygnału (wykonywane tylko dla dopasowania parabolicznego): a_0 , a_1 i a_2 ,
- parametry dopasowania wielomianu zebrane z drugiej połowy sygnału (wykonywane tylko dla dopasowania parabolicznego): a_0 , a_1 i a_2 .

Każda z powyższych cech wygenerowała pojedynczą liczbę, która została użyta jako cecha indywidualna w dalszej analizie. W ten sposób powstało 2214 cech – 756 z oryginalnych szeregów czasowych i 1458 z pochodnych szeregów czasowych.

5.2. Korelacje

Na koniec zdecydowano się wprowadzić współczynniki korelacji pomiędzy szeregi czasowe. Dodatkowe parametry zostały wyprowadzone z korelacji krzyżowych (w tym również autokorelacji) pomiędzy wybranymi parami sygnałów:

- korelacje krzyżowe dla sygnału pobranego przy $t = 0$ oraz tego samego sygnału pobranego przy $t = 0, 100, 200$ i 300 , przy użyciu współczynnika korelacji Pearsona,
- korelacje krzyżowe dla sygnału pobranego przy $t = 0$ oraz tego samego sygnału pobranego przy $t = 0, 100, 200$ i 300 , przy użyciu współczynnika korelacji Spearmana,
- korelacje krzyżowe dla sygnału pobranego przy $t = 0$ oraz tego samego sygnału pobranego przy $t = 0, 100, 200$ i 300 , przy użyciu współczynnika korelacji Kendalla.

Pary sygnałów $x(t)$ i $y(t)$, które wybrano w celu obliczenia współczynników korelacji, obejmowały:

- dowolne pomiary z czujników MM,
- sygnały rozpoczynające się od tych samych prefiksów BA, RH i AN,
- pary powstające tylko w przypadku, jeżeli dwa sygnały miały ten sam prefiks, np. BA z BA, ale nie z innymi.

Prowadziło to do włączenia autokorelacji, ponieważ pozwolono, aby $x(t)=y(t)$. Całkowita liczba cech w końcowym zbiorze wyniosła 4914.

6. KLASYFIKACJA

Jako podstawowego klasyfikatora użyto metody lasów losowych (ang. *Random Forest*) [1], zaimplementowanej w oprogramowaniu Weka [3]. Eksperymentowano też z innymi klasyfikatorami, takimi jak sieci neuronowe (ang. *Neural Networks*), regresja logistyczna (ang. *Logistic Regression*), maszyna wektorów nośnych (ang. *Support Vector Machines*) i inne. Jednakże metoda lasów losowych sprawdziła się najlepiej. Jednym z wyzwań związanych ze skutecznym zastosowaniem tej metody jest wybór optymalnej liczby cech dla każdego drzewa. W przypadku konkursów dokonuje się tego na zasadzie prób i błędów. Eksperymentowano z różną liczbą cech dla jednego drzewa i dla określonego zbioru cech sprawdzała się liczba cech od 60 do 100. Najlepszy wynik uzyskano, kiedy każdy z trzech klasyfikatorów *Random Forest* miał tysiąc drzew. Liczba cech dla każdego drzewa była ograniczona do 80.

7. PODSUMOWANIE

W artykule przedstawiono zastosowanie analityki prognostycznej w oparciu o techniki eksploracji danych używane do prognozowania wybuchów metanu w kopalniach węgla. Opisany algorytm, czyli rozwiązanie nagrodzone w konkursie IJCRS'15 Data Mining Competition, jest zindywidualizowanym podejściem do klasyfikacji wielowymiarowych szeregów czasowych opracowanych w ramach innych konkursów z dziedziny eksploracji danych, które dotyczyły danych wielowymiarowych szeregów czasowych i pozwoliły na osiągnięcie bardzo dobrych wyników (wartość $AUC=0.959$). Wyniki przedstawione w artykule potwierdzają wszechstronność proponowanego podejścia, zgodnie z wynikami artykułu konkursowego [10].

Niespodziewanie ten sam podstawowy klasyfikator, czyli metoda lasów losowych, osiągnęła lepsze wyniki niż inne klasyfikatory. Taką samą prawidłowość zaobserwowano w poprzednim konkursie.

Przedstawione tu wyniki świadczą o tym, że opracowane podejście można łatwo zastosować w rozwiązywaniu innych problemów, dla których dostępne są wielokrotne pomiary w postaci szeregów czasowych.

Literatura

1. Breiman L.: *Random Forests*. „Machine Learning”, 2001, No. 1 (45), pp. 5-32.
2. Boullé M.: *Tagging Fireworkers Activities from Body Sensors under Distribution Drift*, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 389-396.
3. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H.: *The WEKA Data Mining Software: An Update*. SIGKDD Explorations, 2009, No. (1)11.
4. Hall M.A.: *Correlation-based Feature Subset Selection for Machine Learning*, Hamilton, New Zealand 1998.
5. Hanley J.A., McNeil B.J.: *A method of comparing the areas under receiver operating characteristic curves derived from the same cases*. „Radiology”, 1983, No. (3)148, pp. 839-843.
6. Janusz A., Krasuski A., Stawicki S., Rosiak M., Slezak D., Nguyen H.S.: *Key risk factors for Polish State Fire Service: A Data Mining Competition at Knowledge Pit*, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 2014, pp. 345-354.
7. Meina M., Janusz A., Rykaczewski K., Slezak D., Celmer B., Krasuski A.: *Tagging Firefighter Activities at the Emergency Scene: Summary of AIA-15 Data Mining Competition at Knowledge Pit*, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 367-373.
8. Sikora M., Sikora B.: *Improving prediction models applied in systems monitoring natural hazards and machinery*. „International Journal of Applied Mathematics and Computer Science”, 2012, No. 2(22), pp. 477-491.
9. Sikora M., Sikora B.: *Rough natural hazards monitoring*, Rough Sets: Selected Methods and Applications in Management and Engineering 2012, pp. 163-179.
10. Zagorecki A.: *A Versatile Approach to Classification of Multivariate Time Series Data*, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems 2015, pp. 407-410.

Artykuł został zrecenzowany przez dwóch niezależnych recenzentów.