

Adam Maciaszek

Apollogic

Jacek Małyszko, Milena Stróżyna

Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki
Elektronicznej

Autor do korespondencji: Milena Stróżyna, milena.strozyna@ue.poznan.pl

IDENTYFIKACJA ZAWINIĘCIA STATKU DO PORTU Z WYKORZYSTANIEM SYSTEMU AUTOMATYCZNEJ IDENTYFIKACJI (AIS)

Streszczenie: Artykuł dotyczy zagadnienia przetwarzania danych pochodzących z wiadomości AIS na potrzeby identyfikacji zdarzeń polegających na zawijaniu przez statki do portów. Identyfikacja taka może pełnić istotną funkcję w rozwijaniu świadomości sytuacyjnej na morzu różnego rodzaju służb. W artykule omówiono przeprowadzone eksperymenty, wykorzystujące dane pochodzące z systemu AIS oraz dane o wizytach w portach pozyskane z portalu MarineTraffic.com, które posłużyły do ewaluacji wypracowanego rozwiązania. Przeprowadzone badania wskazały na konieczność wypracowania wiarygodnej kolekcji danych, która mogłaby być wykorzystana w dalszych pracach w omawianym kierunku.

Słowa kluczowe: AIS, monitorowanie ruchu morskiego, analiza danych.

Klasyfikacja JEL: C88, L92.

IDENTIFICATION OF PORT CALLS BASED ON AIS MESSAGES ANALYSIS

Abstract: The article concerns the issue of the identification of port calls based on data retrieved from the AIS system. Such identification may play an important role in increasing maritime awareness, for example for the needs of coast guards. In the article we discuss experiments, which were performed and based on data from the AIS system as well as from the MarineTraffic.com portal, where the latter data was used for the purpose of the evaluation of the proposed method of port calls identification. The experiments showed that the development of reliable data collection is vital if further research in this area is to be performed.

Keywords: AIS, maritime surveillance, data analytics.

Wstęp

Transport morski pełni istotną funkcję we współczesnej gospodarce. Ponad 90% globalnego handlu jest transportowane drogą morską [IMO 2012]. W 2012 roku było to ponad 9,2 miliarda ton dóbr. Obecnie w użyciu znajduje się ponad 130 000 statków handlowych, które obsługują ponad 8000 portów na całym świecie. Handel morski odgrywa istotną rolę również w skali Unii Europejskiej, która kontroluje 40% światowego transportu morskiego [Pozo i in. 2010].

Wraz z rozwojem handlu morskiego coraz większe znaczenie przypisuje się kwestii bezpieczeństwa ruchu na morzach i oceanach. Istnieje szereg zagadnień związanych z handlem morskim, z których wynika określone ryzyko polityczne, ekonomiczne, ekologiczne czy militarne. Ryzyko takie wiąże się z trzema głównymi obszarami [Remuss 2009]:

- bezpieczeństwo portów jako infrastruktury krytycznej,
- bezpieczeństwo transportu morskiego na pełnym morzu,
- bezpieczeństwo granic morskich.

W odpowiedzi na zagrożenia związane z wymienionymi obszarami odpowiednie jednostki poszczególnych krajów (takie jak służba celna czy straż przybrzeżna) inwestują znaczne środki w poprawę swych zdolności z zakresu monitorowania ruchu morskiego. Jednym z głównych sposobów na poprawę w tym zakresie jest wykorzystanie nowoczesnych rozwiązań informatycznych w celu generowania, gromadzenia i przetwarzania odpowiednich danych, umożliwiających właściwym służbom utrzymywanie odpowiedniego poziomu

tw. świadomości sytuacyjnej na morzu [Chudoba 2016]. Stosowne rozwiązania informatyczne, mające dostęp do właściwych danych, wspomagają służby w pełnieniu ich funkcji w ramach tzw. pętli OODA (*Observe, Orient, Decide, Act*) [Angermann 2004], na którą składają się:

- obserwacja, tj. gromadzenie odpowiednich danych o sytuacji (*Observe*),
- zrozumienie zgromadzonych danych (*Orient*),
- podjęcie decyzji dotyczącej kroków, które należy przedsięwziąć w odpowiedzi na zaobserwowaną sytuację (*Decide*),
- wdrożenie decyzji (*Act*).

Artykuł prezentuje wyniki badań właśnie z zakresu podnoszenia świadomości sytuacyjnej na morzu na podstawie odpowiednich danych i ich analizy z wykorzystaniem rozwiązań informatycznych. Problemem badawczym podjętym w pracy jest automatyczna identyfikacja tego, czy dany statek zawinął do określonego portu. Danymi wykorzystywanymi w tym celu są wiadomości pochodzące z Systemu Automatycznej Identyfikacji, tzw. AIS (*Automatic Identification System*). Na podstawie danych takich jak współrzędne statku czy jego szybkość (dostarczanych właśnie przez system AIS) przygotowane rozwiązanie ma pozwolić na rozpoznawanie sytuacji, gdy statek wysyłający wiadomość przebywa w pewnym porcie. Przykładowym zastosowaniem proponowanej metody może być rozpoznawanie statków wykorzystywanych do działalności przestępczej, np. do przemytu. Wykorzystując wiedzę ekspercką dotyczącą ruchu morskiego lub opierając się na danych pochodzących z jednostek analizujących ruch morski, można wskazać porty, które są szczególnie narażone na działalność przestępczą. Informacja o tym, że określony statek niedawno zawinął do któregoś z takich podejrzanych portów, może być istotna przykładowo dla urzędów celnych, które dzięki niej będą mogły identyfikować podejrzane okręty wymagające inspekcji w celu wykrycia np. prób przemytu.

Dalsza część artykułu ma następującą strukturę. W pierwszej kolejności zostanie omówiona charakterystyka systemu AIS. Następnie przeprowadzona będzie analiza problemu badawczego ukazująca trudności z nim związane oraz zostaną przedstawione przykłady rozwiązań tego problemu prezentowane w literaturze przedmiotu. W dalszych sekcjach omówione będzie zaproponowane podejście do rozwiązania problemu badawczego, zostaną zaprezentowane dane wykorzystane do przeprowadzenia eksperymentów i ich wyniki. Przedstawiono w nich sposób pozyskania danych i wyniki ich analizy z zakresu możliwości rozpoznawania faktu zawijania do portów na podstawie ustalania wartości granicznych dla określonych danych dotyczących poruszania się statków. Przeanalizowano, jaką poprawność identyfikacji

zawijania do portu dla posiadanych danych można uzyskać w zależności od przyjętych różnych wartości granicznych. Artykuł kończy się krótkim podsumowaniem.

Artykuł w dużej mierze powstał na podstawie badań przeprowadzonych przez Adama Maciaszka podczas przygotowywania jego pracy magisterskiej. Z kolei dane wykorzystywane do eksperymentów zostały zgromadzone w ramach projektu SIMMO¹.

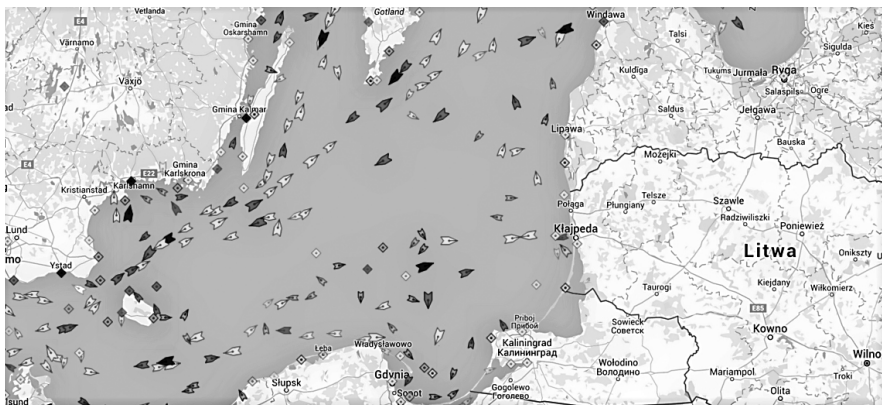
1. System Automatycznej Identyfikacji

System Automatycznej Identyfikacji został utworzony przede wszystkim jako narzędzie mające zapobiegać kolizjom na morzu. System ten działa na zasadzie automatycznej wymiany danych nawigacyjnych (m.in. o tożsamości statku, jego typie, pozycji, kursie i szybkości – patrz wizualizacja na rysunku 1) pomiędzy statkami i innymi urządzeniami. Nadajniki AIS są obowiązkowym wyposażeniem wszystkich jednostek o rozmiarze co najmniej 300 BRT (ton rejestrowych brutto). W systemie AIS wykorzystywane jest pasmo VHF, dzięki któremu wiadomości pomiędzy statkami mogą być wymieniane w odległości do około 20–30 mil morskich. Pozwala to na pozyskanie informacji o ruchu statków na danym obszarze również w przypadku braku kontaktu wzrokowego.

Wysyłane sygnały AIS mogą być odbierane nie tylko przez inne statki, ale również przez stacje lądowe usytuowane na wybrzeżu. Dzięki temu możliwe jest monitorowanie ruchu statków poruszających się w pobliżu lądu. W końcu dane AIS mogą być odbierane również przez satelity operujące na niskiej orbicie. Satelity takie, poruszające się na wysokości około 500 kilometrów, posiadają zasięg odbioru sygnałów z obszaru o średnicy około 3000 kilometrów.

Ponieważ dane pochodzące z AIS pozwalają na ustalenie pozycji statku w określonym momencie, odpowiednio przeprowadzona analiza tych danych powinna pozwolić na identyfikację portów, do których dany statek zawinął w określonym czasie. W związku z tym, jeśli statek ten zawinął do podejrzanych portów, powinno być możliwe również identyfikowanie takich sytuacji. Konieczne jest tutaj dysponowanie odpowiednimi danymi pokrywającymi

¹ Projekt SIMMO (*System for Intelligent Maritime Monitoring*) realizowany przez konsorcjum Uniwersytet Ekonomiczny w Poznaniu i LuxSpace Sarl w ramach międzynarodowego programu badawczo-rozwojowego JIP-ICET 2, finansowanego przez kraje członkowskie Europejskiej Agencji Obrony.



Rysunek 1. Wizualizacja wiadomości AIS dla Morza Bałtyckiego z dnia 20.05.2016 (10:25)

Źródło: MarineTraffic.com.

najlepiej całą kulę ziemską, których analiza mogłaby pozwolić na osiągnięcie założonego celu. Takie dane mogą być pozyskane od jednego z dostawców, takich jak np. LuxSpace (<http://www.luxspace.lu>), którzy dostarczają dane AIS z satelitów i stacji nabrzeżnych.

2. Analiza problemu badawczego

Przy użyciu danych AIS możliwe jest analizowanie pozycji statków w określonym momencie. System AIS cechuje jednak szereg niedoskonałości, utrudniających przeprowadzanie wnioskowań na podstawie danych z niego pozyskiwanych. Warto je w tym momencie przeanalizować.

Po pierwsze, wiele wiadomości AIS nie jest w żaden sposób przechwytywanych ani rejestrowanych. Dotyczy to przede wszystkim wiadomości wysyłanych na pełnym morzu, w oddaleniu od stacji lądowych. Jedynym sposobem na pozyskanie tych danych jest wykorzystanie satelitów. Jednakże pokrycie kuli ziemskiej zasięgiem satelitów działających w ramach systemu AIS aktualnie nie jest pełne; satelity poruszają się po określonym torze i z określonymi odstępami między sobą. Często różne obszary Ziemi mogą nie być pokryte przez wiele godzin, ponieważ żaden z satelitów akurat nie będzie się znajdować wystarczająco blisko. Dodatkowo liczba wiadomości odbieranych przez satelity w wielu wypadkach przekracza możliwości techniczne zamontowanych w nich

odbiorników, co powoduje, że wiele takich wiadomości nie jest prawidłowo rejestrowanych.

Kolejną trudnością z wykorzystaniem danych z systemu AIS w opisywanej pracy jest fakt, że nadajniki AIS w łatwy sposób mogą być wyłączane przez załogę statku. Powinno się to zdarzać tylko w wyjątkowych sytuacjach, jednak decyzja o takim wyłączeniu pozostaje w gestii kapitana statku. W związku z tym istnieje możliwość, że statek przed zawinięciem do portu, jeśli będzie chciał ukryć ten fakt, może wyłączyć nadajnik i włączyć go dopiero po wypłynięciu z portu. W takiej sytuacji nie można jednoznacznie stwierdzić, czy statek ten naprawdę był w danym porcie, czy też nie. Problem ten nie może być obecnie w żaden sposób rozwiązany. Pośrednim sposobem na zapewnienie, że statki uciekające się do takich praktyk i tak zostaną zidentyfikowane jako potencjalnie niebezpieczne, jest wykrywanie sytuacji nieoczekiwanego zniknięcia na dłuższy czas sygnału AIS określonego statku, co może być skutkiem takiego właśnie wyłączenia nadajnika AIS. Zadanie to utrudnia jednak niedoskonałość satelitarnych systemów monitorowania AIS, co zostało opisane w poprzednim akapicie (zniknięcie sygnału może wynikać również z technicznych niedoskonałości systemu).

Zakładając, że opisane powyżej trudności zostaną rozwiązane i do dyspozycji będą kompletne dane, należy jednak zauważyć, że stwierdzenie, czy statek zawinął do pewnego portu, czy też nie, na podstawie danych AIS może być problematyczne. Wynika to m.in. z tego, że porty często nie mogą być traktowane jako punkty czy też obszary o ściśle określonych granicach. Dodatkowo często kilka portów może być w na tyle niewielkiej odległości od siebie, że stwierdzenie, do którego z nich statek faktycznie zawinął, również może nastroczać pewnych trudności. Określenie uniwersalnej odległości granicznej pomiędzy statkiem a lokalizacją portu, poza którą statek byłby uznawany za przebywający „poza portem”, a wewnątrz której statek byłby uznany jako przebywający w porcie, nie jest prawdopodobnie możliwe.

Poza tym problem badawczy analizowany w artykule jest komplikowany dodatkowo przez trudności w przygotowaniu wiarygodnego korpusu danych o faktycznych zawinięciach statków do portów, co odgrywa kluczową rolę w przeprowadzeniu ewaluacji opracowywanego rozwiązania. Jak opisano w poprzednim akapicie, manualne przygotowanie takiego korpusu byłoby bardzo czasochłonne i prawdopodobnie dałoby niedoskonałe wyniki. Innym podejściem może być pozyskanie dodatkowych danych, w których znajdują się informacje o wizytach statków w portach. Dane takie dostępne bywają przykładowo na różnych witrynach internetowych. Nie jest jednak możliwe w prosty

sposób oszacowanie kompletności tak pozyskanych danych. W związku z tym może się okazać, że chociaż pewien statek istotnie zawinął do określonego portu, nie jest to w żaden sposób odwzorowane w danych w ten sposób publikowanych. Utrudnia to bardzo przeprowadzenie wiarygodnej analizy danych, która mogłaby skutkować wypracowaniem w pełni poprawnego modelu pozwalającego na podejmowanie decyzji.

Dane o zawinięciach do portów można próbować dodatkowo uzupełnić, wykorzystując w tym celu dane o porcie docelowym i estymowanym czasie przybycia wysyłane w ramach AIS. Załogi statków powinny deklarować te dane w wiadomościach AIS. Jeśli więc statek zapowiada, że płynie do pewnego portu i dotrze tam o określonej porze, z pewnym prawdopodobieństwem można dodać taką informację do korpusu danych o zawinięciach do portów. Jednakże dostępność i prawdziwość tych danych zależy tylko od dobrej woli załogi danego statku.

3. Identyfikacja zawinięcia statku do portu

Problem identyfikacji zawinięcia statku do portu przy użyciu danych z AIS jest relatywnie nową koncepcją. Informacja o planowanym zawinięciu statku do portu jest przekazywana do portu bezpośrednio przez armatora lub agenta poprzez system awizacji zawinięć [Salomon 2002]. Następnie informacje te są weryfikowane przez władze portu dzięki obserwacji i ręcznej analizie danych z systemu AIS, radarów i kamer, kiedy statek rzeczywiście zawinął do portu. Proces ten może być częściowo wspierany przez systemy informatyczne. Jednakże istniejące automatyczne metody, opisane w literaturze przedmiotu, dotyczą w głównej mierze problemu identyfikacji statków w danym obszarze (np. w porcie) poprzez analizę obrazów radarowych. Dotyczy to zarówno obrazów z tradycyjnych radarów, jak i obrazów satelitarnych z radarów z syntetyczną aperturą (tzw. SAR) [Crisp 2004; Tello, López-Martínez i Mallorqui 2005; Zhu i in. 2010]. Wraz z upowszechnieniem się systemu AIS wyniki zwracane przez te metody są dodatkowo porównywane z danych pochodzącymi z AIS, głównie w celu weryfikacji ich poprawności [Grasso i in. 2009]. Kolejne badania w tym zakresie dotyczą wykrywania ruchu statków na podstawie analizy obrazów wideo przedstawiających sytuację w porcie [Bao, Zinger i Wijnhoven 2013]. Dane z systemu AIS są również wykorzystywane do analizy ruchu statków w porcie i wykrywania wzorców poruszania się statków [Fooladvandi i in. 2009; Tun i in. 2007].

4. Propozycja metody do identyfikacji zawinięcia statku do portu

Wykrywanie, czy statek był w porcie, może być przeprowadzone z wykorzystaniem różnych podejść do analizy danych dotyczących geolokalizacji. Poniżej zaprezentowano dwie propozycje osiągnięcia tego celu:

1. Identyfikacja faktu zawinięcia do portu na drodze badania odległości pomiędzy statkiem a portem oraz szybkości, z jaką się statek porusza. Mając współrzędne statku oraz współrzędne portów, można obliczyć odległości pomiędzy statkiem a poszczególnymi portami. Na podstawie takiej odległości i szybkości statku można porównać te wartości z pewnymi wartościami granicznymi, ustalonymi przykładowo przez eksperta. Dodatkowo wartości te mogą się różnić dla różnych portów. Wadą tego rozwiązania jest to, że dla każdego statku istnieje konieczność określenia jego odległości od wielu tysięcy portów, co może być bardzo kosztowne obliczeniowo.
2. Poprzez analizę geolokalizacji statku bez brania pod uwagę danych o portach, a jedynie z uwzględnieniem wcześniej znanych współrzędnych statków, o których wiadomo, że zawinęły do danego portu. W takiej sytuacji, jeśli statek znajduje się w odpowiednim miejscu, może być rozpoznany jako przebywający aktualnie w pewnym porcie. W tym przypadku możemy mieć do czynienia z dwoma wariantami metody:
 - wykorzystany zostanie klasyfikator binarny, rozpoznający jedną z dwóch sytuacji: statek albo znajduje się w porcie, albo nie, bez rozstrzygania, o który port chodzi; wówczas, w przypadku rozpoznania, że statek był w pewnym porcie, po etapie takiej klasyfikacji należy jeszcze przeprowadzić analizę mającą na celu przypisanie właściwego portu do danego statku i momentu,
 - model klasyfikujący do jednej z wielu klas, gdzie każda klasa odpowiada jednemu portowi oraz istnieje dodatkowa klasa reprezentująca brak zawinięcia statku do portu.

W opisywanym podejściu problematyczne jednak jest to, że rozwiązanie takie będzie mogło działać tylko dla tych portów, dla których są dostępne dane o tym, że określone statki do nich zawijały. Jeśli z jakiegoś powodu nie dysponujemy takimi danymi, nie będziemy też dysponowali żadnymi przykładami uczącymi, które można by wykorzystać na etapie konstruowania klasyfikatora.

W dalszych sekcjach zaprezentowany zostanie eksperyment dla pierwszego z opisanych powyżej podejść.

5. Dane i ich przygotowanie

Podstawowymi danymi wykorzystanymi w opracowywanym rozwiązaniu są wiadomości AIS, dostarczone przez firmę LuxSpace. Składają się na nie m.in. dane pozwalające na identyfikację statku (np. numer MMSI – *Maritime Mobile Service Identity*, morski numer nawigacyjny), jego typu oraz dane określające jego geolokalizację, kurs i szybkość (względem dna).

Dodatkowym zasobem, który jest niezbędny do przeprowadzenia zaplanowanych badań, jest kolekcja danych dotyczących zawinięć do portów. Innymi słowy, konieczne jest dysponowanie korpusem danych o tym, że pewien statek o określonej porze zawinął do określonego portu. Dane takie konieczne są po pierwsze na etapie wypracowywania rozwiązania postawionego problemu, np. na drodze wstępnej analizy danych. Wymagane byłyby też na potrzeby automatycznego konstruowania odpowiednich modeli statystycznych na drodze uczenia maszynowego. Dodatkowo dane takie są niezbędne na etapie ewaluacji rozwiązania, tj. oceny, w jakim stopniu trafnie wypracowana metoda pozwala na identyfikację zawinięć statków do portów. W celu stworzenia takiego korpusu wykorzystano dane pobrane w sposób automatyczny z portalu internetowego MarineTraffic.com. Portal ten jest głównym źródłem danych dotyczących ruchu statków, w tym zawinięć do portów². Trudno jednak oszacować kompletność udostępnianych przez ten portal danych; przeglądając jego zawartość, dość szybko się zauważy, że dla wielu statków portal ten nie udostępnia żadnych danych dotyczących ich zawinięć do portów, co prawdopodobnie wynika po prostu z braku posiadania odpowiednich danych przez sam portal. Dodatkowo wykorzystanie takich danych wymaga przeprowadzenia integracji danych z dwóch różnych źródeł: konieczne jest tu połączenie wiadomości AIS z danymi pobranymi z portalu MarineTraffic.com w taki sposób, aby scalone dane faktycznie dotyczyły tych samych statków.

Pomimo niedoskonałości danych z portalu MarineTraffic.com i konieczności przeprowadzenia ich integracji podjęto decyzję o ich wykorzystaniu, przede wszystkim ze względu na liczne trudności związane z innymi podejściami do przygotowania odpowiedniego korpusu danych. Innym rozwiązaniem mogłaby być ręczna adnotacja, jednak byłaby ona bardzo czasochłonna i prawdopodobnie również dałaby niedoskonałe rezultaty. Poza tym przeprowadzenie badań z wykorzystaniem dostępnych danych może w pewnym stopniu pozwolić również na ocenę kompletności samego źródła i wskazać, czy w przyszłych badaniach dane te warto wykorzystywać ponownie.

² <http://www.sailmagazine.com/gear/electronics-and-navigation/cautiously-optimistic-on-marinetraffic-internet-ais/>.

Dane zostały pobrane z wykorzystaniem skryptu napisanego w języku Python, monitorującego zawartość stron w ramach portalu MarineTraffic.com. Pobrane dane zawierały m.in. informacje o MMSI statku oraz czasie zawinięcia do portu i wypłynięcia z niego, i według tych informacji zostały one połączone z wiadomościami z bazy danych AIS. Dodatkowo dla różnych statków dane z portalu MarineTraffic.com pochodzą z różnych okresów, a to ze względu na długi czas konieczny do automatycznego pobrania danych dla wszystkich statków. Skrypt pobierający dane odwiedzał kolejne strony z danymi o statkach (na każdej stronie były dane o jednym statku) z odstępem pięciu sekund, co przy liczbie statków sięgającej setek tysięcy spowodowało właśnie, że dla różnych statków dane o nich pobierano z momentów odległych od siebie o wiele dni. Konieczne było więc dodatkowe filtrowanie analizowanych wiadomości AIS w taki sposób, aby pochodziły one z okresu, dla którego mamy dane dla konkretnego statku. Przyjęto w tym miejscu, że wiadomość AIS musiała pochodzić z okresu pięciu dni poprzedzających datę, w której pobrano dane o statku z portalu MarineTraffic.com (inaczej wiadomość o zawinięciu statku do konkretnego portu mogła już zniknąć z portalu). Wiadomości z bazy AIS, dla których znaleziony został odpowiadający rekord, czyli taki, gdzie MMSI statku jest takie samo, a wysłanie wiadomości mieści się w przedziale czasu pomiędzy zawinięciem i wypłynięciem z portu, zostały zaliczone do grupy „statek w porcie”. Pozostałe wiadomości (czyli takie, które pochodziły z pięciu dni przed pobraniem danych o danym statku z portalu i dla których w portalu nie było informacji o tym, że statek ten zawinął do portu w określonym momencie) zostały zaliczone do grupy „statek poza portem”. Tak powstał zbiór danych, który mógł następnie stanowić podstawę do przygotowywania i ewaluacji opracowywanych metod.

W poniższym badaniu analizowano aktywność statków w okresie od 1.03.2015 roku do 20.03.2015 roku. Badanie miało na celu stwierdzenie, czy statek znajduje się w pewnym porcie. Na potrzeby analizy wybranych zostało sześć portów – po jednym największym porcie na każdym kontynencie z uwagi na wielkość przeładunku w 2013 roku według American Association of Port Authorities: Szanghaj (Chiny, Azja), Rotterdam (Holandia, Europa), Nowy Orlean (Stany Zjednoczone Ameryki, Ameryka Północna), Richards Bay (Republika Południowej Afryki, Afryka), Port Hedland (Australia), Itaqi (Brazylia, Ameryka Południowa). W przypadku wybranych portów z reguły trudniej jest zidentyfikować, czy statek faktycznie do nich zawinął, ze względu na ich bardzo duże rozmiary (przykładowo dwa statki oddalone od siebie o wiele kilometrów mogą, na podstawie posiadanych danych, ciągle znajdować się w tym samym porcie) i mnogość mniejszych portów, które je zazwyczaj otaczają.

Zbiorem danych w eksperymencie były wszystkie wiadomości z bazy danych AIS przesłane przez statki w promieniu 50 kilometrów od każdego z analizowanych portów we wspomnianym wcześniej okresie. Wiadomości AIS odpowiadających temu kryterium było około jednego miliona. Dla 50 194 rekordów udało się znaleźć dane pochodzące z portalu MarineTraffic.com o tym, że dany statek w momencie, którego dotyczyła wiadomość AIS, znajdował się w porcie. Spośród pozostałych rekordów rozlosowano taką samą liczbę wiadomości AIS (ponownie 50 194), które następnie traktowano jako wiadomości nieodpowiadające zawinięciu do portu. Co prawda założenie to, jak wcześniej wspomniano, nie może być traktowane jako w pełni prawdziwe, niemniej jako jedyne powinno dać możliwość choć orientacyjnej ewaluacji rozwiązania. Łącznie próba wynosiła więc 100 338 rekordy. Z nich na potrzeby eksperymentu losowo wybrano 30% ze względu na duży rozmiar danych i dużą złożoność obliczeń (obliczenia mogły trwać tu wiele godzin) tak, aby przyspieszyć czas uzyskiwania rezultatów.

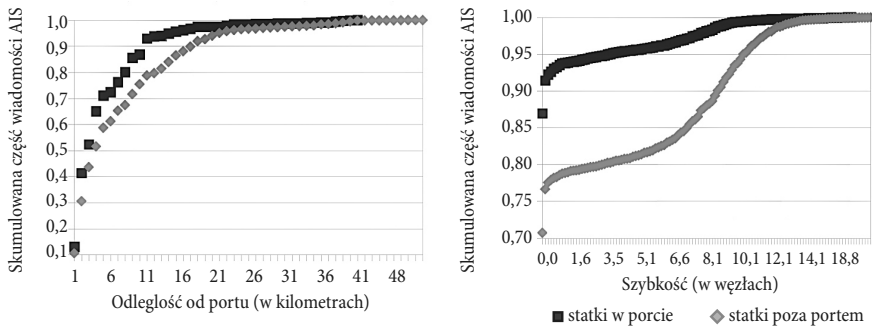
6. Przeprowadzone eksperymenty

Jak wspomniano, podstawowym podejściem do identyfikacji zawinięć do portów może być porównywanie wybranych danych pozyskanych z wiadomości AIS, takich jak odległość od portu i szybkość statku, z ustalonymi wartościami granicznymi. W pierwszej kolejności przeprowadzono analizę rozkładów tych dwóch wartości (odległość od portu i szybkość statku) dla obu populacji: wiadomości AIS, o których wiemy, że pochodzą ze statków w porcie, oraz takich, które traktujemy jako poza portem (w portalu MarineTraffic.com nie było danych o tym, aby dany statek był w danym momencie w porcie). Aby wartości te pozwalały na prawidłową klasyfikację wiadomości, konieczne byłoby tu wystąpienie odpowiednio dużej różnicy w ich rozkładach w obu populacjach.

Skumulowane rozkłady obu wartości dla obu statystyk przedstawiono na rysunku 2. Możliwe jest zaobserwowanie pewnych różnic w obu przypadkach i są one zgodne z oczekiwaniami. Odległość statków od portu w analizowanych danych jest generalnie większa dla statków, które według danych są poza portem, w porównaniu z odległością dla statków przebywających w porcie. Podobnie wygląda to w przypadku szybkości statków. Należy jednak zauważyć, że te różnice są niewielkie.

Przykładowo, gdyby klasyfikować fakt zawinięcia statku do portu tylko na podstawie odległości od portu, to przyjmując jako wartość graniczną

12 kilometrów, jako odpowiadające wizycie w porcie zostałyby zaklasyfikowanych niemal 90% statków faktycznie przebywających w porcie, ale jednocześnie około 80% wiadomości od statków, które w nim nie przebywały. Podejście takie pozwoliłoby więc na uzyskanie wysokiej pełności, ale przy niskiej precyzji, tylko nieznacznie przekraczającej 50%. Potwierdzają to wyniki oceny precyzji i pełności rozwiązania dla różnych wartości granicznych odległości od portu i szybkości statku, co zostało zaprezentowane na rysunku 2.



Rysunek 2. Skumulowane rozkłady odległości od portu oraz szybkości dla statków pozostających w porcie i, zgodnie z posiadanymi danymi, przebywających poza nim

Niewątpliwie wpływ na to może mieć niska pełność danych pozyskanych z portalu MarineTraffic.com dotyczących zawinięć do portów – statki, które metoda może zaklasyfikować jako przebywające w porcie, być może są w nim również wtedy, gdy w portalu MarineTraffic.com nie ma danych pozwalających na potwierdzenie tego.

Dotychczas analizowano jedynie, czy metoda prawidłowo wykrywa sam fakt zawinięcia statku do portu, nie sprawdzano natomiast, czy metoda prawidłowo identyfikuje, o który port chodzi. Zadanie to może być problematyczne ze względu na to, że w wielu przypadkach istnieje wiele portów w niewielkiej odległości od siebie, jak to się dzieje przykładowo z portami Dordrecht oraz Rotterdam. Pomimo prawidłowego przyporządkowania wiadomości do grupy „w porcie” ciągle możliwe jest nieprawidłowe określenie samego portu. W kolejnym eksperymencie oceniono więc, czy metoda może prawidłowo określać takie porty.

Do wiadomości AIS przypisywano ten port, który znajdował się najbliżej podanych w wiadomości współrzędnych statku. Aby określić poprawność

takiego przypisania, porównywano wykryty port z portem zadeklarowanym dla tego statku w portalu MarineTraffic.com. Wartość ta wahała się między 85% a 90% w zależności od wartości granicznych przyjętych podczas klasyfikacji wiadomości AIS jako przebywających w porcie bądź nie.

W podsumowaniu wyniki wstępnej analizy danych pozwalają stwierdzić, że podstawowym problemem w analizowanym obszarze jest zebranie odpowiednio wiarygodnych danych. W pewnym stopniu dane, które były analizowane w przeprowadzonych eksperymentach, zgadzają się z intuicją mówiącą, że statki będące blisko portu i nieporuszające się, prawdopodobnie do niego zawinęły, ale można było się spodziewać, że takie wnioskowanie będzie poprawne w większej liczbie przypadków. Prawdopodobnie więc wiele faktów zawinięcia statków do portów nie zostało odnotowanych w źródle, z którego dane pochodziły (a należy pamiętać, że było to jedno z największych ogólnodostępnych źródeł podobnych danych).

Zakończenie

W artykule przebadano możliwość rozpoznawania (na podstawie danych z systemu AIS) faktu zawijania statku do portu oraz jego identyfikacji. W artykule określono problemy związane z pozyskaniem odpowiednich danych do nauki klasyfikatora oraz na potrzeby jego ewaluacji. Wykazano, że nawet dane pozyskane z przodującego portalu publikującego dane o statkach (MarineTraffic.com) mają prawdopodobnie niską pełność w tym zakresie. Zaproponowano również metodę identyfikacji faktu zawinięcia statku do portu, biorącą pod uwagę jego odległość od danego portu oraz szybkość, z jaką się porusza. Podejście takie pozwala, przy ustaleniu odpowiednich wartości granicznych, na otrzymanie wysokiej pełności, jednak precyzja dla posiadanych danych nie przekracza progu 60%, co należy uznać za niski wynik. Prawdopodobnie jest to spowodowane niewielką pełnością danych wykorzystywanych na potrzeby ewaluacji (tj. pobranych z portalu MarineTraffic.com).

Dalsze prace w omawianym zakresie powinny przede wszystkim zmierzać w kierunku wygenerowania korpusu zaadnotowanych danych o wysokiej pełności. Dane z portalu MarineTraffic.com mogą posłużyć jako jedno ze źródeł danych dla takiego korpusu. Jednakże konieczne byłoby uzupełnienie go o dane dodatkowe, np. dane przesyłane w wiadomościach AIS o tym, do jakiego portu dany statek płynie (deklarowany cel podróży) i jaki jest oczekiwany czas przybycia. Być może istnieją również inne źródła, dzięki którym możliwe byłoby pozyskanie danych potrzebnych na etapie konstruowania

takiego korpusu. W ostateczności dane mogą być również adnotowane ręcznie przez ekspertów, co jednak wiązałoby się ze znacznym nakładem pracy.

Posiadanie takiego korpusu umożliwiłoby przeprowadzenie bardziej wiarygodnej ewaluacji zaproponowanej metody. Możliwe byłoby również wypracowanie nowych metod identyfikacji faktu zawijania do portów z wykorzystaniem wybranych podejść uczenia maszynowego. Metody te, dysponując odpowiednim korpusem uczącym, mogłyby pozwolić na automatyczną konstrukcję klasyfikatora o wyższej skuteczności niż metoda zaproponowana w przedstawionej pracy.

Bibliografia

- Angerman, W., 2004, *Coming Full Circle with Boyd's OODA Loop Ideas: An Analysis of Innovation Diffusion and Evolution (Technical Report)*, DTIC Document.
- Bao, X., Zinger, S., Wijnhoven, R., 2013, *Ship Detection in Port Surveillance Based on Context and Motion Saliency Analysis*, in: *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, s. 1–8.
- Chudoba, R., 2016, *Wojskowa świadomość sytuacji na morzu*, Przegląd Sił Zbrojnych, nr 2, s. 116–119.
- Crisp, D.J., 2004, *The State-of-the-art in Ship Detection in Synthetic Aperture Radar Imagery*, Defence Science and Technology Organisation Salisbury (Australia) Info Sciences Lab.
- Fooladvandi, F., Brax, C., Gustavsson, P., Fredin, M., 2009, *Signature-based Activity Detection Based on Bayesian Networks Acquired from Expert Knowledge*, in: *12th International Conference on Information Fusion*, Seattle, WA, s. 436–443.
- Grasso, R., Mirra, S., Baldacci, A., Horstmann, J., Coffin, M., Jarvis, M., 2009, *Performance Assessment of a Mathematical Morphology Ship Detection Algorithm for SAR Images through Comparison with AIS Data*, in: *Ninth International Conference on Intelligent Systems Design and Applications*, Pisa, s. 602–607.
- IMO – International Maritime Organisation, 2012, *International Shipping Facts and Figures – Information Resources on Trade, Safety, Security, Environment (Technical Report)*, International Maritime Organization, London.
- Pozo, F., el, Dymock, A., Feldt, L., Hebrard, P., Sanfelice di Monteforte, F., 2010, *Maritime Surveillance in Support of CSDP (Technical Report)*, European Defence Agency, Brussels.
- Remuss, N., 2009, *Space and Internal Security – Developing a Concept for the Use of Space Assets to Assure a Secure Europe (Technical Report)*, ESPI Reports 20, European Space Policy Institute, Vienna.

- Salomon, A., 2002, *EDI (Electronic Data Interchange) w polskich portach morskich*, Inżynieria Morska i Geotechnika, no. 1, s. 41–48.
- Tello, M., López-Martínez, C., Mallorqui, J.J., 2005, *A Novel Algorithm for Ship Detection in SAR Imagery Based on the Wavelet Transform*, Geoscience and Remote Sensing Letters, vol. 2, no. 2, s. 201–205.
- Tun, M.H., Chambers, G.S., Tan, T., Ly, T., 2007, *Maritime Port Intelligence Using AIS Data*, Recent Advances in Security Technology, no. 33, s. 33–43.
- Zhu, C., Zhou, H., Wang, R., Guo, J., 2010, *A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features*, Transactions on Geoscience and Remote Sensing, vol. 4, no. 9, s. 3446–3456.