Studium Magisterskie

Kierunek: Metody ilościowe w ekonomii i systemy informacyjne
Specjalność: Modele i metody decyzyjne

Bartosz Zawieja
nr albumu: 77010

Optymalizacja aktywnych inwestycji giełdowych przy użyciu metod uczenia ze wzmacnianiem.

Praca magisterska
pod kierunkiem naukowym
prof. dr. hab. Tomasza Szapiro
Zakład Wspomagania i Analizy Decyzji

Warszawa 2022

Spis treści

1 Wprowadzenie

Celem pracy była konstrukcja procedury wspierania decydenta w problemie optymalizacji aktywnych inwestycji giełdowych. Przedstawiona argumentacja ekonomiczna zdecydowała o sprecyzowaniu tak postawionego wyzwania. Przytaczane badania empiryczne dowodziły, że uzasadnione było wnioskowanie o dodatniej korelacji między progresem gospodarczym, a rozwojem rynku akcji, por. Rozdział 2. Z tego względu cel pracy przeformułowano do problemu optymalizacji struktury zarządzanego aktywnie portfela tego rodzaju walorów.

Rozważany problem badawczy w odniesieniu do skali makroekonomicznej warunkował sprawność dokonywanej przez rynek finansowy alokacji kapitału – uznawanej przez ekspertów za najważniejszą z jego funkcji, por. Stiglitz (1989). W skali mikroekonomicznej tożsamy był z wyrazem efektywności działania rynków poszczególnych produktów. Determinował on także bezpośrednio ilość użyteczności znajdującej się w zasięgu poszczególnych konsumentów. Z tych wszystkich powodów można było go uznawać za istotny.

Argumenty z zakresu finansów zdecydowały o tym, że do problemu optymalizacji struktury zarządzanego aktywnie portfela akcyjnego zdecydowano się odnieść w sposób zgodny z paradygmatem analizy technicznej. Takie podejście stało w zgodzie z teorią dziedziny, por. Rozdział 2. Decydowało ono o skalowalnym i kwantyfikowalnym charakterze wykorzystywanych metod. Z uwagi na bieżące trendy wydawało się również innowacyjne. Ze względu na – realizujący się szczególnie dynamicznie w ostatnich latach – duży potencjał rozwojowy sztucznej inteligencji, w pracy podjęto decyzję o wykorzystaniu metod uczenia ze wzmacnianiem.

Analizując dotykające podobnych problemów badawczych artykuły naukowe podjęto decyzję o wyborze metod wykorzystywanych przy realizacji obranego w pracy celu. Tym sposobem zdecydowano o optymalizacji przy użyciu programów implementujących w środowisku problemu algorytmy Advantage Actor Critic (A2C) i Proximal Policy Optimization (PPO) decydujące o modyfikacjach sieci neuronowych typu Long short-term memory (LSTM).

Przy realizacji procedury badawczej istotne było osiągnięcie maksymalnej w ramach dostępnych mocy obliczeniowych optymalności względem rozpatrywanego zbioru danych. Zdecydowano o wykorzystaniu notowanych w interwale dziennym informacji o cenach akcji spółek reprezentujących jak najlepiej całe spektrum światowej gospodarki.

Po niniejszym Wprowadzeniu, w Rozdziale 2. zamieszczono, zbudowane na podstawie przeglądu literatury, argumenty ekonomiczne i finansowe wprowadzające w istotę rozpatrywanego problemu badawczego.

Rozdział 3. przedstawia wykorzystane w pracy metody badawcze – w szczególności algorytmy uczenia ze wzmacnianiem. Obejmuje opis historii ich rozwoju oraz matematyczno-programistyczny schemat działania.

Rozdział 4. zawiera informacje dotyczące zbioru danych oraz implementacji metod. Przedstawiono w nim również otrzymane wyniki, ich interpretację i estymację ograniczeń wykonanej pracy.

W Rozdziale 5. umieszczono uwagi końcowe i komentarze. Dotyczą one dostrzeganych możliwości kontynuacji podjętych rozważań badawczych i wynikają z szacowanych sposobów wykluczenia ich niedoskonałości.

Kolejne rozdziały poświęcono odpowiednio: spisowi cytowanej literatury (Rozdział 6 – Bibliografia) oraz wykazom rysunków, tabel i kodów (Rozdziały 7, 8 i 9). Pracę zamyka streszczenie (Rozdział 10).

2 Rynek finansowy w ujęciu ekonomiczno-analitycznym

W rozdziale zawarto teoretyczno-historyczne fundamenty rozważań podejmowanych w dalszych etapach pracy. Składa się on z 2 części. Pierwsza z nich – Podrozdział 2.1. – dotyczy ekonomicznego uzasadnienia istotności gospodarczej rozpatrywanego problemu badawczego. Przedstawia argumentację przemawiającą za istnieniem pozytywnego wpływu, utożsamianego z optymalną alokacją kapitału, efektywnego rynku akcji na rozumiany wielowymiarowo rozwój gospodarczy. Podrozdział 2.2. odnosi się do możliwych w ujęciu finansowym sposobów podejścia do problemu badawczego. Dla paradygmatów analizy fundamentalnej oraz technicznej rozważono w nim kwestie dotyczące: poprawności założeń oraz istnienia trwałej możliwości osiągania ponadprzeciętnego zysku.

2.1 Znaczenie rynku finansowego dla gospodarki

Kwestia znaczenia rynku finansowego w życiu ekonomicznym społeczeństwa jest podejmowana w bardzo wielu publikacjach. Na poziomie ogólnym stanowi względnie abstrakcyjny przedmiot zainteresowania filozofów i naukowców. Warunki gospodarki wolnorynkowej decydują jednak o tym, że poza sferami teoretycznymi, rozpatrywana jest również w wielu kontekstach czysto empirycznych. Z makroekonomicznego punktu widzenia, jest to praktyczny problem instytucji władzy ustawodawczej, organów regulujących i organizacji międzynarodowych. Równie namacalne, na poziomie mikroekonomicznym, jest oddziaływanie rynku finansowego na wszystkie podmioty gospodarcze z osobna.

Historia rynków finansowych może w pewnym sensie rozpoczynać się wraz z historią pieniądza w gospodarce. Podstawowe koncepcje przepływów pieniężnych i pożyczek były znane i praktycznie wykorzystywane już w czasach cywilizacji starożytnych. Pierwsze obligacje emitowano około 2400 lat przed naszą erą, na terenie Mezopotamii. Za ich pośrednictwem wystawca zobowiązywał się do odroczonej w czasie płatności w zbożu, por. Cummans (2014). Organizacje oferujące podstawowe usługi bankowe istniały już wtedy, jednak pierwszy, w pełni funkcjonalny bank powstał dopiero w dwunastowiecznej Wenecji (w roku 1156). W tamtym okresie, we Włoszech rozpowszechniał się napędzany rozwojem handlu obrót wekslami[1].

[1] Początkowo ekspansja działalności pożyczkowej na starym kontynencie wstrzymywana była przez kościół katolicki utożsamiający ją z lichwą i uznający za niemoralną. Stanowisko religii w tej sprawie zliberalizował jednak już w trzynastym wieku Tomasz z Akwinu, por. Kaźmierczak (1993).

Pierwsze akcje wyemitowano w 1603 roku, a więc dopiero około czterech tysięcy lat po emisji pierwszych obligacji. Wydarzenie było wyrazem ogromnego ciężaru gospodarczego projektu Holenderskiej Kompanii Wschodnioindyjskiej. Spółka dysponowała gwarantowanym przez władze monopolem na działalność kolonialną w Azji, por. Gelderblom i in. (2013).

Na przestrzeni siedemnastego wieku założone zostały kolejno: Bank Amsterdamski, Bank Szwecji oraz Bank Anglii dzielące między sobą zależnie od przyjętej definicji w różnych proporcjach miano pierwszego w historii banku centralnego, por. Quinn i Roberds (2006), Crowe i Meade (2007), Roseveare (1991). Okres ten można wskazać jako początek funkcjonowania pewnej formy polityki monetarnej. W dalszych latach swobodnemu rozwojowi rynku finansowego sprzyjała demokratyzacja społeczeństw, rozwój liberalnego kapitalizmu oraz następujące po sobie rewolucje przemysłowe. Poważnych wątpliwości dotyczących słuszności tego procesu dostarczyły dopiero kryzysy dwudziestego wieku.

Powstanie spójnego, globalnego systemu monetarnego możliwe było dzięki spotkaniu w Bretton Woods z 1944 roku[2]. Zorganizowany wtedy schemat przestał obowiązywać ostatecznie w latach siedemdziesiątych dwudziestego wieku wraz z zawieszeniem wymienialności dolara na złoto. Od różnych dat końca dwudziestego wieku, w zależności od różnych poddziedzin rynku finansowego liczy się natomiast początek jego epoki współczesnej.

Wspominane w literaturze gospodarcze zadania współczesnego rynku finansowego to między innymi sprawny obrót kapitałowy, wycena projektów gospodarczych i stanie w roli barometru koniunktury rynkowej, por. Dębski (2014). Co jednak zaznacza amerykański noblista, Joseph Stiglitz (1989), zdecydowanie najważniejszą funkcją rynku finansowego w gospodarce jest alokacja kapitału. Z tego względu w dalszej części tekstu będzie on traktowany równoznacznie z mechanizmem pełniącym tę rolę. Efektywne działanie rynku finansowego utożsamiane będzie z efektywną alokacją kapitału.

Paul Samuelson stworzył powszechnie akceptowaną definicję problemu rzadkości, określanego czasem też problemem niedostatku, który: *„…odnosi się do podstawowego faktu przyrodniczego, że istnieje tylko skończona ilość zasobów ludzkich i pozaludzkich, które najlepsza wiedza techniczna jest w stanie wykorzystać do wytworzenia jedynie ograniczonych maksymalnych ilości każdego dobra gospodarczego…”*, por. Samuelson i Nordhaus (1980; tłumaczenie własne). Lionel Robbins, zdefiniował natomiast przy użyciu tego pojęcia

[2] System z Bretton Woods przyniósł ujednolicenie światowej waluty rezerwowej, opartej na parytecie złota i ustabilizował relacje bilansów handlowych największych europejskich gospodarek przemysłowych. Tym samym położył kres agresywnym politykom dewaluacyjnym skonfliktowanych ze sobą państw starego kontynentu.

ekonomię. Według niego *„…ekonomia jest nauką badającą zachowanie człowieka jako relację między celami a rzadkimi środkami, które mają alternatywne zastosowania…”*, por Robbins (1932; tłumaczenie własne). W zgodzie z kontekstami filozoficznym, psychologicznym oraz mikroekonomicznym, wspomniane cele człowieka utożsamiać można ze spełnianiem ujętych w piramidzie Maslowa potrzeb. Przy tym jak zaznaczają Roman Milewski i Eugeniusz Kwiatkowski: *„…gospodarowanie, czyli działalność gospodarcza ludzi odbywa się ciągle, ze względu na odnawialność i rozwój ludzkich potrzeb…”*, por. Milewski i Kwiatkowski (2000). Okazuje się, więc, że celem ekonomicznym społeczeństwa gospodarującego jest dążenie do pokonania problemu rzadkości. Oznacza to spełnianie nieskończenie odnawiających się i rozwijających potrzeb przy dostępie do skończonej ilości zasobów. Alternatywnie, przyjmując wobec takich warunków założenie o heterogeniczności potrzeb dochodzimy do wniosku, że celem gospodarczym społeczeństwa jest spełnianie nieskończonej liczby potrzeb przy dostępie do skończonej liczby zasobów.

Zgodnie z wywodzącą się ze szkoły austriackiej, subiektywną teorią wartości, kwotę pieniężną, za którą sprzedano produkt lub usługę postrzegać można jako dolne ograniczenie miary użyteczności płynącej ze spełnienia potrzeby konkretnego konsumenta, w sprecyzowanym miejscu i czasie. Jeżeli dla uproszczenia przyjęlibyśmy idealne różnicowanie cen na rynku, to utożsamilibyśmy wspominane dolne ograniczenie z dokładną miarą użyteczności. Okazałoby się, że przy odjęciu zmian inflacyjnych pieniądze stanowią równowartość użyteczności zachowaną w czasie. W tym kontekście przepływy kapitałowe na rynku finansowym to niejako strumienie wartości równoważne wytworzonej przez społeczeństwo użyteczności. Te trafiają natomiast za pośrednictwem rynku finansowego do przedsięwzięć zapewniających najwyższy poziom stopy zwrotu z inwestycji względem podejmowanego ryzyka. W ten sposób istniejąca wartość kierowana jest w stronę działań gospodarczych generujących wartość w sposób najbardziej efektywny.

W uproszczeniu gospodarowanie realizuje się więc poprzez wzbogacanie pierwotnie istniejących zasobów o wartość na drodze wytwarzania produktów i świadczenia usług, które ostatecznie zostają sprzedane i w domyśle skonsumowane. Podobny schemat znajduje swoje praktyczne odzwierciedlenie w metodzie produkcyjnej obliczania PKB. W takim ujęciu, kluczowym czynnikiem procesu gospodarowania jest więc rosnąca w czasie funkcja wartości mierzącej użyteczność płynącą ze spełniania potrzeb. Funkcja postępu gospodarczego jest jej dodatnią pochodną. Przyjmując taki model, nawet przy dostępie do skończonej ilości zasobów możliwe jest spełnianie co raz większej i potencjalnie nieskończonej liczby potrzeb. W ujęciu

teoretyczno-filozoficznym, efektywny rynek finansowy jest więc dla społeczeństwa gospodarującego bardzo ważny w rozwiązywaniu problemu rzadkości.

Obserwując gospodarkę całościowo, z bardziej praktycznego punktu widzenia, zauważyć można, że opisywany wcześniej priorytet rozwiązywania problemu rzadkości dzieli się na wiele innych wymiarów. Makroekonomia w formie znanej dzisiaj, urodziła się w latach czterdziestych dwudziestego wieku jako odpowiedź intelektualistów na Wielki kryzys. Przyświecała jej pierwotnie misja lepszego zrozumienia mechanizmów rynkowych w celu zapobiegnięcia wystąpienia podobnych zjawisk w przyszłości, por. Lucas (2003). Pierwsze z tradycyjnych modeli makroekonomicznych rodziły się więc z nadrzędnym celem badania zjawiska fluktuacji rynkowych. W dalszej kolejności, jako uzupełnienie do tej idei pojawiały się kolejne, opisujące konkretne wycinki gospodarki. W związku z tym, na zawsze w tradycji makroekonomicznej oraz statutach instytucji regulujących rynek zapisały się cele związane przeciwdziałaniem efektom gorączki spekulacyjnej lat trzydziestych. Chodzi więc tradycyjnie o stabilny wzrost gospodarczy, pełne zatrudnienie i unormowany poziom cen.

W obecnych realiach gospodarczych rynek finansowy jest niezbędny do utrzymywania stabilnego poziomu cen. Banki centralne kontrolują podaż pieniądza wpływając na stopy procentowe za pośrednictwem operacji otwartego rynku oraz działań depozytowo-kredytowych. Okazuje się jednak, że przynajmniej niektóre składowe efektywnie działającego rynku finansowego promują również bezpośrednio rozwój gospodarczy, na co wskazuje badana od lat dodatnia korelacja obu tych zmiennych. Pozytywną oddziaływanie między rozwojem rynku finansowego, a wzrostem gospodarczym opisywał już Joseph Schumpeter (1911). Biorąc pod uwagę opisany wcześniej schemat teoretyczny, oraz prace austriackiego ekonomisty, we wspomnianej korelacji doszukiwać można się przynajmniej częściowej przyczynowości od strony rynku finansowego w stronę rozwoju gospodarczego. W końcu dwudziestego stulecia, istnienie tego typu oddziaływania starało się dowodzić empirycznie wielu ekonomistów. Używali przy tym wielu modeli statystycznych na przestrzeni wielu zbiorów danych, por. Mckinnon (1973), Shaw (1973), Greenwood i Jovanovic (1990), Pagano (1993), King i Levine (1993). Wnioski z ich badań określić można jednak jako niekonkluzywne.

W ramach dostępnej literatury kompromisem naukowym odnośnie kwantyfikowania wzrostu gospodarczego wydaje się być używanie miar spójnych z tymi wykorzystywanymi przez rozliczne urzędy statystyczne i instytucje międzynarodowe. Wymienić tutaj należy przede wszystkim różne warianty PKB. Większe wyzwanie stanowi ocena wzrostu rynku finansowego. Składa się on z wielu podsektorów, których wartości bywają skorelowane ze sobą

nawet prawie idealnie ujemnie[3]. Autorzy publikowanych badań dysponują więc względnie precyzyjnymi miarami wzrostu gospodarczego ujmującymi zjawisko w sposób zgodny z teorią ekonomiczną i intuicją matematyczną. W przypadku wzrostu rynku finansowego jest inaczej. Dlatego drugi z opisywanych wymiarów danych jest często dostosowywany do pierwszego – brane jest zestawienie wzrostu gospodarczego ze wzrostem jedynie określonego wycinka rynku finansowego. Ograniczając się do instrumentów, których ceny reagują na różne bodźce makroekonomiczne w podobny sposób udaje się uchwycić zależności między określonymi w ten sposób dwoma zmiennymi. Często omawianą w tym kontekście składową rynku finansowego jest rynek instrumentów udziałowych. Silnej pozytywnej korelacji między wzrostem rynku akcji i wzrostem gospodarczym w warunkach współczesnych rynków finansowych dowiedziono empirycznie na przestrzeni licznych prac naukowych, por. Atje i Jovanovic (1993), Korajczyk (1996), Levine i Zervos (1998).

Patrząc z praktycznej perspektywy makroekonomicznej, efektywny rynek finansowy jest więc bardzo ważny, ponieważ pomaga realizować przynajmniej dwa spośród trzech głównych celów gospodarczych społeczeństwa w tym wymiarze. Mechanizmy rynku międzybankowego umożliwiają bankom centralnym prowadzenie sprawnej polityki monetarnej i w konsekwencji przyczyniają się do stabilizacji poziomu cen. Alokacja kapitału na rynku akcyjnym promuje natomiast bezpośrednio wzrost gospodarczy. Wskazują na to prace teoretyczno-naukowe oraz występowanie udowodnionej empirycznie, pozytywnej korelacji między wzrostem rynku akcji, a wzrostem gospodarczym.

Rozpatrując ekonomiczną perspektywę pojedynczych przedsiębiorstw, funkcjonujących w ramach rozróżnialnych struktur rynkowych, zagadnienia makroekonomiczne można rozbić na niezliczoną liczbę podproblemów. W takim ujęciu, przedmiotem zainteresowania przestaje być ogół społeczeństwa oraz jego wyzwania – wzrostu gospodarczego, eliminacji bezrobocia i stabilizacji cenowej. Kluczowe stają się problemy wynikające z gry konkurencyjnej wiążącej uczestników poszczególnych rynków. Początków mikroekonomii neoklasycznej dopatruje się w pracach Léona Walrasa (1874) na temat teorii równowagi ogólnej. Sam jej koncept odnosił się do kwestii makroekonomicznych, jednak dał początek rozważaniom na temat skalowania go do poziomu mikro. W najbardziej podstawowej wersji teorii[4], definicja stanu równowagi ogólnej pokrywa się z koncepcją pochodzącą

[3] Niektóre klasy walorów giełdowych są skorelowane ze sobą niemal idealnie ujemnie ze względu na istnienie instrumentów pochodnych. Zależność tego typu dotyczy akcji i opcji sprzedaży wystawianych na te akcje.

[4] Teoria równowagi ogólnej skupia się w głównej mierze na mechanizmach determinujących przechodzenie układów połączonych ze sobą rynków od wybranych stanów początkowych do wspominanych stanów równowagi.

z pierwszego fundamentalnego twierdzenia ekonomii dobrobytu. Mowa więc o wolumenach produkcji i związanych z nimi poziomach cen, przy których zestaw rynków doskonale konkurencyjnych osiąga kolektywnie optimum w sensie Pareto[5], por. Hammond (1997).

Koncepcję równowagi ogólnej przenieść można do poziomu mikroekonomicznego otrzymując teorię równowagi cząstkowej. Początek tego typu idei przypisuje Alfredowi Marshallowi (1890). Stworzona przez niego teoria dotyczy pojedynczych rynków widzianych z perspektywy pojedynczych przedsiębiorstw – całą resztę gospodarki obejmuje się przy tym klauzulą ceteris paribus, uznając ją za stałą. Według tego typu modeli, przedsiębiorstwa, niezależnie od obowiązującego schematu konkurencji, dążą zawsze do swojego optimum ekonomicznego, wyznaczanego przez punkt maksymalizacji całkowitego zysku. Przy ustaleniu zdeterminowanego w ten sposób poziomu produkcji model osiąga swój stan równowagi rynkowej. W takim ujęciu, generowanie jak największych przychodów, przy jak najniższych kosztach jest więc celem nadrzędnym każdego z podmiotów. Zbliżając się do rzeczywistości finansowo-rachunkowej należy rozszerzyć te rozważania o złożoność struktur kapitałowych rozpatrywanych spółek. Wtedy wcześniej wspominany cel utożsamiać można z maksymalizacją stopy zwrotu z kapitału własnego. Przy tanim dostępie do finansowania zewnętrznego przedsiębiorstwa mogą efektywnie podnosić wartość wskaźnika ROE (Return On Equity). Alokacja kapitału dokonywana przez rynek finansowy zapewnia spółkom o wysokich wskaźnikach rentowności niskokosztowe finansowanie zewnętrzne. Widać więc, że patrząc na pojedynczy rynek z perspektywy pojedynczego przedsiębiorstwa, efektywnie działający rynek finansowy jest niezwykle ważny. Zwiększa on przewagi konkurencyjne bardziej rentownych przedsiębiorstw – a więc wykorzystujących administrowany kapitał w bardziej efektywny gospodarczo sposób. Dzięki temu rynek działa bardziej efektywnie.

Odnosząc się z kolei do mikroekonomicznej teorii wyboru konsumenta, stwierdzić można, że przy założeniu jego aktywności inwestycyjnej, efektywniej działający rynek finansowy oznacza wyższe stopy zwrotu z oszczędzanego kapitału. To z kolei podnosi wartości funkcji ograniczenia budżetowego pozwalając konsumować szersze koszyki dóbr oraz w konsekwencji zapewniając więcej użyteczności.

Podsumowując, na podstawie istniejących prac uznanych w dziedzinie ekonomii filozofów i teoretyków oraz udokumentowanych badań empirycznych naukowców z różnych

[5] W optimum Pareto nie można zwiększyć zysku jednego producenta (lub konsumpcji jednego konsumenta) bez zmniejszenia zysku innego producenta (lub konsumpcji innego konsumenta), por. Samuelson i Nordhaus (1980).

uniwersytetów udaje się dowieść bardzo ważnej roli rynku finansowego dla gospodarki. Zastosowanie ma przy tym uproszczenie jego roli do mechanizmu alokacji kapitału. Biorąc przecięcie zbioru wszystkich części rynku finansowego, których dotyczy omawiana zależność należy się ograniczyć do rynku akcji. Okazuje się więc, że efektywna alokacja kapitału w ramach rynku akcji jest bardzo ważna dla gospodarki – z perspektywy teoretycznej, makroekonomicznej i mikroekonomicznej. Dlatego, w konsekwencji, wyrażająca efektywność alokacji kapitału, jak najlepsza optymalizacja portfeli akcyjnych poszczególnych inwestorów giełdowych jest ważna dla gospodarki. W dalszej część pracy istota problemu wagi rynku finansowego dla gospodarki rozpatrywana będzie już tylko z najbardziej intuicyjnej dla jednostki, mikroekonomicznej perspektywy inwestora giełdowego.

2.2 Rynek akcji z perspektywy inwestora

Opisywana na początku pierwszego rozdziału, ogólna historia rynków finansowych, wspomina wyemitowanie na początku siedemnastego wieku przez pochodzącą z Niderlandów spółkę pierwszych akcji. W tej samej części Europy, w roku 1688 opublikowana została książka uważana dzisiaj za pierwsze, całkowicie skupione na tematyce giełdowej, dzieło literackie, por. Castro (1781), Corzo i in. (2014). Jej autor przedstawia w treści mechanizmy działania poszczególnych instrumentów finansowych, dynamikę zmian koniunktury na rynku akcji oraz zachowania inwestorów. Co ciekawe, nawet tak dziewiczy opis zjawiska zawiera liczne porady autora mające umożliwić czytelnikowi osiąganie zysków na rynku. Uzasadnionym jest więc przypuszczenie, że historia spekulacji giełdowej jest tak długa jak historia samej giełdy. Uczestnikom rynku, wymieniającym między sobą walory finansowe, od samego początku zależało na optymalizacji tego procesu pod kątem maksymalizacji zysku. Wspomnianą książkę można uznać, za pewien kamień milowy w rozwoju nauk związanych z finansami behawioralnymi i wyceną papierów wartościowych. Dzisiejsze metody wyceny akcji podzielić można na te związane z analizą fundamentalną oraz te dotyczące analizy technicznej.

Analiza fundamentalna służy wycenie jakości spółki, określanej, względem otaczających ją warunków rynkowych, zdolnością generowania jak największej rentowności kapitału w jak najdłuższym okresie. Jakość tę bada się w oparciu o sprawozdania finansowe spółki oraz dane dotyczące ogółu rynku i całej gospodarki. Wykorzystuje się więc pełen przekrój dostępnych informacji – zarówno ilościowych, jak i jakościowych, tych związanych z opisem przeszłości, jak i tych estymujących przyszłość. Ze względu na charakter wykorzystywanych źródeł informacji, analizę fundamentalną określić należy badaniem

statycznym – nastawionym na uchwycenie pewnego obraz rzeczywistości, w pewnym momencie.

Początki analizy fundamentalnej łączą się naturalnie z samym momentem pojawienia się pierwszych akcji giełdowych. Trudno wskazać przy tym jakieś historycznie wydarzenia wpływające znacząco na metodykę jej sporządzania. O zasadności użycia decydują dwa aksjomaty bazowe.

1. Istnienie wyceny jakości spółki, która jest w danym momencie jedyną obiektywną.
2. Zbieganie wraz z upływem czasu bieżącej wyceny rynkowej do wyceny obiektywnej w sytuacji stałości tej drugiej na przestrzeni zadanego okresu.

Trwała możliwość osiągania ponadprzeciętnego zysku przy tych założeniach wynika z ewentualnej zdolności określania wyceny bardziej obiektywnej niż ta wskazywana przez rynek. Oznacza więc: dostęp do większej ilości informacji lub sprawniejsze ich dyskontowanie. Co istotne, druga z możliwości implikuje, że rynek w roli mechanizmu wyceny nie jest maksymalnie efektywny. Przy założeniu wyżej opisanego obrazu rzeczywistości oraz spełnieniu wymienionych warunków, analiza fundamentalna umożliwiałaby sprzedaż akcji spółek przecenionych oraz zakup akcji spółek niedocenionych. Działanie takie generowałoby zysk w długim okresie – definiowanym, w tym wypadku, czasem potrzebnym rynkowi na korektę różnicy występującej pierwotnie między wyceną bieżącą, a wyceną obiektywną.

Empirycznie nie można stwierdzić w jakim stopniu założenia bazowe analizy fundamentalnej są spełnione. Przy obecnym poziomie rozwoju nauki, rzecz ma charakter filozoficzny i powinna być rozpatrywana jedynie w takim kontekście. Przyjmując nawet całkowitą trafność obu stwierdzeń, kwestią sporną pozostaje trwała możliwość osiągania ponadprzeciętnych zysków przy użyciu metod analizy fundamentalnej. Jest ona natomiast głównym narzędziem profesjonalnych inwestorów, wykorzystywanym w procesie selekcji spółek, por. Harrington (2003). Okazuje przy tym, że jak mówią raporty branżowe[6], aż *"…79,6% funduszy akcji krajowych (amerykańskich) pozostało w tyle za (indeksem) S&P Composite 1500 w roku 2021…"*, por. SPIVA (2021; tłumaczenie własne). Istotną wadą praktyczną analizy fundamentalnej jest pracochłonność połączona niemożliwością wystąpienia

[6] Wydawany corocznie raport SPIVA (Standard & Poors Index Versus Active) porównuje wyniki osiągane przez fundusze aktywnie wybierające swoje akcyjne lokaty kapitału z wynikami indeksów giełdowych, por. S&P Global (https://www.spglobal.com/spdji/en/research-insights/spiva/about-spiva/, dostęp: 15.07.22r.).

stałości warunków rynkowych na przestrzeni jakiegokolwiek okresu. Wysokim kosztom realizacji towarzyszy szybkie tempo dezaktualizacji[7].

Analiza techniczna służy przewidywaniu ruchów cen akcji spółki ze względu na ich zapis historyczny oraz liczby zawieranych w przeszłości transakcji rynkowych, por. Kirkpatrick i Dahlquist (2006). Te dwa indykatory uznaje się przy tym powszechnie za podstawowe dla uchwycenia istoty zjawiska. Bardziej liberalne definicje rozszerzają pojęcie analizy technicznej o badania dowolnej liczby zmiennych kwantyfikowalnych, fluktuujących na przestrzeni czasu w sposób bieżący. W każdym z założeń dotyczy ona wykorzystywania jedynie danych o charakterze ilościowym, związanych z opisem przeszłości. Rodzaj źródeł informacji przesądza o dynamicznej naturze analizy technicznej. Jej celem jest ciągłe w ujęciu czasowym monitorowanie rzeczywistości.

Podobnie jak w przypadku analizy fundamentalnej – początków analizy technicznej upatrywać można już w czasach notowań pierwszych akcji giełdowych[8]. Moment przełomowy w rozwoju pojęcia analizy technicznej nastąpił jednak dopiero w drugiej połowie osiemnastego wieku. W tamtym okresie, amerykański dziennikarz finansowy i inwestor, Charles Dow, napisał między pewien cykl artykułów[9] kolektywnie nazywanych dzisiaj syntezą teorii zwanej „Dow theory", a za razem podstawą współczesnej analizy technicznej[10]. Obecnie przyjmuje się, że o zasadności jej użycia decydują przedstawione niżej założenia bazowe, por. Internet[11].

1. Rynek dyskontuje maksymalnie efektywnie wszystkie dostępne informacje.
2. Ceny poruszają się w ramach trendów.
3. Ruchy cen są w jakimś stopniu powtarzalne, co powoduje nawracanie trendów.

Dostrzec tutaj można istotną sprzeczność z ideą trwałej możliwości osiągania zysku przy użyciu analizy fundamentalnej. W świecie, w którym pierwsze założenie analizy technicznej są całkowicie spełnione i przy wykorzystywaniu jedynie ogólnodostępnych informacji, okazuje się być nierzeczywista.

---

[7] Uwagę na problem wysokich kosztów działalności operacyjnej współczesnych funduszy zwracał wielokrotnie między innymi CEO Berkshire Hathaway, Warren Buffett, por. Floyd (2019).
[8] Elementy opisu analizy technicznej ujmowała już wspomniana siedemnastowieczna książka Josepha de la Vegi.
[9] 255 tekstów opublikowano pierwotnie na łamach Wall Street Journal.
[10] Charles Dow, na podstawie notowań cen akcji i wolumenów transakcyjnych, wyliczał średnie ruchome, definiował powtarzające się trendy giełdowe i poszukiwał zdarzeń wybijających ruch cen ze zmian schematycznych. Wszystkie elementy połączył w spójny obraz teoretyczny własnej hipotezy dotyczącej rzeczywistości rynkowej. Warto przy tym podkreślić wagę słów teoretyczny oraz hipoteza – autor nigdy nie przedstawiał swoich rozważań w postaci praktycznej strategii spekulacyjnej.
[11] CFA Institute (https://www.cfainstitute.org/en/membership/professional-development/refresher-readings/technical-analysis; dostęp: 17.07.22r.).

Pozostałe dwa aksjomaty odnoszą się do w głównej mierze do psychiki inwestorów. Uznaje się za ich pośrednictwem, że mózg człowieka, chcąc jak najlepiej pojmować rzeczywistość przy jak najmniejszym wysiłku, poszukuje jej uproszczeń. Dzięki szacowaniu rzeczywistości zestawem wzorów i schematów może znacząco zwiększać efektywność reakcji na bodźce zewnętrzne[12]. Inwestorzy giełdowi, dokonując retrospekcji, naturalnie mieliby reagować na podobne w swojej ocenie bodźce w sposób, który w odniesieniu do zdarzeń przeszłych zapewniłby im zysk. Zakłada się tu, że silniejszym bodźcem z ich perspektywy jest sama zmiana cen. W związku z tym stwierdzić można, że u podstaw analizy technicznej leży założenie o szablonowym myśleniu ludzi i wynikającym z niego szablonowym funkcjonowaniu rynków.

Trwała możliwość osiągania ponadprzeciętnego zysku przy założeniach analizy technicznej wynika z ewentualnej zdolności predykcji ruchu bieżącej ceny. W przeciwieństwie do analizy fundamentalnej, wspomniane działanie nie jest w żadnym stopniu związane z wyceną rzeczywistych cech spółki. Dotyczy ono klasyfikacji bodźców cenowych oddziałujących na inwestorów giełdowych i wyprzedzenia ich, interpretowanej za możliwą do przewidzenia, reakcji. Zakładając poprawność ujętego wyżej obrazu rzeczywistości oraz trafne wychwytywanie schematyczności myślenia inwestorów, analiza techniczna umożliwiałaby sprzedaż akcji spółek, których cena będzie spadać oraz zakup akcji spółek, których cena będzie rosnąć. Zauważalne jest przy tym skupienie analizy technicznej na krótkim okresie – definiowanym czasem reakcji inwestorów na poszczególne bodźce.

Podobnie jak w przypadku analizy fundamentalnej, empirycznie nie można stwierdzić tego jak bardzo założenia bazowe analizy technicznej są spełnione w rzeczywistości. Nie wiadomo także, czy nawet w takim wypadku trwałe osiąganie ponadprzeciętnych zysków przy jej użyciu byłoby możliwe. Istnieje wiele badań dotyczących tej kwestii natomiast nie dostarczają one konkluzywnych wniosków, por. Olser (2000), Lo i in. (2000), Scott i Park (2007). W praktyce, analiza techniczna jest często wykorzystywana komplementarnie względem, stosowanej do selekcji spółek, analizy fundamentalnej. Pomaga podejmować decyzje dotyczące samych momentów zakupu i sprzedaży akcji. Wykorzystywanie analizy technicznej nawet w tej ograniczonej roli, kojarzone jest czasem wśród zawodowych inwestorów amerykańskich w sposób pejoratywny,

[12] Założenia analizy technicznej można by interpretować zgodnie z powiedzeniem, o tym, że "historia lubi się powtarzać". To jednak nie ujmowałoby ich istoty. Postulują one raczej to, że mózg człowieka "lubiłby" jeżeli historia by się powtarzała, więc interpretuje ją w taki sposób.

por. Harrington (2003). Niewątpliwą jej zaletą jest względna prostota automatyzacji i wynikające z niej niskie koszty realizacji. Ze względu na, między innymi: całkowicie kwantyfikowalny charakter, wykorzystywanie dużych zbiorów danych, połączenie z teoriami behawioralnymi oraz polaryzację opinii ekspertów, analiza techniczna pozostaje bardzo popularnym tematem artykułów i prac naukowych, por. Harrington (2003), Mizrach i Weerts (2007), Azzopardi (2010).

Wspominane nieefektywności analizy technicznej i analizy fundamentalnej tłumaczyć można czystą losowością. W takim ujęciu uzasadnione jest dążenie do ich minimalizacji poprzez zwiększanie skali opcji inwestycyjnych oraz wprowadzanie narzędzi statystycznych i probabilistycznych. Tak postawionymi wyzwaniami optymalizacyjnymi zajmuje się w szerokiej definicji współczesna matematyka finansowa. Za jej istotę, uznaje się użycie zaawansowanych akademicko metod wnioskowania przy modelowaniu zjawisk związanych z rynkiem finansowym w celach rozwiązywania problemów optymalizacyjnych.

Początki matematyki finansowej związane są z wyceną opcji akcyjnych. Według ogólnego konsensusu wyznacza je rok 1900 – data obrony pracy doktorskiej francuskiego matematyka, Louisa Bacheliera[13]. Zawarta w niej idea zastosowania zaawansowanej probabilistyki matematycznej przy próbie rozwiązania problemu optymalizacyjnego z dziedziny finansów była w tamtych czasach bardzo innowacyjna, por. Weatherall (2013). Nie przełożyła się jednak na trwałe zainteresowanie naukowców tym tematem. Dopiero przeszło siedemdziesiąt lat później, powstał model Blacka-Scholesa[14]. Od tego czasu datować można początek dynamicznego rozwoju matematyki finansowej, por. Hayes i in. (2022). Postępująca komputeryzacja wyposażyła naukowców w moc obliczeniową niezbędną do realizacji ich pomysłów. Aspekty analityczne i algebraiczno-geometryczne problemów modelowania zjawisk fizycznych rozwiązywano przy pomocy co raz efektywniejszych metod numerycznych. Wraz z dalszym rozwojem technologii informacyjnych wiele idei matematycznych, dla których wcześniej nie znajdowano zastosowania, okazywało się możliwych do wykorzystania w praktyce. Stopniowo, główne determinanty konkurencyjności na rynkach finansowych przesuwały się w stronę zdolności efektywnego przetwarzania

[13] Rozprawa Louisa Bacheliera została opublikowana jako: „Théorie de la Spéculation". Autor, jako pierwszy, przedstawił w niej model procesu stochastycznego opisującego zjawisko fizyczne zwane ruchami Browna. Wykorzystał go następnie w swoich rozważaniach do wyceny kontraktów opcyjnych na zakup akcji, por. Encyclopedia Of Mathematics (https://encyclopediaofmath.org/images/f/f1/LouisBACHELIER.pdf; dostęp: 23.07.22r.).

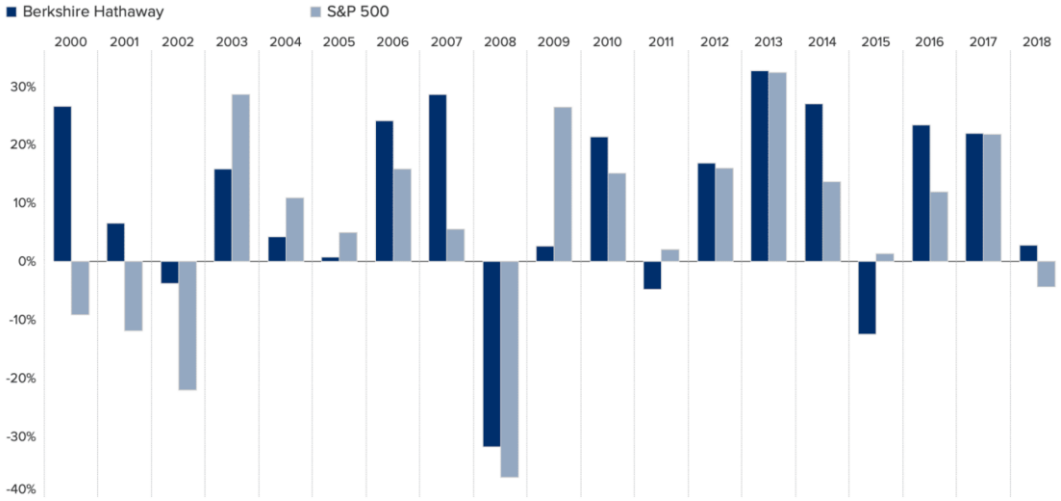
[14] Fischer Black, Robert Merton i Myron Scholes użyli w celu wyceny opcji akcyjnych modelu procesu stochastycznego geometrycznych ruchów Browna. Wyniki ich prac, opublikowane dokładnie w roku 1973, wykorzystywane są do dzisiaj z tym samym przeznaczeniem w praktyce.

informacji. Co raz więcej procesów poddawano automatyzacji i optymalizacji. Znalazło to swój wyraz w, charakterystycznym dla dzisiejszych czasów, intensywnym rozwoju handlu algorytmicznego, oraz giełdowej implementacji metod sztucznej inteligencji.

Podstawowe paradygmaty analizy fundamentalnej i analizy technicznej, pozostają niepewne. Nawet przy założeniu ich względnej poprawności, trwałą możliwość osiągania ponadprzeciętnego zysku przy wykorzystaniu którejś z nich należy poddawać w wątpliwość. Taki stan rzeczy potwierdzają tysiące historii spektakularnych porażek inwestycyjnych. Warto jednak wspomnieć na koniec, że istnieją również empiryczne dowody mogące przemawiać za skutecznością obu podejść.

Wspominane już w tym podrozdziale, największe przedsiębiorstwo zarządzające kapitałem (ang. *asset manager*) na świecie, Berkshire Hathaway pokonywało wiele razy, ujmowany często w roli barometru amerykańskiej gospodarki, indeks S&P 500.

Inwestycja w Berkshire Hathaway oraz inwestycja w S&P 500.

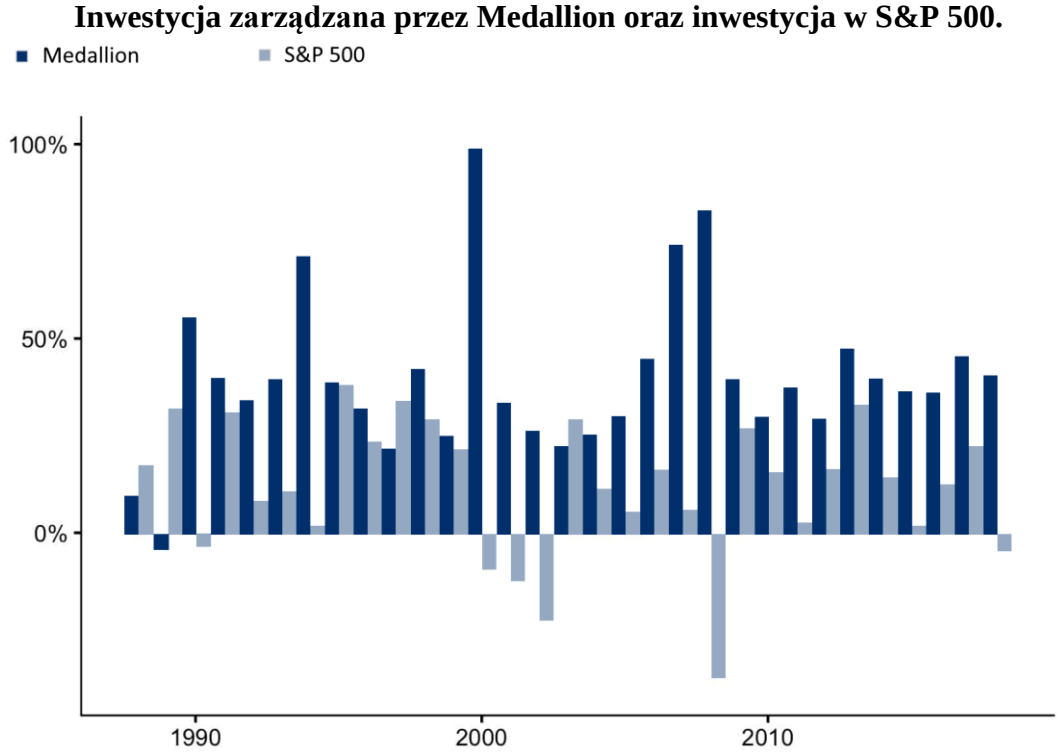


Źr: opr. wł. na podst. Rosenbaum (2021).

Rysunek 2.1. *Na osi pionowej odłożono roczne stopy zwrotu z inwestycji w ujęciu procentowym. Na osi poziomej określono wartości zmiennej czasu. Skonstruowane zestawienie jest miarodajne ponieważ inwestowanie zarówno w akcje Berkshire Hathaway, jak i kontrakty terminowe na indeks S&P 500 nie wiąże się z dodatkowymi opłatami*

Rysunek 2.1. przedstawia porównanie rocznych stóp zwrotu z inwestycji w Berkshire Hathaway oraz S&P 500 w latach 2000 – 2018. Pokazuje, że spółka zapewniała swoim inwestorom w takim porównaniu wyższe zwroty roczne z inwestycji na przestrzeni piętnastu spośród osiemnastu branych okresów. Decydowała o tym wybitnie fundamentalna w swojej naturze strategia analityczna jego zarządców – Warrena Buffetta i Charlesa Mungera. Bardziej

spektakularny sukces na szerszym przedziale czasowym osiągał należący do Renaissance Technologies fundusz Medallion.



Źr: opr. wł. na podst. Maggiulli (2019).

Rysunek 2.2. *Roczne stopy zwrotu z inwestycji wyrażono w ujęciu procentowym i odłożono na osi pionowej. Wymiar poziomy równoznaczny jest z osią czasu. Przedstawione na rysunku wyniki Medallionu nie obejmują pobieranych przez fundusz kosztów prowadzenia rachunku. Inwestowanie w kontrakty terminowe na indeks S&P 500 nie wiąże się natomiast z dodatkowymi opłatami.*

Rysunek 2.2. ilustruje zestawienie rocznych stóp zwrotu z inwestycji zarządzanej przez Medallion oraz inwestycji w S&P 500 na przestrzeni lat 1988 – 2018. Jak widać na powyższym rysunku, jego roczne zwroty z inwestycji pokonywały wyniki indeksu S&P 500 w ramach dwudziestu ośmiu z trzydziestu branych okresów. Tak niezwykłe osiągnięcia Medallionu łączyły się z przyjmowanym przez zarządzających rodzajem wysoce zautomatyzowanej algorytmicznie strategii inwestycyjnej, zgodnej przynajmniej częściowo z paradygmatem analizy technicznej. Neo Yi Peng streścił swój dotyczący funduszu artykuł następującymi słowami: *„…Grupa utalentowanych matematyków i informatyków zastosowała uczenie maszynowe do modelowania rynków finansowych, stawiając na strategie krótkoterminowe, które od 1988 roku zwracały 66% rocznie…”*, por. Peng (2020; tłumaczenie własne). Gregory Zuckerman stwierdził, że: *“Medallion (…) szuka przede wszystkim zależności między grupami akcji, między sektorami, między indeksami. Oni nie zagłębiają się w analizę fundamentalną,*

często nawet nie wiedzą czym się zajmują spółki, które mają w portfelu”, por. Zuckerman (2019; tłumaczenie własne) [15].

Reasumując, rozwojowi rynku finansowego od samego początku towarzyszył postęp myśli analitycznej mającej w założeniu przekładać się na trwałą możliwość osiągania ponadprzeciętnych zysków. Jeden z podziałów metod klasyfikuje je w ramach dwóch podejść: fundamentalnego oraz technicznego. Nie istnieją dowody empiryczne mogące potwierdzać, bądź obalać prawdziwość założeń lub samą skuteczność działania narzędzi wykorzystywanych przez żadne z nich. W związku z tym trudno mówić o jednoznacznej wyższości jednego z podejść nad drugim. Istotniejsza, w kontekście uznania wśród przedstawicieli największych uczestników rynku, jest analiza fundamentalna. Szybki rozwój technologiczny w połączeniu dorobkiem naukowym matematyki finansowej napędza ekspansję zainteresowania analizą techniczną.

Efektywna alokacja kapitału wyrażana jak najlepsza optymalizacja portfeli akcyjnych inwestorów giełdowych jest wielowymiarowo istotna dla gospodarki, por. Podrozdział 2.1. Rozwiązywanie problemu optymalizacji portfela akcyjnego przy użyciu narzędzi analizy technicznej jest natomiast podejściem niebezpodstawnym, osadzonym w faktycznej metodologii rynkowej oraz potencjalnie innowacyjnym.

[15] Wyniki osiągane przez fundusz Medallion w ujęciu zagregowanym istotnie dystansowały te wyznaczane notowaniami cen walorów S&P 500 oraz Berkshire Hathaway. Z przedstawionych na Rysunkach 2.1 i 2.2 rocznych stóp zwrotu wynika, że 1$ zainwestowany w S&P 500 na okres 1988 – 2018 urósłby do 20$. 1$ zainwestowany w Berkshire Hathaway urósłby w tym czasie do około 100$. 1$ zarządzany przez Medallion na tym samym przedziale czasowym urósłby do ponad 20 000$ (bez opłat).

3 Uczenie ze wzmacnianiem

Niniejszy rozdział dotyczy zastosowanych w pracy metod badawczych. Podzielono go na 3 obszary. Pierwszy z nich – Podrozdział 3.1. – dotyczy historii rozwoju sztucznej inteligencji oraz wyprowadzanej z niej koncepcji uczenia ze wzmacnianiem. Prezentuje ekspotencjalne tempo wzrostu możliwości optymalizacyjnych programów modyfikujących sieci neuronowe w procesach treningowych. Podrozdział 3.2. poświęcono teoretycznemu omówieniu schematów działania algorytmów uczenia ze wzmacnianiem. Rozważania wyprowadzono od terminów przyjętej definicji bazowej. Ostatecznie opisano interpretowany w pojęciach łańcuchów Markowa model probabilistyczny, w którym wyprowadzono wzory podstawowych dla zjawiska funkcji. W Podrozdziale 3.3. dokonano przeglądu artykułów naukowych dotykających, podobnych do podejmowanego w pracy, problemów badawczych. W ten sposób dokonano ostatecznego wyboru używanych do badań dwóch metod uczenia ze wzmacnianiem – A2C i PPO. Kolejne 2 części Podrozdziału 3.3. omawiają ogólną budowę programistyczną oraz działanie obu algorytmów.

3.1 Historia sztucznej inteligencji i koncepcje uczenia maszynowego

Jak stwierdził w 1957 roku Hubert Simon[16] *„…obecnie na świecie istnieją maszyny, które myślą, uczą się i tworzą. Co więcej, ich zdolność do robienia tych rzeczy będzie szybko rosła, aż w widocznej przyszłości zakres problemów, z którymi mogą sobie poradzić, będzie współmierny do zakresu, do którego ludzki umysł został zastosowany…”*, por. Crevier (1993; tłumaczenie własne). Słowa te odnosiły się do pojęcia sztucznej inteligencji. Z perspektywy roku 2022 widać, że w pewnym sensie poprawnie przewidywały przyszłość. Można stwierdzić, że dzisiejsze maszyny faktycznie myślą, uczą się i tworzą. Radzą sobie z wieloma złożonymi wyzwaniami logicznymi o wiele lepiej niż ludzie. Ekstrapolując aktualne trendy trudno sobie wyobrazić, żeby za 50 lat mogły istnieć jakiekolwiek, w których by im ustępowałyby. Paradoksalnie jednak, mimo, że przepowiednia Simona spełnia się, to nie w sposób, który miał na myśli wypowiadając te słowa.

Pewne wczesne koncepty sztucznej inteligencji powstawały w pierwszej kolejności jako całkowicie abstrakcyjne twory fikcji literackiej[17]. Zjawisko było wtedy całkowicie oderwane od rzeczywistości naukowej. Pojawiające się znacznie później, formalne definicje ewoluowały

[16] Hubert Simon był amerykańskim ekonomistą, laureatem Nagrody Nobla z dziedziny ekonomii oraz Nagrody Turinga – odpowiednika Nagroda Nobla w dziedzinie informatyki.

[17] Mówiąc o wczesnej fikcji literackiej odnoszącej się do pojęcia sztucznej inteligencji, wspomnieć można na przykład postać Frankensteina, por. Mary Shelley (1818).

naturalnie wraz z rozwojem samego konceptu – począwszy od teoretycznych badań nad możliwościami pierwszych komputerów w latach trzydziestych dwudziestego wieku, aż do dzisiaj. Za okres krystalizacji pojęć związanych z tą dziedziną do form zgodnych z obecnym stanem rzeczy uznaje się przełom dwudziestego i dwudziestego pierwszego wieku. W związku z tym, jako definicje obecnego pojęcia sztucznej inteligencji przywołać można te bazujące na słowach pisanych wtedy książek. Stwierdzają one, że sztuczna inteligencja jest nauką o budowie inteligentnych agentów, nazwanych też często aktorami. Agent to coś wchodzącego w interakcje z otoczeniem – jak pies, samolot, człowiek, czy społeczeństwo. Inteligentny agent to coś wchodzącego w interakcje ze środowiskiem w sposób inteligentny. Inteligencja objawia się poprzez to, że działania agenta są przez niego optymalizowane ze względu na cele i możliwości wyznaczane ograniczeniami percepcji i mocy obliczeniowych. Inteligentny agent, uczy się przy tym z własnych doświadczeń i pozostaje elastyczny wobec zmieniającego się środowiska. Oceny inteligencji agenta można więc dokonać na podstawie jego działań. Głównym celem nauki zwanej sztuczną inteligencją jest zrozumienie warunków, które decydują o tym, że tak definiowane inteligentne działania agenta są możliwe, por. Poole i in. (1998). Powyższy opis stoi w zgodzie z definicjami formułowanymi przez autorów licznych podręczników z tamtego okresu, por. Russell i Norvig (2003), Nilsson (1998), Legg i Hutter (2007). Na przestrzeni tej pracy, określenia agenta i aktora stosowane były zamiennie. Określenia agenta, aktora, obserwatora i środowiska odnoszące się do schematów technicznych oznaczano wielką literą – w przeciwieństwie do tych samych słów używanych w kontekście ogólnym. Określenie „środowiska problemu" było prawie równoznaczne z określeniem „Środowiska". Występowało natomiast w mniej ścisłych kontekstach opisowych.

Chcąc uchwycić sam początek ciągu przyczynowo-skutkowego historii rozwoju sztucznej inteligencji warto rozpocząć od samej matematyki. Rozwój tej nauki doprowadził do narodzin teorii obliczeń – starającej się odpowiadać na pytania dotyczące teoretycznych możliwości oraz ograniczeń maszyny, por. Sipser (2013). Wykorzystuje przy tym modele opisujące schematy w jakich dla ustalonej funkcji i danych wejściowych (argumentów) obliczane są dane wyjściowe (wartości). Modele te, zwane maszynami, kojarzyć można w pewnym sensie z matematyczną reprezentacją współczesnego komputera. Jednym z nich jest ten opisany w 1936 przez Alana Turinga[18], por. Hodges (2012). Prace naukowe nad rozwojem

[18] Alan Turing był brytyjskim matematykiem, pionierem współczesnej informatyki.

teorii obliczeń doprowadziły do sformułowania hipotezy Churcha-Turinga mówiącej, że każde obliczenie może być przełożone na obliczenie równoważne wykorzystujące maszynę Turinga, por. Internet[19]. Oznaczało to, że maszyna posługująca się jedynie symbolami zero-jedynkowymi może przeprowadzić poprawnie każdy ciąg wnioskowania matematycznego, por. Berlinski (2000). Taka konkluzja położyła teoretyczne podstawy pod próby stworzenia fizycznie istniejącego, sztucznego systemu zdolnego rozwiązywać nawet niewyobrażalnie szerokie spektrum niewyobrażalnie złożonych problemów.

Wobec braku zauważalnych w matematyce sprzeczności logicznych, naturalnym kierunkiem rozwoju były rozważania na temat możliwości stworzenia sztucznej inteligencji na wzór inteligencji ludzkiej. Za pierwszy artefakt historyczny świadomego rozwoju metod sztucznej inteligencji uznaje się powszechnie powstały już w 1943 roku model matematyczny ludzkiego neuronu[20], por. Russell Norvig (2009). Cechowała go znaczna prostota przez co względem późniejszych, udoskonalonych wersji, jego użyteczność okazywała się mocno ograniczona[21], por. Chandra (2018). Na przestrzeni następnych trzydziestu lat, kolejno pojawiające się prace naukowe dotyczące sztucznej inteligencji dało się zaklasyfikować względem wyrażanego w nich paradygmatu do jednej z dwóch grup. Rozwój sztucznej inteligencji podzielił się wtedy na nurty symbolicznej sztucznej inteligencji oraz koneksjonizmu.

Założeniem symbolicznej sztucznej inteligencji było rozwijanie, wspominanej w definicji, idei inteligentnego agenta wokół wzoru ludzkiego sposobu pojmowania rzeczywistości. Podejście takie było wczesnym wyrazem kształtowania się paradygmatu programowania obiektowego. Jego założenia wyrażać można więc terminami związanymi z dominującą dzisiaj na rynku formą pisania programów komputerowych. Mianowicie, zakładało się, że ludzie widzą świat jako symbole – a więc obiekty, będące instancjami klas oraz wchodzące ze sobą w interakcje za pomocą metod. Tworzone w ten sposób wynalazki, działały w prosty do zrozumienia dla człowieka, hierarchiczny sposób. Ich proces uczenia się był dokonywany całkowicie przez ludzi. Był też niezwykle pracochłonny, ponieważ nie skalował się efektywnie względem złożoności otoczenia rzeczywistego, por. Garnelo i Shanahan (2019).

[19] Wolfram MathWorld (https://mathworld.wolfram.com/Church-TuringThesis.html; dostęp: 25.07.22r.).

[20] Praca dotycząca modelu napisana została przez neurobiologa – Warrena McCulloch oraz logika matematycznego – Waltera Pittsa. Autorzy starali się imitować budowę ludzkiego mózgu.

[21] Pierwszy model matematyczny ludzkiego neuronu był zarazem najbardziej podstawową wersją sieci neuronowej, zdolną odwzorowywać działanie funkcji boolowskich liniowo separowalnych – takich jak na przykład: negacja, koniunkcja lub alternatywa.

Koneksjonizm wiązał się z rozwojem idei inteligentnego agenta wokół schematu biologicznej budowy ludzkiego mózgu. Wyrażało się to skupieniem na systemach nazywanych sieciami neuronowymi. Tworzone w duchu koneksjonizmu wynalazki działały w sposób niezrozumiały dla człowieka. Ich proces uczenia się zachodził za pośrednictwem algorytmu. Przez to nie pochłaniał pracy programistów. Wymagał natomiast ogromnych zbiorów danych i dużej mocy obliczeniowej komputerów. Co najważniejsze jednak, ostatecznie skalował się efektywnie względem złożoności faktycznego otoczenia rzeczywistego. Najwcześniejszym konceptem związanym z nurtem koneksjonizmu był wspominany już, najstarszy wynalazek całej dziedziny sztucznej inteligencji – Neuron McCullocha-Pittsa. Bazując na tym osiągnięciu, Frank Rosenblatt[22] udoskonalił jego budowę oraz stworzył w 1957 roku najstarszą klasyczną sieć neuronową – perceptron jednowarstwowy – wraz z algorytmem uczącym ją rozwiązywania prostych problemów klasyfikacyjnych[23], por. Loiseau (2019).

Aż do połowy lat dziewięćdziesiątych symboliczna sztuczna inteligencja była paradygmatem dominującym nad koneksjonizmem[24], por. Russell i Norvig (2009). Mimo sukcesów Franka Rosenblatta, prace nad algorytmami uczącymi sieci neuronowe postępowały do ostatniej dekady dwudziestego wieku względnie powoli. Istotnym problemem było pojawianie się bariery koncepcyjnej, która przez około piętnaście lat hamowała rozwój dziedziny. Wynikała z tego, że w roku 1969 udowodniono istnienie znacznych ograniczeń jednowarstwowego perceptronu[25]. Impas przełamały definitywnie dopiero prace naukowe z końca lat osiemdziesiątych – opis efektywnego algorytmu propagacji wstecznej[26] oraz wynalazek wielowarstwowej sieci neuronowej[27].

Od wspominanych początków sztucznej inteligencji, aż do połowy lat siedemdziesiątych naukowcy skupieni w ramach paradygmatu symbolicznej sztucznej

[22] Frank Rosenblatt był amerykańskim psychologiem, pionierem sztucznej inteligencji.
[23] Rosenblatt wykazał, że jego wynalazek poza odwzorowywaniem działań funkcji boolowiskich mógł uczyć się rozwiązywania zagadnień klasyfikacji binarnej. Proces nauki realizowany był za pośrednictwem algorytmu. Wspomniana prostota problemów wynikała z ograniczeń perceptronu pozwalających na poprawną klasyfikację jedynie zbiorów liniowo separowalnych, por. Freund i Schapire (1999).
[24] Dominacja symbolicznej sztucznej inteligencji wynikała ze wczesnych sukcesów tego podejścia oraz jego spójności z dziedzinami matematyki i filozofii, por. Manyika (2022). Koneksjonizm był czymś nowym i w tamtych czasach nieintuicyjnym. Dodatkowo, komputery nie dysponowały wtedy dostatecznymi mocami obliczeniowymi, a na świecie brakowało baz danych o wymaganych wielkościach.
[25] Marvin Minsky i Seymour Papert w pracy pt. „Perceptrons: An Introduction to Computational Geometry" wykazali, że jednowarstwowy perceptron nie pozwalał osiągnąć aproksymacji uniwersalnej (oznaczającej odwzorowywanie działania funkcji ciągłych między przestrzeniami euklidesowymi. Nie umożliwiał nawet imitowania funkcji boolowskiej alternatywy rozłącznej, por. Marsland (2014).
[26] Algorytm został stworzony w 1986 roku przez: Davida Rumelharta, Geoffrey'a Hintona i Ronalda Williamsa.
[27] George Cybenko (1989) udowodnił, że wielowarstwowa sieć pozwala osiągnąć aproksymację uniwersalną.

inteligencji osiągali duże sukcesy, mające praktyczne aplikacje[28]. Rozbudzali swoimi przewidywaniami wielkie nadzieje amerykańskich kongresmenów decydujących o finansowaniu badań sztucznej inteligencji. W końcu dotarli jednak do oczywistych z dzisiejszej perspektywy ograniczeń symbolicznej sztucznej inteligencji[29]. W efekcie dużego rozczarowania możliwościami rozwoju ówczesnej sztucznej inteligencji nastąpił znaczny spadek dotyczącego niej zainteresowania naukowego i finansowania badań[30]. Na faktyczne odżycie dziedziny czekać trzeba było ponad pięć lat, do wczesnych lat osiemdziesiątych. Ponowne zainteresowanie symboliczną sztuczną inteligencją wynikło wtedy z dyfuzji technologii i komercjalizacji jej zastosowań. Intensywny rozwój komputerów pozwolił na rozszerzenie ich zastosowań automatyzacyjnych w różnych przedsiębiorstwach. Powstające w tamtym okresie Lisp-maszyny służyły jako nośniki systemów nazywanych ekspertowymi[31]. Ich zmierzch przyniosła jednak już pod koniec lat osiemdziesiątych popularyzacja komputerów osobistych. Proces ten zdecydował również o zaprzestaniu produkcji systemów ekspertowych, por. McCorduck (2004). Osiągnięcia związane z dziedziną wykorzystywane były cały czas w innych kontekstach teoretycznych i inżynieryjnych. Samo pojęcie rozwoju sztucznej inteligencji zamarło jednak wśród głównych sfer nauki na przeszło dziesięć lat[32].

Od połowy lat dziewięćdziesiątych sztuczna inteligencja zaczęła się odradzać pod postacią koneksjonizmu. Na początku dwudziestego pierwszego wieku teoretyczny poziom zaawansowania dziedziny, moce obliczeniowe komputerów oraz rozwój Internetu pozwalały już na efektywne wykorzystywanie algorytmów do uczenia sieci neuronowych na dużych zbiorach danych. Okres ten wyznacza wczesne początki dynamicznego wzrostu dziedziny we wszystkich metrykach. Od okolic 2012 roku wyniki osiągane przez sieci neuronowe przewyższają te cechujące tradycyjnie programowane systemy w ramach większości wyzwań analitycznych[33]. Ostatnie lata przynosiły rozpowszechnienie technologii chmurowych oraz znaczny rozwój badań rozszerzających jeszcze bardziej ich możliwości[34].

---

[28] Wyrazem optymizmu z lat pięćdziesiątych są przywołane we wstępie tego podrozdziału słowa Huberta Simona.
[29] Zderzenie naukowców z barierą spotkało się w czasie z publikacją krytycznego raportu dotyczącego dziedziny, autorstwa Jamesa Lighthill'ego (1973).
[30] Okres stagnacji dziedziny nazywa się dzisiaj pierwszą zimą sztucznej inteligencji (ang. *first AI winter*).
[31] Systemy ekspertowe pozwalały emulować proces podejmowania decyzji przez człowieka – eksperta w jakiejś dziedzinie. Były programami ułatwiającymi interakcje ludzi z bazami danych za pomocą interfejsów. Pozwalały na skalowanie rozwiązań schematycznych problemów, per Russell i Norvig (2009r), Luger i Stubblefield (1997).
[32] Koniec lat osiemdziesiątych wyznaczał początek drugiej zimy sztucznej inteligencji (ang. *second AI winter*).
[33] Raport McKinsey (2017) pod tytułem „Ask the AI experts: What's driving today's progress in AI?", cytuje słowa Adama Coatesa, dyrektora biura do spraw sztucznej inteligencji Baidu, który stwierdził, że definitywny moment określanego w ten sposób przesilenia technologicznego nastąpił w okolicach roku 2012.
[34] Według raportu UNESCO (2021), na przestrzeni lat 2015 – 2019 liczba publikacji naukowych dotyczących sztucznej inteligencji rosła średnio o 50% w skali każdych kolejnych dwunastu miesięcy.

Pojęcie uczenia maszynowego związane jest bezpośrednio z systemami budowanymi zgodnie z paradygmatem koneksjonistycznego nurtu sztucznej inteligencji. Tom Mitchell[35], zdefiniował je jako *„…dziedzinę badań poświęconą zrozumieniu i budowaniu metod, które się "uczą", to znaczy metod, które wykorzystują dane w celu poprawy wydajności (mierzonej) na pewnym zestawie zadań…”*, por. Mitchell (1997; tłumaczenie własne). Słowa te można traktować jako poprawne uchwycenie idei, ale również pewną generalizację, czy też uproszczenie. Zlepiają one trzy drobniejsze pojęcia techniczne w jedno – nazywane metodą. Traktują równoważnie algorytm oraz zawierający go program wyposażony w sieci neuronowe. W praktyce jeden program może zawierać wiele rozłącznych sieci neuronowych i wiele rozłącznych algorytmów uczących. Dla potrzeb tej pracy wykorzystane zostanie założenie rozróżniające te pojęcia. Przyjmowane będzie jednak za każdym razem, że jeden program ma jeden algorytm uczący i jedną sieć neuronową.

Pierwszym krokiem w rozwoju uczenia maszynowego był wspominany już algorytm napisany przez Franka Rosenblatta. Począwszy od lat dziewięćdziesiątych, a więc okresu, w którym koneksjonizm zdominował dziedzinę sztucznej inteligencji, zauważaalna jest postępująca ekspansja uczenia maszynowego. Na przestrzeni ostatnich lat postępowała ona szczególnie dynamicznie. W roku 1997 roku Deep Blue, system ekspertowy działający na specjalnie zaprojektowanej maszynie IBM, pokonał w szachy ówczesnego mistrza świata – Garrego Kasparova, por Saletan (2007). Stał się wtedy pierwszym komputerem w historii, któremu udało się przezwyciężyć w królewskiej grze posiadacza tego tytułem. Wydarzenie to w oczach wielu wyznaczało pewien kamień milowy rozwoju całej dziedziny sztucznej inteligencji. Blisko dwadzieścia lat później naukowcy z DeepMind[36] zaczęli prezentować światu pełne możliwości swoich wynalazków. Były one efektem lat silnej dominacji światowej podejścia konseksjonistycznego, rozbudowy architektur sieci neuronowych i ostatecznie – wspominanego renesansu uczenia maszynowego. W kolejności chronologicznej przedstawiane były programy: AlphaGo, AlphaGo Master, AlphaGo Zero i AlphaZero.

W roku 2016 AlphaGo pokonał w go Lee Sedola – sklasyfikowanego na najwyższym, dziewiątym stopniu rankingu profesjonalistów, jednego z przedstawicieli ścisłej grupy najlepszych graczy na świecie, por. Koch (2016). Zaznaczyć należy, że konkurencja w ramach wywodzącej się z Azji Wschodniej gry stanowi dla komputerów znacznie większe wyzwanie

[35] Tom Mitchell jest amerykańskim naukowcem, ekspertem w dziedzinie sztucznej inteligencji.
[36] Od 2014 rokuDeepMind jest spółką zależną Google (Alphabet).

niż to, któremu wiele lat wcześniej sprostał Deep Blue[37]. W pierwszych pięciu dniach roku 2017 AlphaGo Master pokonał ówcześnie najwyżej notowanego w rankingu światowym gracza oraz osiemnastu innych zdobywających w przeszłości tytuły mistrzowskie, por. Internet[38]. Parę miesięcy później AlphaGo Zero pokonał AlphaGo Master na przestrzeni stu gier – wygrał każdą z nich, por. Knight (2017). Jeszcze przed końcem 2017 roku AlphaZero odniósł zwycięstwo nad AlphaGo Zero, przezwyciężając wcześniejszą wersję programu na dystansie stu gier z wynikiem sześćdziesiąt do czterdziestu, Silver i in. (2017r).

Kolejnym polem konkurencji między ludzką inteligencją, a uczonymi maszynowo sieciami neuronowymi stawały się bardziej skomplikowane gry komputerowe o charakterze konkurencyjnym. Rozwijany przez OpenAI program OpenAI Five stworzony został by grać w Dotę 2, mierząc się na raz z drużyną pięciu przeciwników. W kwietniu 2019 roku pokonał on zespół ówczesnych mistrzów świata stając się pierwszym systemem komputerowym wygrywającym z posiadaczami tego tytułu w jakiejkolwiek nowoczesnej grze e-sportowej, por. Internet[39]. Stworzony przez DeepMind program AlphaStar uczony był rywalizacji w ramach uniwersum Starcraft 2 – bardzo złożonej gry z gatunku strategii czasu rzeczywistego, por. Internet[40]. Konkurując przez Internet z ludźmi, w sierpniu 2019 program zdobył tytuł Grandmastera, plasując się wśród najlepszych 0,2% graczy, por. Statt (2019).[41]

Uczenie maszynowe podzielić można głębiej na wiele różnych poddziedzin w odniesieniu do wielu różnych kryteriów. Dokonując podstawowej separacji ze względu na paradygmat schematu uczenia, zdefiniować należy trzy grupy algorytmów realizujące podejścia nazywane: uczeniem nadzorowanym[42], uczeniem nienadzorowanym[43] oraz uczeniem ze wzmacnianiem.

---

[37] Zakładając średnią długość rozgrywki w przypadku szachów gracz ma do dyspozycji ponad 10^{120} kombinacji ruchowych, por. Shannon (1950). W przypadku go, mowa o okolicach 10^{360} możliwych kombinacji ruchowych.

[38] DeepMind (https://www.deepmind.com/research/highlighted-research/alphago/alphago-vs-alphago, dostęp: 30.07.22r.).

[39] Open AI (https://openai.com/blog/openai-five/ ; dostęp: 30.07.22r.).

[40] DeepMind (https://www.deepmind.com/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii; dostęp: 30.07.22r.).

[41] Obie współczesne gry komputerowe są wielokrotnie bardziej złożone niż szachy, czy go.

[42] Uczenie nadzorowane polega na przedstawieniu programowi danych wejściowych w połączeniu z pożądanymi wartościami danych wyjściowych. Zadaniem algorytmu jest zbudowanie modelu wykrywającego zależności oraz uogólniającego je do przypadków innych danych wejściowych, por. Stawka (2020). Zastosowaniami uczenia nadzorowanego są m. in. klasyfikacja i analiza regresji. Prostym przykładem realizującym pierwsze z zadań jest wspominany już wcześniej perceptron i uczący go algorytm Franka Rosenblatta.

[43] Uczenie nienadzorowane polega na prezentowaniu maszynie danych wejściowych, bez sugestii dotyczących danych wyjściowych. W ramach działania programu algorytm sam uczy sieć neuronową pożądanego działania, por. Stawka (2020).

Spośród wymienionych trzech podejść, uczenie ze wzmacnianiem realizuje paradygmat najbliższy ogólnemu zamysłowi sztucznej inteligencji. Algorytm modyfikujący sieć neuronową wchodzi w tym wypadku w interakcje z otoczeniem. Wybierając kolejne działania programu, otrzymuje pozytywne lub negatywne informacje zwrotne. System jest więc behawioralnie warunkowany do optymalizacji swoich decyzji względem celu maksymalizacji wartości otrzymywanych nagród. W praktyce efektywne działanie algorytmu opiera się między innymi na optymalnym balansowaniu dwóch klas aktywności – eksploracji i eksploatacji. Pierwsza z nich oznacza próbowanie nowych rozwiązań i jest realizowana przez podejmowanie decyzji losowych. W krótkim okresie jest to opcja bardziej ryzykowna, niesprzyjająca poszukiwaniom spójnego rozwiązania. Pozwala jednak odkrywać większe części przestrzeni decyzyjnej co przekłada się na dużą wartość dla procesu uczenia. W ten sposób stwarza większe szanse lepszej optymalizacji rozwiązania w długim okresie. Druga wyraża wykorzystanie starych rozwiązań, a więc podejmowanie decyzji optymalnej w ramach dostępnej wiedzy. W krótkim okresie jest więc bezpieczniejsza. Powoduje jednak ograniczenie eksploatacji przestrzeni decyzyjnej, co skutkuje potencjalnie mniejszymi możliwościami długookresowej optymalizacji rozwiązania. Przy założeniu minimalizacji kosztów, kwestia balansu przeradza się w pytanie o to jak bardzo optymalne musi być w danym wypadku optymalne rozwiązanie, por. Kaelbling i in. (1996r). Możliwe jest częściowe nadzorowanie uczenia ze wzmacnianiem. Udostępniając programowi zestaw danych wejściowych połączonych z danymi wyjściowymi do eksploatacji determinuje się w jego działaniu pewne możliwe do przewidzenia tendencje, por. Pandian i Noel (2018r). Uczenie ze wzmacnianiem w przeciwieństwie do wcześniej wymienionych dwóch podejść nie wymaga koniecznie dużych zbiorów danych, por. Internet[44]. Algorytm może modyfikować sieci neuronowe programu wykorzystując informacje generowane z władnych doświadczeń, w procesie treningu.

Schemat uczenia ze wzmacnianiem eksploatowany jest w ostatnich latach przez bardzo innowacyjne i skuteczne algorytmy. Przykładami są te zawarte w przywoływanych wcześniej systemach stworzonych przez DeepMind. Już przy AlphaGo Master przewaga programu nad ludźmi była ogromna. Szokował on często swoich przeciwników niekonwencjonalnymi posunięciami. Używał przy tym dokładnie algorytmu częściowo nadzorowanego uczenia ze wzmacnianiem. W pierwszej części procesu trenowania sieci neuronowej algorytm posługiwał się bazą danych znanych ludziom zagrań optymalnych względem zagrań

[44] Axcelerate (https://www.axcelerate.com.au/post/alphago-master-vs-alphago-zero-the-power-of-reinforcement-learning; dostęp: 02.08.22r.).

przeciwnika. W ten sposób znacząco ograniczał wielkość rozpatrywanego drzewa decyzyjnego[45]. Najwyraźniejszy w sensie wyników gier bezpośrednich skok jakościowy na przestrzeni kolejnych wersji programu nastąpił jednak między AlphaGo Master, a AlphaGo Zero dlatego, że usunięto wtedy z niego część nadzorowaną procesu uczenia sieci neuronowej. Maszyna podczas pierwszej gry treningowej wykonywała całkowicie losowe ruchy. W trakcie początkowych rozgrywek radziła sobie oczywiście gorzej niż swój poprzednik, jednak po jakimś czasie okazywała się już znacznie lepsza. Wyszło więc na to, że pominięcie setek lat ludzkiego doświadczenia w go pozwoliło algorytmowi uniknąć bardzo wielu nieefektywności[46]. Podobnych zależności dowodzą przykłady rozwoju algorytmów AlphaStar[47] oraz OpenAI Five[48].

Podsumowując, pojęcie sztucznej inteligencji na przestrzeni ostatnich dwustu lat przeszło drogę od fikcji literackiej, przez samą ideę filizoficzno-matematyczną, podejście symboliczne i związane z nim systemy ekspertowe, aż do algorytmów uczenia maszynowego i sieci neuronowych. W ostatnich latach szczególnie zauważalny jest rozwój uczenia ze wzmacnianiem. Pozwala on na co raz efektywniejsze rozwiązywanie kwantyfikowalnych problemów optymalizacyjnych. Zjawisko można zobrazować na przykładach programów doskonalących swoje zdolności w środowiskach złożonych gier kompetytywnych. Systemy nastawione na uczenie się jedynie z własnych doświadczeń, bez wsparcia ludzkich teorii, mogą dochodzić przy tym do lepszych rozwiązań problemów optymalizacyjnych dzięki efektywniejszej eksploatacji przestrzeni decyzyjnej.

---

[45] AlphaGo Master "wiedział", że przeciwnik najprawdopodobniej przy danym układzie wybierze zagranie A lub B – analiza równoległych układów na tym etapie gry nie miała sensu (redukcja szerokości drzewa). Analogicznie, "wiedział", że doprowadzając rozgrywkę do danego układu, najprawdopodobniej już wygra – analiza głębszych układów na tym etapie gry nie miała sensu (redukcja wysokości drzewa), por. Axcelerate (https://www.axcelerate.com.au/post/tree-branches-the-secret-to-alphagos-impressive-win; dostęp: 04.08.22r.). W ramach zmniejszonego drzewa algorytm z jednej strony posługiwał się schematem Monte Carlo Tree Search, a z drugiej realizował równocześnie działania eksploracyjne. Z próbowania nowych rozwiązań wynikały nietypowe dla ludzi i bardzo efektywne zagrania, por. LiveBook Manning (https://livebook.manning.com/book/deep-learning-and-the-game-of-go/chapter-14/; dostęp: 05.08.22r.).

[46] Axcelerate (https://www.axcelerate.com.au/post/alphago-master-vs-alphago-zero-the-power-of-reinforcement-learning; dostęp: 02.08.22r.).

[47] AlphaStar wyposażony był w algorytm częściowo nadzorowanego uczenia ze wzmacnianiem. Gra była bardzo złożona i w przypadku rozpoczynania od strategii absolutnie losowych trening dostarczałby prawie wyłącznie negatywnych informacji zwrotnych skutkując minimalnymi zawężeniami przestrzeni decyzyjnej. Program wyposażono w bazę danych związanych z zagraniami profesjonalistów. Od tego poziomu sam generował doświadczenia, por. Kelion (2019).

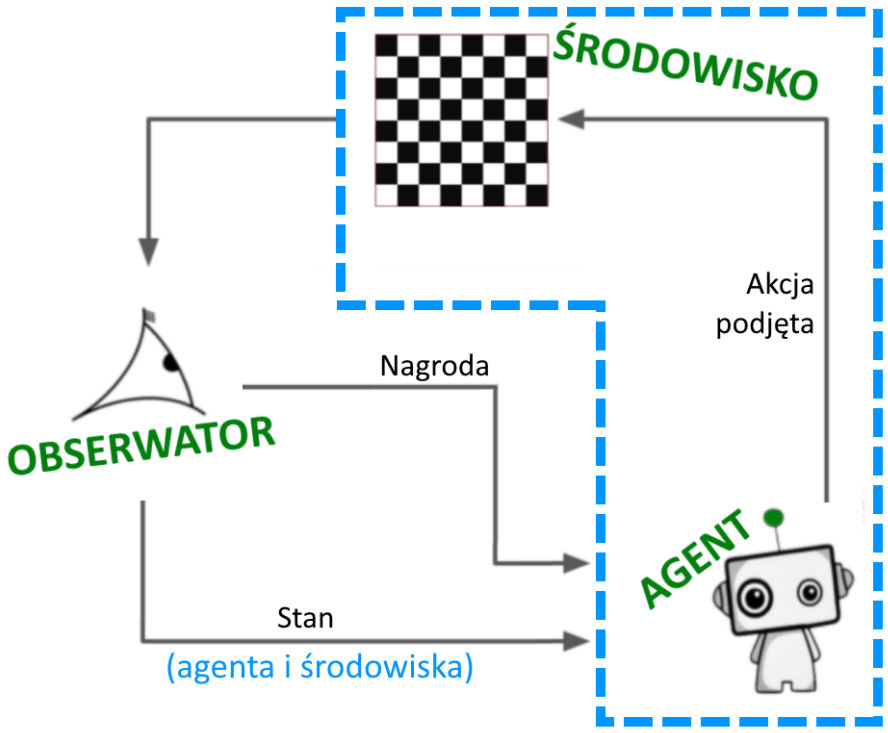
[48] OpenAI Five, zaopatrzono w algorytm wykorzystujący od samego początku wyłącznie własne doświadczenia. Rozgrywka była w większym stopniu możliwa do podziału na mniejsze problemy optymalizacyjne. Koszt czasu uczenia programu był i tak duży. Jak mówią twórcy, do osiągnięcia poziomu pozwalającego wygrywać z mistrzami świata potrzebował równowartości 10tys. lat ciągłego treningu, por. Open AI (https://openai.com/blog/openai-five/ ; dostęp: 30.07.22r.).

Jak już argumentowano we wcześniejszym rozdziale, efektywna alokacja kapitału wyrażana jak najlepszą optymalizacją portfeli akcyjnych inwestorów giełdowych jest wielowymiarowo istotna dla gospodarki. Rozwiązywanie problemu optymalizacji portfela akcyjnego przy użyciu narzędzi analizy technicznej można określić natomiast podejściem niebezpodstawnym, osadzonym w faktycznej metodologii rynkowej. Wykorzystanie najnowszych algorytmów uczenia ze wzmacnianiem jest natomiast pomysłem ciekawym, aktualnym i potencjalnie rozwojowym naukowo.

3.2 Schemat działania algorytmów uczenia ze wzmacnianiem

Schemat działania generycznego algorytmu uczenia ze wzmacnianiem opisać można przy pomocy prostego diagramu odwołującego się do pojęć znanych z definicji sztucznej inteligencji, przywoływanej w pierwszym rozdziale.

Uczenie ze wzmacnianiem.



Źródło: opracowanie własne.

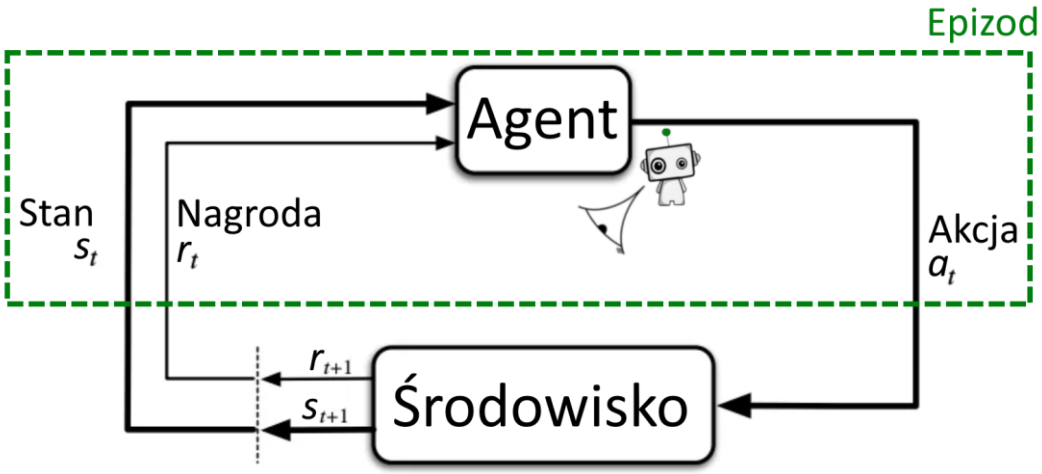
Rysunek 3.1. *Tradycyjnie pojęcia Agenta, Środowiska i Obserwatora symbolizuje się odpowiednio robotem, szachownicą i okiem. Zachowując zgodność z definicją i znaczenie schematu postać Agenta można zastąpić Aktorem.*

Na Rysunku 3.1. widać interakcje zachodzące między Agentem, Środowiskiem i Obserwatorem. Można przyjąć, że przedstawiony symbolicznie ciąg wydarzeń zaczyna się kiedy Agent otrzymuje od Obserwatora informację o stanie Środowiska i nagrodę. Jest

względem wiadomości przekazywanych mu przez Obserwatora częścią Środowiska (co wyraża niebieska ramka). Na podstawie treści z pierwszego rozdziału określić można go programem wyposażonym w algorytm uczenia ze wzmacnianiem oraz architekturę sieci neuronowej (odpowiednie dla rozpatrywanego problemu optymalizacyjnego i Środowiska). Jego cel to jak największa nagroda. Podejmuje akcję, która wpływa na Środowisko. Obserwator, widząc to, przekazuje Agentowi sygnał zwrotny dotyczący zaktualizowanego stanu Środowiska i nagrody za wykonaną akcję.

Upraszczając schemat, można założyć, Agent ma bezpośredni, pełen przegląd Środowiska, wtedy rola Obserwatora jest z nim utożsamiana. Aby rozpatrywać matematyczny model niezbędne jest zadeklarowanie zmiennych opisujących zachodzące w nim interakcje. Chcąc natomiast analizować ciągi zdarzeń dłuższe niż jeden cykl należy wprowadzić pojęcie czasu. Ten najłatwiej mierzyć w ujęciu dyskretnym względem kolejno zachodzących faktów.

Uczenie ze wzmacnianiem z dyskretnym ujęciem czasu.



Źródło: opracowanie własne.

Rysunek 3.2. *Przeniesienie symbolu oka obserwującego Środowisko w pobliże symbolu robota oznacza przypisanie Agentowi roli Obserwatora.*

Rysunek 3.2. przedstawia zmatematyzowany schemat działania algorytmu uczenia ze wzmacnianiem przy dyskretnym ujęciu czasu, gdzie:

- a_t to akcja podjęta przez Agenta w momencie t,
- s_{t+1} to stan wynikający z akcji a,
- r_{t+1} to nagroda wynikająca z akcji a,
- A to zbiór dostępnych Agentowi akcji,
- S to zbiór stanów,
- R to zbiór nagród.

Ujęcie dyskretne czasu opiera się na założeniu, że w każdym momencie t do Agenta dociera informacja o aktualnym stanie s_t i nagrodzie r_t. Jego celem w tym ujęciu jest maksymalizacja wartości otrzymywanych nagród, Wybiera ze zbioru dostępnych akcji A pojedynczą akcję a_t, którą działa na Środowisko, wywołując jego zmianę. Determinowana jest wtedy informacja o nowym stanie aktualnym s_{t+1} i nagrodzie r_{t+1}. Tak zdefiniowane, ograniczone w ramach jednej wartości t, przedziały czasowe nazywa się epizodami.

Można wprowadzić pojęcie skumulowanej (wartości) nagrody G_t definiowane jako:

$$G_t = \gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots = \sum_{k=0} \gamma^k r_{t+k+1}$$

, gdzie $\gamma \in [0,1)$ jest czynnikiem dyskontującym. W ten sposób z nagrodami bliższymi czasowo wiąże się większą wartość.

- Agent charakteryzujący się wartością γ bliską 1 jest ‘cierpliwy’ – ocenia wartość odległych nagród podobnie do wartości nagród bliskich.
- Agent charakteryzujący się wartością γ bliską 0 jest ‘niecierpliwy’ – ocenia wartość odległych bardzo nagród nisko względem wartości nagród bliskich.

Agent jest behawioralnie warunkowany do modyfikacji swoich zachowań za pośrednictwem informacji zwrotnych otrzymywanych ze Środowiska. Można więc stwierdzić, że jego proces uczenia się ma na celu wyznaczenie schematu wyboru akcji przekładającego się na maksymalizację skumulowanej (wartości) nagrody G_t otrzymywanej na przestrzeni kolejnych okresów.

Wprowadzając do naszych rozważań probabilistykę ogólny schemat wyboru akcji opisać można przy pomocy następującej funkcji:

- $\pi: A \times S \to [0,1]$;
- $\pi(a,s) = \Pr(a_t = a \mid s_t = s)$.

Przeprowadza ona zmienne stanu Środowiska i dostępnej Agentowi akcji na wartość prawdopodobieństwa obrania tej akcji w tym stanie. Będący celem Agenta optymalny schemat wyboru akcji – przekładający się na maksymalizację skumulowanej wartości nagrody G_t – oznaczyć można jako π^*. W źródłach anglojęzycznych funkcję π nazywa się *policy*.

Przejścia między stanami w przedstawionym modelu matematycznym interpretować można jako łańcuchy Markowa. Z teorii prawdopodobieństwa wiadomo, że w takim razie rozkład prawdopodobieństwa każdego przejścia zależy jedynie od rozkładu

prawdopodobieństwa przejścia poprzedniego. Funkcja rozkładu prawdopodobieństwa przejścia ze stanu s i przy wykonaniu akcji a do stanu s' związanego z nagrodą r oznaczona jest przez:

$$p(s', r | s, a) = \Pr(s_t = s',\ r_t = r \mid s_{t-1} = s,\ a_{t-1} = a)$$

Funkcja nagrody związanej z przejściem ze stanu s do stanu s' w wyniku akcji a wskazywana jest przez:

$$r(s, a, s') = \mathbb{E}[r_t \mid s_{t-1} = s, a_{t-1} = a, {s'}_t = s']$$

Ze względu na ujęcie przejść między stanami jako skończonych łańcuchów Markowa cel Agenta, określony wyznaczeniem funkcji π^* maksymalizującej skumulowaną nagrodę G_t utożsamić można z wyznaczeniem funkcji π' maksymalizującej nagrodę w każdym ze stanów ($\pi^* = \pi'$).

W takim modelu określić można funkcje wartości: $V^\pi(s)$ oraz $Q^\pi(a, s)$, które definiuje się w następujący sposób.

1 $V^\pi(s) = \mathbb{E}_\pi[G_t \mid s_t = s]$

2 $Q^\pi(a, s) = \mathbb{E}_\pi[G_t \mid s_t = s,\ a_t = a]$

Funkcja $V^\pi(s)$ przeprowadza zmienną stanu Środowiska na wartość oczekiwaną skumulowanej (wartości) nagrody G_t przy założeniu stanu początkowego s i zastosowaniu funkcji π. Funkcja $Q^\pi(a, s)$ przeprowadza zmienne akcji i stanu na wartość oczekiwaną skumulowanej (wartości) nagrody G_t przy założeniu stanu początkowego s, akcji a i zastosowaniu funkcji π.

Z definicji funkcji wartości wynika, że między jednoznacznym określeniem każdej z nich, a jednoznacznym określeniem funkcji π zachodzi implikacja obustronna. Wyznaczenie jakiejś postaci funkcji: $V^\pi(s)$ albo $Q^\pi(a, s)$ równoznaczne jest wyznaczeniu jakiejś postaci funkcji π. W takim razie, cel Agenta związany z wyznaczeniem optymalnej funkcji π^* utożsamiać można z wyznaczeniem optymalnej w tym sensie funkcji $V^{\pi^*}(s)$ albo $Q^{\pi^*}(a, s)$ (i na odwrót). Problem wyznaczenia funkcji π^* można rozpatrywać dokonując po kolei:

1 wyprowadzenia równania Bellmana dla funkcji $V^\pi(s)$ albo funkcji $Q^\pi(a, s)$,
2 jego optymalizacji, przy pomocy równania optymalności Bellmana, do postaci odpowiadającej funkcji $V^{\pi^*}(s)$ lub $Q^{\pi^*}(a, s)$.

W źródłach anglojęzycznych każda z nich nazywana jest *value function*. Dodatkowo pierwszą nazywa się *state value function*, a drugą *state-action value function*.

Z poprzednich rozważań wiemy, że funkcję $V^{\pi}(s)$, w zależności od sposobu zapisania skumulowanej (wartości) nagrody G_t określić można jako:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[G_t \mid s_t = s]$$

lub:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[\gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots \mid s_t = s]$$

lub:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s]$$

możemy wyciągnąć pierwszą nagrodę z sumy, wtedy:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+1+k+1} \mid s_t = s]$$

i zmieniając zapis:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[r_{t+1} + \gamma G_{t+1} \mid s_t = s]$$

z liniowości wartości oczekiwanej, wiemy, że możemy rozdzielić ją względem sumy:

$$V^{\pi}(s) = \mathbb{E}_{\pi}[r_{t+1} \mid s_t = s] + \mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s]$$

przekształcamy pierwszy składnik sumy:

$$\mathbb{E}_{\pi}[r_{t+1} \mid s_t = s] = \sum_{a} \left[\pi(s,a) \sum_{s'} [p(s', r \mid s, a) r(s, a, s')] \right]$$

Jest to wartość oczekiwana nagrody w momencie $t+1$ przy zadanym stanie s. Można wyrazić ją sumując po wszystkich możliwych akcjach a i wszystkich możliwych zwróconych przez Środowisko stanach s' mnożonych przez nagrody wiązane z przejściem ze stanu s do stanu s' w wyniku akcji a. Obliczamy drugi składnik sumy:

$$\mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s] = \mathbb{E}_{\pi}[\gamma \sum_{k=0}^{\infty} \gamma^k r_{t+1+k+1} | s_t = s]$$

γ możemy wyciągnąć poza wartość oczekiwaną:

$$\mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s] = \gamma \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} \gamma^k r_{t+1+k+1} | s_t = s]$$

analogicznie do poprzedniego składnika sumy, mamy, że:

$$\mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s] = \gamma \sum_{a} \left[\pi(s,a) \sum_{s'} \left[p(s', r \mid s, a) \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} \gamma^k r_{t+1+k+1} | s_{t+1} = s'] \right] \right]$$

z definicji funkcji $V^{\pi}(s)$ wynika, że:

$$V^{\pi}(s') = \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} \gamma^{k} r_{t+1+k+1} | s_{t+1} = s']$$

i w takim razie:

$$\mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s] = \gamma \sum_{a} \left[\pi(s,a) \sum_{s\prime} [p(s', r \mid s, a) V^{\pi}(s')] \right]$$

w końcu mamy drugi składnik sumy:

$$\mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s] = \sum_{a} \left[\pi(s,a) \sum_{s\prime} [p(s', r \mid s, a) \gamma V^{\pi}(s')] \right]$$

łączymy sumę składników i wychodzi nam ostateczna forma równania Bellmana dla funkcji $V^{\pi}(s)$:

$$V^{\pi}(s) = \sum_{a} \left[\pi(s,a) \sum_{s\prime} \left[p(s', r \mid s, a) \big(r(s,a,s') + \gamma V^{\pi}(s') \big) \right] \right]$$

Proces przebiega zupełnie analogicznie do ciągu wnioskowań związanego z wyprowadzeniem równania Bellmana dla $V^{\pi}(s)$. Z poprzednich rozważań wiemy, że funkcję $Q^{\pi}(s)$ określić można jako:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[G_t \mid s_t = s,\ a_t = a]$$

lub:

$$Q(s,a) = \mathbb{E}_{\pi}[\gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots \mid s_t = s,\ a_t = a]$$

lub:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \mid s_t = s,\ a_t = a]$$

możemy wyciągnąć pierwszą nagrodę z sumy, wtedy:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[r_{t+1} + \ \gamma \sum_{k=0}^{\infty} \gamma^{k} r_{t+1+k+1} \mid s_t = s,\ a_t = a]$$

i zmieniając zapis:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[r_{t+1} + \gamma G_{t+1} \mid s_t = s,\ a_t = a]$$

z liniowości wartości oczekiwanej, wiemy, że możemy rozdzielić ją względem sumy:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[r_{t+1} \mid s_t = s,\ a_t = a] + \mathbb{E}_{\pi}[\gamma G_{t+1} | s_t = s,\ a_t = a]$$

od razu mamy pierwszy składnik sumy:

$$\mathbb{E}_\pi[r_{t+1} \mid s_t = s,\ a_t = a] = \sum_{s\prime}[p(s', r \mid s, a) r(s, a, s')]$$

liczymy drugi składnik sumy, w którym γ możemy wyciągnąć poza wartość oczekiwaną:

$$\mathbb{E}_\pi[\gamma G_{t+1} | s_t = s,\ a_t = a] = \gamma \mathbb{E}_\pi[\ G_{t+1} | s_t = s,\ a_t = a]$$

wychodzi na to, że:

$$\mathbb{E}_\pi[\gamma G_{t+1} | s_t = s,\ a_t = a] = \gamma \sum_{s\prime}\left[p(s', r \mid s, a) \sum_{a\prime}\left[\pi(s', a') \mathbb{E}_\pi[\ G_{t+1} | s_{t+1} = s',\ a_t = a]\right]\right]$$

z definicji funkcji $Q^\pi(s, a)$ wynika, że:

$$Q^\pi(s', a') = \mathbb{E}_\pi[\ G_{t+1} | s_{t+1} = s',\ a_t = a]$$

i w takim razie:

$$\mathbb{E}_\pi[\gamma G_{t+1} | s_t = s,\ a_t = a] = \gamma \sum_{s\prime}\left[p(s', r \mid s, a) \sum_{a\prime}[\pi(s', a') Q^\pi(s', a')]\right]$$

w końcu mamy drugi składnik sumy:

$$\mathbb{E}_\pi[\gamma G_{t+1} | s_t = s,\ a_t = a] = \sum_{s\prime}\left[p(s', r \mid s, a) \gamma \sum_{a\prime}[\pi(s', a') Q^\pi(s', a')]\right]$$

łączymy sumę składników i wychodzi nam ostateczna forma równania Bellmana dla funkcji $Q^\pi(s, a)$:

$$Q^\pi(s, a) = \sum_{s\prime}\left[p(s', r \mid s, a)\left[r(s, a, s') + \gamma \sum_{a\prime}[\pi(s', a') Q^\pi(s', a')]\right]\right]$$

Równanie optymalności Bellmana mówi, że w przypadku optymalnej funkcji $V^{\pi^*}(s)$, dla każdego ze stanów Środowiska s zachodzi następująca zależność:

$$V^{\pi^*}(s) = \max_\pi V^\pi(s)$$

Funkcja $V^{\pi^*}(s)$, zgodnie z definicją $V^\pi(s)$ musi przeprowadzać zmienną stanu s na wartość oczekiwaną skumulowanej (wartości) nagrody G_t przy założeniu stanu początkowego s i zastosowaniu funkcji π^*. Jako, że funkcja $V^{\pi^*}(s)$ wykorzystuje funkcję π^* i w konsekwencji spełnia wyżej opisane równanie optymalności Bellmana, to dla każdego ze stanów s zwracana przez nią wartość oczekiwana skumulowanej (wartości) nagrody G_t musi być równa wartości

oczekiwanej wynikającej z podjęcia najlepszej akcji w tym stanie – czyli akcji takiej, dla której wartość oczekiwana skumulowanej (wartości) nagrody G_t jest największa. W takim razie, mamy że:

$$V^{\pi^*}(s) = \max_{\pi} V^{\pi}(s) = \max_{a} Q^{\pi^*}(s,a)$$

więc z definicji $Q^{\pi}(s,a)$:

$$V^{\pi^*}(s) = \max_{a} \mathbb{E}_{\pi^*}[G_t \mid s_t = s,\ a_t = a]$$

więc można rozpisać jak poprzednio:

$$V^{\pi^*}(s) = \max_{a} \mathbb{E}_{\pi}[r_{t+1} + \gamma G_{t+1} \mid s_t = s,\ a_t = a]$$

i z rozważań o $V^{\pi^*}(s)$, mamy że:

$$V^{\pi^*}(s) = \max_{a} \mathbb{E}_{\pi}\left[r_{t+1} + \gamma V^{\pi^*}(s_{t+1}) \,\middle|\, s_t = s,\ a_t = a\right]$$

Kierując się takim samym ciągiem wnioskowania jak przy wyprowadzaniu równania Bellmana dla funkcji $V^{\pi}(s)$, otrzymujemy ostateczną postać równania optymalności Bellmana dla funkcji $V^{\pi^*}(s)$:

$$V^{\pi^*}(s) = \max_{a} \sum_{s\prime} \left[p(s', r \mid s, a)\left(r(s,a,s') + \gamma V^{\pi^*}(s')\right)\right]$$

Posługując się analogicznym ciągiem wnioskowania i znając schemat wyprowadzenia równania Bellmana dla funkcji $Q^{\pi}(s,a)$, otrzymać można ostateczną postać równania optymalności Bellmana dla funkcji $Q^{\pi^*}(s,a)$:

$$Q^{\pi^*}(s,a) = \mathbb{E}_{\pi}\left[\, r_{t+1} + \gamma \max_{a} Q^{\pi^*}(s_{t+1}, a') \,\middle|\, s_t = s,\ a_t = a\right]$$

$$Q^{\pi^*}(s,a) = \sum_{s\prime} \left[p(s', r \mid s, a)\left(\mathrm{r}(s,a,s') + \gamma \max_{a\prime} Q^{\pi^*}(s', a')\right)\right]$$

We wszystkich rozważaniach przyjęty został model rzeczywistości, w którym Agent będąc w stanie s i podejmując działanie a nie wie do jakiego stanu $s+1$ to doprowadzi i nie wie jaką dostanie nagrodę. Agent jest więc doskonałym Obserwatorem, ale nie może idealnie przewidywać wpływu swoich akcji na Środowisko i wysokości nagród (jest zmuszony je szacować). W literaturze często przyjmuje się upraszczające założenie o tym, że z perspektywy Agenta wybór akcji a w danym stanie a jest deterministyczny względem stanu przyszłego $a+1$. W takim, mniej złożonym, wypadku ostateczne równania wyglądają

nieco inaczej. Ogólna Idea pozostaje jednak taka sama. Ostatecznym celem Agenta jest estymacja przedstawionych, optymalnych postaci funkcji wartości i na tym zadaniu, zgodnie z powyższym modelem, skupiają się algorytmy uczenia ze wzmacnianiem.

3.3 Selekcja algorytmów

Zagadnienie wykorzystania uczenia ze wzmacnianiem do automatyzacji inwestycji giełdowych jest szczególnie w ostatnich latach rozwojowe. Ma to swoje odzwierciedlenie w publikacjach naukowych.

W swojej pracy opublikowanej w 2017 roku Luca Di Persio i Oleksandr Honchar porównali 3 rodzaje architektury sieci neuronowych: zwykłej wielowarstwowej RNN (Recurrent Neural Network, jest to sieć będąca następcą klasycznej wielowarstwowej FNN – Feedforward Neural Network), LSTM (Long Short Term Memory) i GRU (Gated Recurrent Unit). Celem badań była ocena ich względnej efektywności. Mierzono ją na podstawie dokonywanych przez poszczególne sieni neuronowe przewidywań ruchów ceny akcji Google (Alphabet). Wyniki wykazały znaczną przewagę architektury typu LSTM nad pozostałymi dwoma w środowisku rozpatrywanego problemu. Osiągała ona w testach aż do 72% poprawności decyzji na przedziałach pięciodniowych. Okazało się, że ten rodzaj architektury pozwala na sprawne przetwarzanie nie tylko pojedynczych zbiorów danych, ale i ciągów zbiorów danych Badanie prowadzone było na przedziale 5 lat i wykazało też, że sieci neuronowe do względnie stabilnego działania wymagają bardzo wielu okresów treningowych (zwanych epokami), a więc również szerokiego, pierwotnie branego, przedziału czasowego, por. Persio i Honchar (2017).

W 2019 roku naukowcy – Yang Li, Wanshan Zheng, Zibin Zheng – przedstawili wyniki swoich badań dotyczących optymalizacji budowy systemów uczenia ze wzmacnianiem w środowisku giełdowym. Rozpatrywali efektywność algorytmów DQL (Deep Q-learning) i A3C (Asynchronous Advantage Actor Critic) przy współpracujący z różnymi typami architektury sieci neuronowych. Tworzonym w ten sposób systemom dawali trenować na zbiorze danych dotyczącym notowań 3 amerykańskich spółek giełdowych z różnych sektorów gospodarczych (Apple, IBM, Procter&Gamble) i kontraktów terminowych pochodnych względem 2 indeksów (S&P 500 i HS300). Proces nauki programów przebiegał więc zgodnie z paradygmatem analizy technicznej. Autorzy pracy nie wyposażali ich przy tym w żadne znane ludziom zasad, reguł czy sugestie dotyczące strategii inwestycyjnych. Zbudowane systemy w ogólności świetnie sobie z wyzwaniem. Każdy z rozpatrywanych wariantów maszyny pokonywał generyczną strategię Buy and Hold w ramach każdego

z rozpatrywanych rynków. Badanie uwierzytelniło wnioski prac innych naukowców – LSTM ponownie okazał się rodzajem architektury sieci neuronowej bardziej adekwatnym do zastosowań giełdowych niż standardowe FNN, czy CNN (Convolutional Neural Network). Potwierdziło się również przypuszczenie autorów, mówiące o tym, że w środowisku giełdowym algorytmy typu policy-based sprawdzą się lepiej niż algorytmy typu value-based. W ramach każdego z rozpatrywanych rynków A3C zwracał lepsze wyniki niż DQL. Taki stan rzeczy wynika z tego, że rozpatrywany problem jest zbyt skomplikowany żeby program mógł nauczyć się efektywnie aproksymować funkcję wartości. Algorytmy operujące natomiast bezpośrednio w przestrzeni funkcji π (policy, więcej o tym w kolejnym rozdziale) pozwalają na względnie racjonalne przybliżanie wartości optymalnych. Algorytmy typu actor-critic, takie jak A3C, są natomiast połączeniem algorytmów policy-based oraz value-based, por. Li i in. (2019).

Praca Jonathana Sadighiana z 2019 roku odnosiła się do pomysłu stworzenia programu działającego według założeń uczenia ze wzmacnianiem i postawienia go w roli animatora rynku kryptowalutowego handlującego w ramach pojedynczych walorów giełdowych. Autor testował w niej możliwości dwóch algorytmów: PPO (Proximal Policy Optimization) oraz A2C (Advantage Actor Critic) przy zastosowaniu architektury wielowarstwowej sieci neuronowej typu MLP (Multilayer Perceptron, jest to rodzaj klasycznej FNN). Do procesu treningowego programów, w zgodzie z paradygmatem analizy technicznej, wykorzystywał jedynie dane historyczne dotyczące transakcji giełdowych na 3 parach walutowych – zestawiających USD z BTC, ETH oraz LTC. Nie przedstawiał im żadnych ludzkich wskazówek dotyczących strategii inwestycyjnych. Wyniki okazały się bardzo dobre. Programy generowały dodatnie dzienne wyniki finansowe na różnych zbiorach danych, por. Sadighian (2019).

Opublikowane w 2019 roku owoce wysiłków naukowych Evgeny'a Ponomareva, Ivana Oseledetsa i Andrzeja Cichockiego dotykały bardzo podobnego problemu. Autorzy badali zdolności maszyny realizującej ideę uczenia ze wzmacnianiem do osiągania zysku na giełdzie moskiewskiej. Opisane przez nich programy wykorzystywały algorytm A3C oraz różne rodzaje architektury wielowarstwowej sieci neuronowej typu LSTM (Long short-term memory). Uczyły się od początku samodzielnie na zbiorach danych historycznych transakcji giełdowych zawieranych na kontrakty futures pochodne względem rosyjskiego indeksu RTX. Najefektywniejsze z nich osiągnęły bardzo satysfakcjonujące wyniki w postaci rocznej stopy zwrotu na poziomie 66% po uwzględnieniu kosztów transakcji, por. Ponomarev i in. (2019).

W roku 2020 Thibaut Théate i Damien Ernst przedstawili swoją pracę dotyczącą zastosowania programów uczenia ze wzmacnianiem do automatyzacji inwestycji w ramach pojedynczych walorów giełdowych. Autorzy testowali możliwości rozwiązania bazującego na algorytmie DQL z użyciem architektury wielowarstwowej sieci neuronowej klasycznego typu FNN. Stworzony przez nich system trenował swoje umiejętności całkowicie samodzielnie, na danych dotyczących cen historycznych 25 akcji spółek i kontraktów terminowych pochodnych względem 5 indeksów reprezentujących wspólnie 5 sektorów gospodarki na przestrzeni 3 regionów świata. Badanie zwróciło obiecujące wyniki. Program wykorzystujący uczenie ze wzmacnianiem okazał się przeciętnie lepszy od programów trzymających się ściśle 4 tradycyjnych, znanych ludziom strategii inwestycyjnych związanych z metodami analizy technicznej. W uwagach do swojego badania autorzy zwrócili uwagę na to, że analizując wyniki prac innych naukowców doszli do wniosku, że wykorzystanie przez nich algorytmu PPO oraz architektury sieci neuronowej typu LSTM mogłoby zwrócić jeszcze lepsze rezultaty, Théate i Ernst (2020).

W 2020 roku światu ukazał się FinRL – biblioteka open-source poświęcona zastosowaniom algorytmów uczenia ze wzmacnianiem w środowiskach giełdowych. Implementowała ona miedzy innymi takie algorytmy jak: DQL, PPO, A2C, SAC (Soft Actor-Critic), określane w opisie FinRL najaktualniejszymi osiągnięciami nauki (ang. *state-of-the-art*). Wskazując podstawowe zastosowania swojej biblioteki autorzy wymienili handel akcjami jednej spółki, handel akcjami wielu spółek oraz ogólną optymalizację portfela inwestycyjnego, por. Liu i in. (2020).

Gang Huang, Xiaohua Zhou i Qingyang Song opublikowali w 2021 roku swoją pracę dotyczącą możliwości optymalizacji akcyjnego portfela inwestycyjnego przy pomocy programu działającego według założeń uczenia ze wzmacnianiem. Autorzy wykorzystali algorytm DDPG (Deep Deterministic Policy Gradient) oraz architekturę sieci neuronowej typu VGG (Visual Geometry Group, jest to rodzaj klasycznej wielowarstwowej sieci neuronowej FNN). W procesach treningowych swojego systemu wykorzystywali, zgodnie z paradygmatem analizy technicznej, dane historyczne dotyczące transakcji giełdowych związanych z wybranymi całkowicie losowo 4 spółkami chińskiego indeksu CSI300. Zwracane przez program wyniki okazywały się bardzo korzystne. Przy zarządzanym aktywnie portfolio inwestycyjnym przewyższały notowania rynkowe na różnych zbiorach danych nawet przy uwzględnieniu kosztów transakcyjnych, por. Huang i in. (2021).

W 2021 roku wyniki swoich badań przedstawili Badr Hirchoua, Brahim Ouhbi i Bouchra Frikh. Dotyczyły one automatyzacji procesu inwestycyjnego w ramach pojedynczych walorów giełdowych przy pomocy maszyny realizującej ideę uczenia ze wzmacnianiem. Grupa naukowców użyła przy tym algorytmów DQL i PPO oraz standardowej architektury wielowarstwowej sieci neuronowej typu FFN. Pozwolili trenować programowi swoje umiejętności od początku bez względu na znane ludziom metody analizy technicznej. Dane treningowe dotyczyły cen historycznych akcji 7 spółek i jednej pary walutowej. System okazał się niezwykle skuteczny. W ramach handlu każdym z 8 walorów giełdowych zwracał względnie stabilny wzrost zakumulowanego w czasie zysku wobec występującego wtedy na rynku trendu bocznego. Uwagi autorów dotyczyły przewidywanego przez nich wysokiego potencjału podobnych badań z ukierunkowaniem na optymalizację portfela inwestycyjnego, por. Hirchoua i in. (2021).

Przytoczone przykłady prac badawczych legitymizują wykorzystanie algorytmów uczenia ze wzmacnianiem przy rozwiązywaniu problemu optymalizacji struktury zarządzanego aktywnie portfela akcyjnego w paradygmacie analizy technicznej.

Jest to zagadnienie poruszane rzadziej niż ruchy inwestycyjne w ramach pojedynczego waloru giełdowego, a przy tym potencjalnie bardziej złożone. Rozważanie wielu spółek mapujących przekrój wielu sektorów gospodarczych można uznać za wyzwanie szczególnie rozwojowe. Przytaczane badania empiryczne pozwalają również estymować najlepsze rozwiązania względem samej metodologii. Uzasadnione jest przypuszczenie, że w przypadku środowiska giełdowego algorytmy typu policy-based sprawdzają się lepiej niż algorytmy typu value-based. PPO (policy-based) oraz A2C (między policy-based, a value-based) to narzędzia wykorzystywane w tego typu problemach optymalizacyjnych. Autorzy są również zgodni co do tego, że architektura sieci neuronowej typu LSTM sprawdza się najlepiej do takiego rodzaju implementacji.

Ostatecznie, metodami, które zostały wybrane do osiągnięcia celu są programy uczenia ze wzmacnianiem wyposażone w algorytmy PPO i A2C oraz sieci neuronowe typu LSTM. Mają one trenować swoje umiejętności jedynie na podstawie własnych doświadczeń.

3.3.1 Algorytm A2C – Advantage Actor Critic

Pseudokod algorytmu A2C.

1. Inicjalizuj parametry funkcji π oraz stan oznaczane jako: θ, s
2. Powtarzaj (…):
   2.1. Losuj akcję a korzystając z funkcji π_ϑ
   2.2. Podejmij akcję a
   2.3. Odbieraj nagrodę r i następny stan s'
   2.4. Licz Advantage (TD Error) zgodnie z formułą:
   $$\delta_{t_k} = r(s_t, a_t) + \gamma V(s_{t+1}) - V(s_t)$$
   2.5. Aktualizuj parametry Aktora zgodnie z formułą:
   $$\theta \leftarrow \theta + \omega(\frac{1}{N}\sum_{i=1}^{N}[\sum_{t=0}^{T}\gamma^t \nabla_\theta \log \pi_\theta(a_{i,t} \mid s_{i,t})(r(s_t, a_t) + \gamma V(s_{t+1}) - V(s_t)))$$
   2.6. Aktualizuj $V(s_t)$ używając celu:
   $$r(s_t, a_t) + \gamma V(s_{t+1})$$
   (…) Do czasu aż s nie jest stanem oczekiwany

Źr: opr. wł. na podst. Zhang T. i in. (2020).

Cechą charakterystyczną algorytmu A2C (Advantage Actor Critic) jest uczenie dwóch komponentów, które opisuje się jako osobne podmioty.

1. Krytyk (Critic) – szacuje zagregowaną wartość nagród (V-value) osiągalną potencjalnie w przyszłości dzięki przebywaniu w stanie s.
2. Aktor (Actor) – przedstawia rozkład prawdopodobieństwa wykonania akcji a w zależności od stanu s.

Uczenie polega na równoległej aktualizacji Krytyka i Aktora.

1. Zadaniem krytyka jest jak najlepsza estymacja optymalnych wartości zagregowanych nagród (V-value) – definiowanych maksymalnymi zwrotami osiągalnymi z danego stanu.
2. Aktor ma zwracać taką dystrybucję akcji dla stanów, aby wykonywane akcje pozwalały na maksymalizację zdyskontowanych sum nagród – osiąganie optymalnych wartości V-value.

Sieć neuronowa Aktora jest modyfikowana tak, aby maksymalizować prawdopodobieństwo wykonania danej akcji dla danego stanu, jeśli związane z tym Advantage (TD error) przyjmuje wartość pozytywną, oraz minimalizować, jeżeli charakteryzuje je wartość negatywna. W trakcie aktualizacji liczone są pochodne logarytmu prawdopodobieństwa wykonania danej akcji a w danym stanie s dla każdego parametru sieci neuronowej. Następnie każdy parametr, za pomocą wspomnianych różniczek przemnożonych przez skalar oraz wartości Advantage, aktualizowany jest poprzez dodawanie.

Wartości zwracane przez Krytyka dla pary (stan s, akcja a) są reewaluowane po otrzymaniu nagrody r za wykonanie akcji a w danym stanie s oraz następnego stanu s' i akcji a', którą algorytm wykona dla następnego stanu (s'). Wartość zwracana przez Krytyka dla pary (stan s, akcja a) po aktualizacji ma zbliżyć się do sumy nagrody i ułamka wartości zwracanej przez Krytyka dla pary (następny stan s', następna akcja a').

Aktualizacja Krytyka działa podobnie do aktualizacji Aktora. Odbywa się dzięki liczeniu pochodnych. Inne są natomiast funkcje, które się różniczkuje oraz czynniki przez które pochodne są mnożone przed dodaniem ich do obecnych wartości przyjmowanych przez parametry sieci.

3.3.2 Algorytm PPO – Proximal Policy Optimization

Pseudokod algorytmu PPO.

1. Inicjalizuj parametry funkcji π oraz funkcji wartości V oznaczane jako: θ_0, φ_0
2. Dla kolejnych iteracji (k = 0, 1, 2, …) rób:
   2.1. Zbieraj trajektorie $D_k = \{\tau_i\}$ działając zgodnie z funkcją $\pi_k = \pi(\theta_k)$ w Środowisku
   2.2. Dyskontuj przyszłe nagrody $\hat{R}_t$
   2.3. Licz estymacje Advantage $\hat{A}_t$ (używając dowolnej metody estymacji Advantage) bazując na bieżącej wartości funkcji wartości V_{φ_k}
   2.4. Aktualizuj funkcję π zgodnie z formułą:

$$\theta_{k+1} = \arg\max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi_{\theta_k}}(s_t, a_t),\ \ g(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)) \right)$$

   Zwykle na podstawie algorytmu stochastycznego wzrostu gradientowego

2.5. Dopasuj funkcję wartości przez regresję błędu średniokwadratowego, zgodnie z formułą:

$$\phi_{k+1} = \arg\min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t \right)^2$$

Zwykle na podstawie algorytmu spadku gradientowego

Źr: opr. wł. na podst. Internet[49].

Schemat działania PPO (Proximal Policy Optimization) jest bardzo podobny do A2C (Advantage Actor Critic). Można więc ograniczyć się do omawiania jedynie występującą między nimi różnicę. W przypadku A2C, aby aktualizować wartości zwracane przez Aktora, liczone są pochodne logarytmów prawdopodobieństw wykonania danych akcji a w danych stanach s. PPO zakłada natomiast ewaluację pochodnej w jednym z dwóch wariantów.

- Ratio – stosunku prawdopodobieństwa wykonania akcji a teraz do prawdopodobieństwa jej wykonania w przeszłości.
- Ograniczonego Ratio – stosunku definiowanego tak samo jak powyższy, jednak przyjmującego wartości jedynie z przedziału $[1 - \varepsilon, 1 + \varepsilon]$, gdzie: $\varepsilon \in (0, \infty)$.

Tak określone pochodne mnoży się przez wartości Advantage.

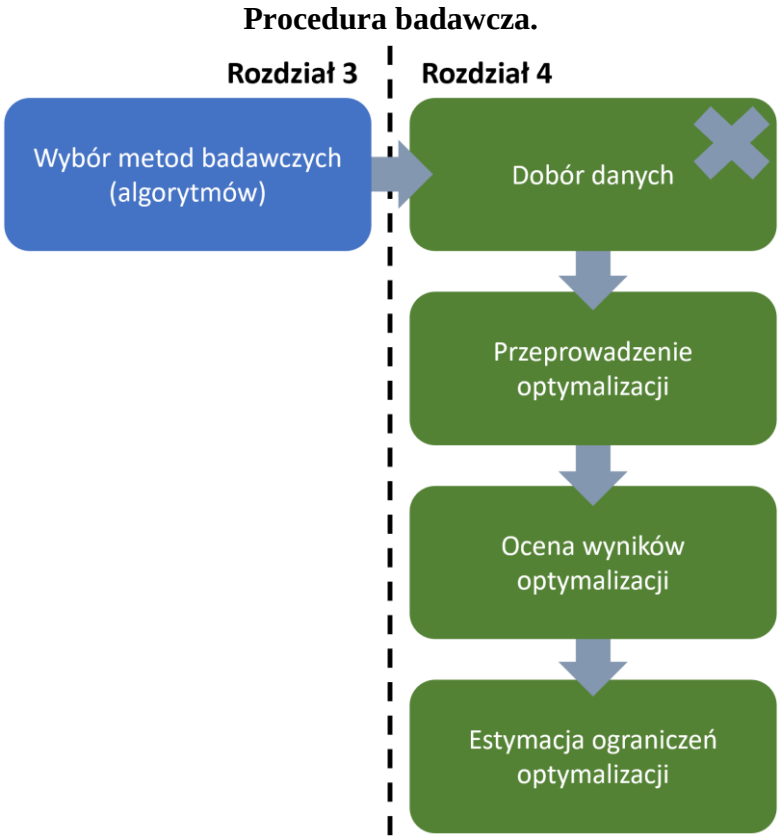
[49] Open AI Spinning Up (https://spinningup.openai.com/en/latest/algorithms/ppo.html; dostęp: 25.05.22r.).

4 Optymalizacja portfela akcyjnego

W rozdziale przedstawiono realizowane w pracy badanie mające na celu optymalizację struktury zarządzanego aktywnie portfela akcyjnego. W szczególności, na przestrzeni Podrozdziału 4.1. przedstawiono pierwotny zbiór danych, schemat selekcji decydujący o kolejnych przekształceniach oraz jego ostateczną formę. Podrozdział 4.2. dotyczy przeniesienia wcześniej przedstawionego schematu funkcjonowania wybranych algorytmów uczenia ze wzmacnianiem do środowiska rozważanego problemu. W Podrozdziale 4.3. przedstawiono wyniki badania oraz płynące z nich wnioski wraz z estymacją ograniczeń.

4.1 Schemat procedury badawczej i dobór danych

Opracowano szablon postępowania mający na celu optymalizację portfela akcyjnego przy użyciu algorytmów uczenia ze wzmacnianiem. Można było wyrazić go w ujęciu blokowym.



Źródło: opracowanie własne.

Rysunek 4.1. *Na niebiesko zaznaczono dokonany już w Rozdziale 3. wybór metod badawczych. W Rozdziale 4 zaplanowano zaznaczone na zielono, 4 kolejne kroki: dobór danych, przeprowadzenie optymalizacji, ocenę jej wyników oraz oszacowanie ograniczeń badania. Pierwszy z nich dotyczy tego podrozdziału, co zaznaczono krzyżykiem.*

Przedstawiony na Rysunku 4.1. schemat procedury badawczej złożono z 5, następujących po sobie, działań. Oznaczonego jako pierwszy element ciągu, wyboru metod badawczych dokonano na drodze przedstawionej w poprzednim rozdziale analizy prac naukowych. Zdecydowano się wtedy na użycie 2 algorytmów uczenia ze wzmacnianiem: A2C i PPO. W tym podrozdziale przedstawiono dobór danych. Na przestrzeni kolejnego opisano schemat przeprowadzania optymalizacji portfela akcyjnego definiowany zastosowaniem wybranych algorytmów do utworzonego zbioru danych. Ostatni podrozdział dotyczy wyników ich oceny. Zawarto w nim także szacowne ograniczenia realizowanej optymalizacji wynikające z charakterystyki zbioru danych, metod i wykorzystywanego paradygmatu analizy finansowej.

Chcąc skonstruować adekwatny do badania zbiór danych rozważano w pierwszej kolejności podmioty jakich powinien on dotyczyć. Aby imitować jak najlepiej szerokie spektrum gospodarki, wybrano 400 największych na świecie spółek giełdowych według kapitalizacji rynkowej w dolarach amerykańskich z dnia 13.07.22r.[50] Określony w ten sposób zbiór pokrywał dużą część światowego przekroju sektorów działalności operacyjnej oraz regionów geograficznych. Ostatecznie planowano wzięcie danych dotyczących historycznych notowań dziennych cen akcji z okresu czasowego znajdującego się w przedziale od 5 lat do 10 lat. Dolne ograniczenie wynikało z przeglądu literatury, a górne z estymowanych możliwości obliczeniowych dla stawianego sobie problemu badawczego rozpatrywanego względem grupy około 35 podmiotów. Do tak ograniczonego zbioru możliwie jak najbardziej reprezentatywnej informacyjnie grupy spółek planowano dojść na drodze selekcji ze względu na kategorie działalności operacyjnej, dostępność danych oraz ich korelację.

[50] Do wyboru 400 największych na świecie spółek giełdowych posłużono się aktualizowanymi codziennie danymi, por. Companies Market Cap (companiesmarketcap.com; dostęp: 13.07.22r.)

W pierwszej kolejności spółki podzielono względem rodzaju prowadzonej działalności operacyjnej na 44 kategorie.

Wyróżnione kategorie spółek.

Zarządzanie kapitałem[51]	Z.K. z nastawieniem na rynek nieruchomości	Budownictwo	Chemia i nawozy
Części samochodowe i baterie litowe	Energetyka i infrastruktura energetyczna	FMCG	Gospodarowanie odpadami
Hotelarstwo i turystyka	HR	Infrastruktura kolejowa	Konglomerat
Konsulting operacyjny, zarządczy i podobne [52]	Logistyka	Media i treści rozrywkowe	Motoryzacja
Obsługa rynków i informacji finansowych	Obsługa transakcji finansowych	Odzież, obuwie i okulary	Opieka zdrowotna
Oprogramowanie dla firm i osób fizycznych [53]	Oprogramowanie konsumenckie [54]	Podzespoły elektroniczne	Półprzewodniki
Produkcja gazów przemysłowych	Produkcja i sprzedaż artykułów luksusowych	Produkcja maszyn przemysłowych i podob. [55]	Produkcja sprzętu AGD
Produkty elektroniczne	Przemysł farmaceutyczny i podobne [56]	Przemysł lotniczy [57]	Przemysł zbrojeniowy [58]
Sprzedaż artykułów spożywczych i domowych	Marketplace	Technologia medyczna i podobne [59]	Telekom i infrastruktura teleinformacyjna
Usługi bankowe dla firm i osób fizycznych	U.B.F.O.F. kontro. przez rząd Arabii Saudyjskiej	U. B. F. O. F. kontrolowane przez rząd Chin	U. B. F. O. F. kontrolowane przez rząd Indii
U. B. F. O. F. kontrolowane przez rząd Indonezji	U. B. F. O. F. kontrolowane przez rząd Rosji	Wydobycie metali i metalurgia	Wydobycie węglowo. i przetwórstwo petrochem.

Źródło: opracowanie własne.

Tabela 4.1. *Kategorie działalności operacyjnej spisano w kolejności alfabetycznej względem wierszy tabeli.*

Wszystkie wyróżnione rodzaje podsiębiorstw wymieniono w Tabeli 4.1. Utworzenie takich kategorii pozwoliło na zgromadzenie spółek w klastrach o zbliżonych dynamikach zmian cen

[51] Przedsiębiorstwo zarządzające kapitałem jest tłumaczeniem własnym angielskiego terminu *asset manager*.
[52] Konsulting operacyjny, zarządczy i analityka biznesowa.
[53] Oprogramowanie dla firm i osób fizycznych rozumiane jako służące modyfikacja plików, zarządzaniu pracą i bazami danych, a także consulting IT, cyberbezpieczeństwo.
[54] Oprogramowanie konsumenckie rozumiane jako gry komputerowe, aplikacje internetowe oraz sieci społecznościowe.
[55] Produkcja maszyn przemysłowych, rolniczych i budowlanych.
[56] Przemysł farmaceutyczny, biotechnologia i produkty medyczne jednorazowego użytku.
[57] Przemysł głównie lotniczy i w mniejszym stopniu mniej zbrojeniowy.
[58] Przemysł głównie zbrojeniowy i w mniejszym stopniu lotniczy.
[59] Technologia medyczna i produkcja sprzętu medycznego.

akcji względem czasu. Liczyły one od 1 do 38 podmiotów (w przypadku Usług bankowych dla firm i osób fizycznych). Konglomeraty zostały na tym etapie selekcji odrzucone z badania. Działalność operacyjna tych spółek nie ograniczała się w znacznej większości do jedynie jednej z klas. Co za tym idzie, ceny ich akcji nie spełniały przesłanek do przewidywania znaczących różnic dynamiki względem wszystkich kategorii poza jedną. W ramach każdej z 44 kategorii wybrano jej reprezentację do badania.

Schemat wyboru spółek.

Liczba zgromadzonych w kategorii spółek	Liczba spółek-reprezentantów
[21, 38] →	4
[11, 20] →	3
[6, 10] →	2
[3, 5] →	1
[0, 2] →	0

Źródło: opracowanie własne.

Tabela 4.2. *Zgodnie z przypuszczeniami, wśród 4 typów kategorii posiadających reprezentantów najczęściej występującym był ten liczący od 3 do 5 spółek, co zaznaczono błękitną elipsą.*

Tabela 4.2. opisuje schemat według którego liczebność grupy spółek zgromadzonych w kategoriach działalności operacyjnej determinowała liczebność grupy jej spółek-reprezentantów. Do każdego z zawężonych w ten sposób zbiorów następne podmioty dobierano w kolejności zgodnej z kapitalizacją rynkową. Wymogiem dotyczącym każdej z utworzonych reprezentacji kategorii było również względne rozproszenie geograficzne. W ramach każdej z nich dopuszczano maksymalnie dwie spółki z siedzibami w tym samym kraju. Aby nie wypaczać wyników selekcji posługiwano się główną lokalizacją działalności operacyjnej, co nie zawsze było jednoznaczne z oficjalnym miejscem rejestracji spółki. W ten sposób uniknięto nadreprezentacji krajów ponadprzeciętnie atrakcyjnych podatkowo. Grupę pierwotnie wybranych 400 podmiotów ograniczono do 71.

W dalszym etapie rozważań analizowano długości dostępnych zakresów czasowych danych. Już na tym etapie w grupie nie było przedsiębiorstw notowanych na giełdzie moskiewskiej. W przeciwnym wypadku konieczne byłoby uwzględnienie ograniczenia danych występującego od dnia 28 lutego 2022r. Porównano daty debiutów giełdowych wszystkich 71 spółek. Za dolną granicę przedziału czasowego wybrano na tej podstawie dzień 26.09.2014r. Wyznaczał on tydzień opóźnienia względem IPO chińskiego Alibaby. Decyzja ta wiązała się z wymianą dwóch innych spółek na najsilniej skorelowane w ramach kategorii alternatywy.

Użyto klasycznego współczynnika korelacji Pearsona liczonego dla cen z dni równoległych notowań obu spółek od momentów IPO podmiotów odrzucanych do dnia 15.07.2022. Tym sposobem wśród reprezentantów wydobycia węglowodorowego i przetwórstwa petrochemicznego miejsce Saudi Aramco (debiut dnia 11.12.2019) zajęło China Shenhua Energy (debiut dnia 09.10.2007). W kategorii części samochodowych i baterii litowych CATL zastąpiono O'Reilly Automotive. Górne ograniczenie wszystkich analizowanych w pracy okresów oparto o datę wybraną maksymalnie względem samego badania. Łącznie rozważano więc okresy danych dziennych ograniczone w ramach długości blisko 7 lat.

Przy dalszej selekcji spółek giełdowych używano danych dotyczących dziennych cen akcji. Posługiwano się przy tym wartościami „adjusted close". W przeciwieństwie do zwykłych cen z momentu zamknięcia rynku, jest to miara uwzględniająca również ewentualne wydarzenia takie jak split akcji, por. Internet[60]. Z tego względu można uznawać ją za bardziej optymalną dla celów podejmowanego badania.

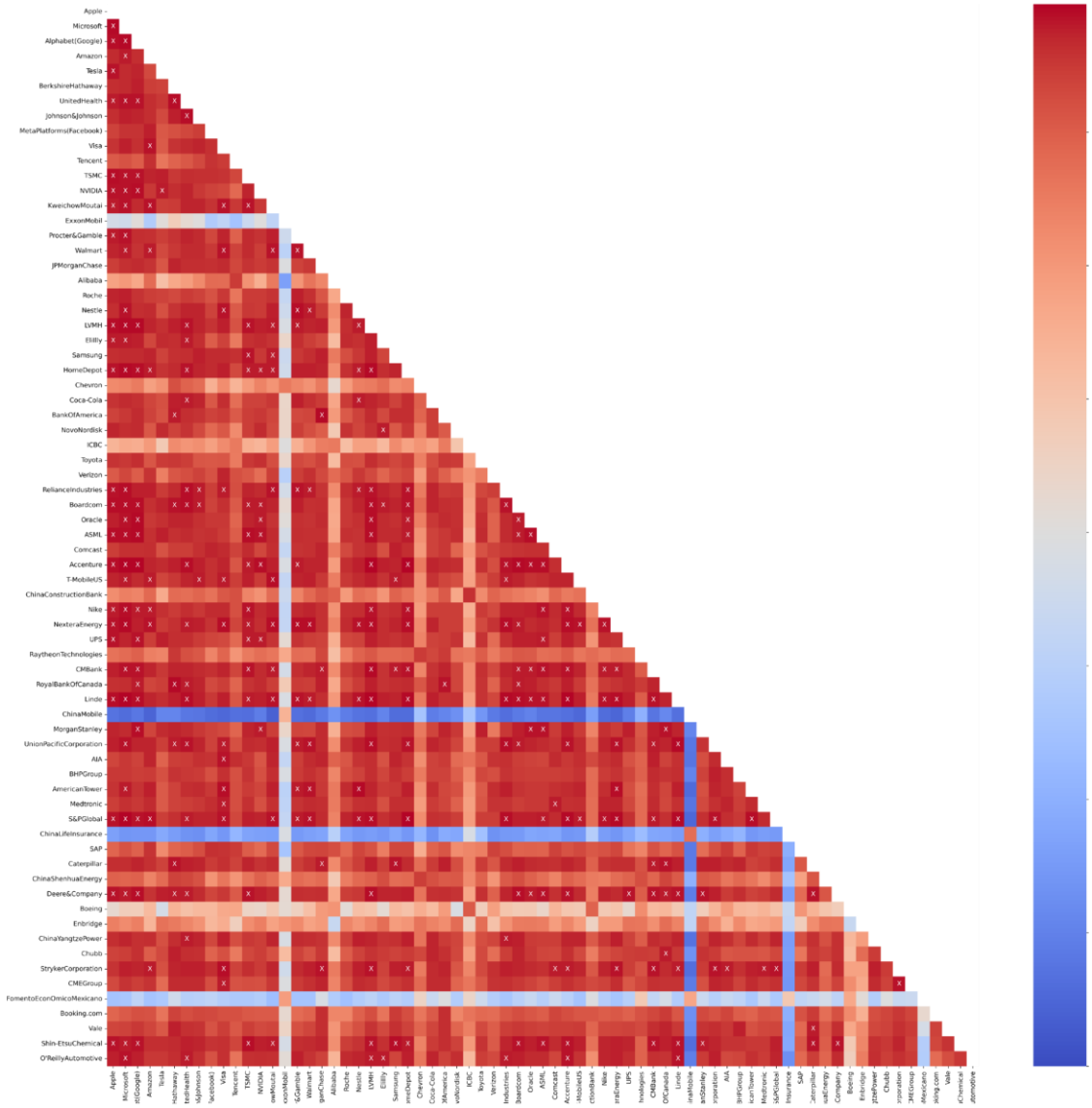
Wszystkie dane wykorzystane w pracy pochodziły z serwisu Yahoo Finance. Jest to wykorzystywane w wielu pracach naukowych rzetelne i publicznie dostępne repozytorium informacji dotyczących historycznych notowań giełdowych.

Utworzono 71 szeregów czasowych dotyczących notowań historycznych cen za dni będące przecięciem zbioru wszystkich handlowych w ramach rozpatrywanych giełd. Różne kraje mają inaczej względem siebie rozłożone dni wolne i należało to uwzględnić chcąc zachować spójność czasową danych. Szeregi czasowe cen akcji par spółek poddane takiej samej modyfikacji zostały określone wcześniej cenami z dni równoległych notowań. Będą tak nazywane również w dalszej części tekstu.

Dla opracowanych szeregów czasowych stworzono macierz korelacji. Ponownie wykorzystano przy tym klasyczny współczynnik korelacji Pearsona.

[60] Yahoo (https://help.yahoo.com/kb/SLN28256.html; dostęp: 28.05.22r.).

Macierz korelacji.



Źródło: opracowanie własne.

Rysunek 4.2. *W wymiarze pionowym i poziomym odłożony jest ten sam zbiór 71 spółek. Macierz prezentuje poziomy korelacji odpowiadające każdej z ich par. Cieplejsze kolory świadczą o większych wartościach. Pary o współczynniku korelacji przekraczającym poziom 0,95 zaznaczono krzyżykami.*

Rysunek 4.2. przedstawia macierz korelacji liczoną dla szeregów czasowych cen historycznych z dni równoległych notowań akcji 71 spółek. Po opracowaniu takiego zestawienia, w ramach każdej z par o współczynniku korelacji przekraczającym poziom 0,95 odrzucono spółki o niższej kapitalizacji giełdowej. Grupę rozważanych podmiotów ograniczono tym sposobem do 36 i opracowano dla niej ponownie szeregi czasowe cen z dni równoległych notowań.

Spółki w ostatecznym zbiorze danych.

AIA	Alibaba	Amazon
Apple	BerkshireHathaway	BHPGroup
Boeing	Booking.com	Chevron
ChinaConstructionBank	ChinaLifeInsurance	ChinaMobile
ChinaShenhuaEnergy	ChinaYangtzePower	Chubb
CMEGroup	Comcast	Enbridge
ExxonMobil	FomentoEconOmicoMexicano	ICBC
Johnson&Johnson	JPMorganChase	MetaPlatforms
Nestle	NovoNordisk	Oracle
O'ReillyAutomotive	RaytheonTechnologies	Roche
Samsung	SAP	Tencent
Toyota	Vale	Verizon

Źródło: opracowanie własne.

Tabela 4.3. *Spółki wymieniono w kolejności alfabetycznej względem wierszy tabeli.*

Otrzymano zbiór danych dotyczących spółek przedstawionych w Tabeli 4.3. Budowały go szeregi czasowe o długości równej 1719 dniom handlowym rozłożonym na przestrzeni ograniczonej datami: 26.09.2014, 14.07.2022.

Do automatyzacji poboru danych oraz wyboru przecięć zbiorów rekordów posługiwano się programem Microsoft Excel. Współczynniki korelacji liczono wykorzystując bibliotekę Pandas Pythona w środowisku Jupyter Notebook.

4.2 Algorytmy A2C i PPO w środowisku problemu badawczego

Aby implementować algorytmy w środowisku rozpatrywanego problemu badawczego stworzono 2 programy. Procesie modyfikacji wykorzystywanych sieci neuronowych jeden z nich używał A2C, a drugi PPO. Opis działania obu programów podzielić można na cztery podpunkty.

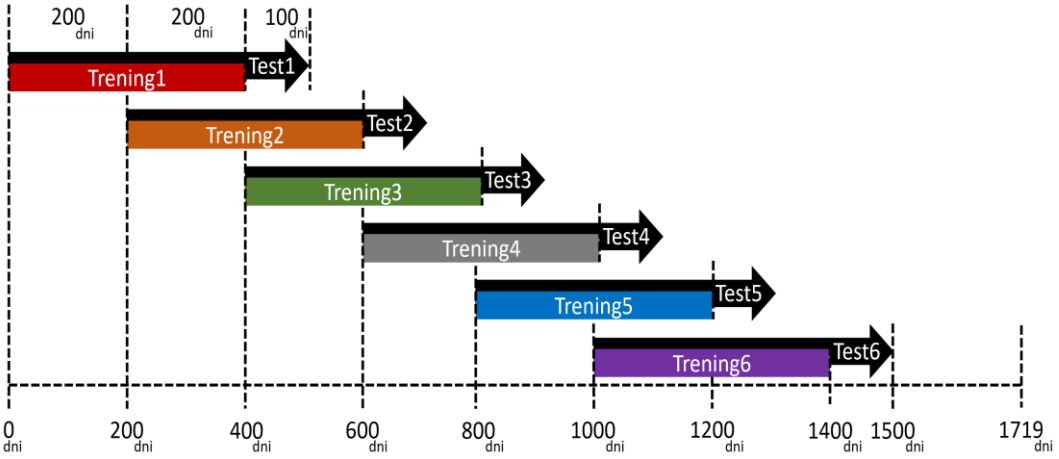
1. Wyrażany w terminach środowiska opis działania algorytmów.
2. Podział wykorzystywanego zbioru danych względem realizowanego procesu treningowo-testowego.
3. Budowa programów w uproszczonym ujęciu.
4. Kod programistyczny.

W odniesieniu do podpunktu pierwszego, zarządzany przez każdy z algorytmów trening przebiegał zgodnie ze schematem przedstawionego w poprzednim rozdziale pseudokodu. Aktor, na podstawie obserwacji Środowiska mapującej jego stan bieżący, określał wybieraną

przez siebie akcję. Akcja równoznaczna była z nowym kształtem portfela akcyjnego. Po jej wyrażeniu otrzymywał informacje zwrotne dotyczące nowego stanu Środowiska i otrzymywanej nagrody. Obserwacją były ceny z ostatnich 20 dni notowań normalizowane osobno dla każdego z walorów finansowych. Akcję wyrażał wektor zmiennych związanych warunkiem określającym sumę ich wartości bezwzględnych jako stale równą 1. Dodatkowo, każda z nich, brana osobno, musiała być większa niż –1, a przy tym mniejsza niż 1 (wyłącznie). Wartość bezwzględna danej zmiennej reprezentowała ułamek portfela mający zostać poświęcony inwestycji w dany instrument finansowy. Znak zmiennej określał stronę zajmowanej pozycji handlowej. Ujemne wartości świadczyły o sprzedaży / zajęciu krótkiej pozycji. Dodatnie natomiast wyrażały zakup / zajecie pozycji długiej. Jako nagrodę zdefiniowano stosunek ewaluacji portfela inwestycyjnego na dzień po jego modyfikacji do ewaluacji w momencie zmiany.

Chcąc wykorzystywać poprawnie algorytmy A2C i PPO rozplanowano adekwatny do zbioru danych proces treningowo-testowy.

Proces treningowo-testowy.



Źródło: opracowanie własne.

Rysunek 4.3. *Wymiar poziomy równoznaczny jest z mierzoną dniami osią czasu. Każda ze strzałek reprezentuje pojedyncze okno czasowe. Dalsza od grotu część, w kolorze innym niż czarny, oznacza trening. Część zawierająca grot to proces testowy.*

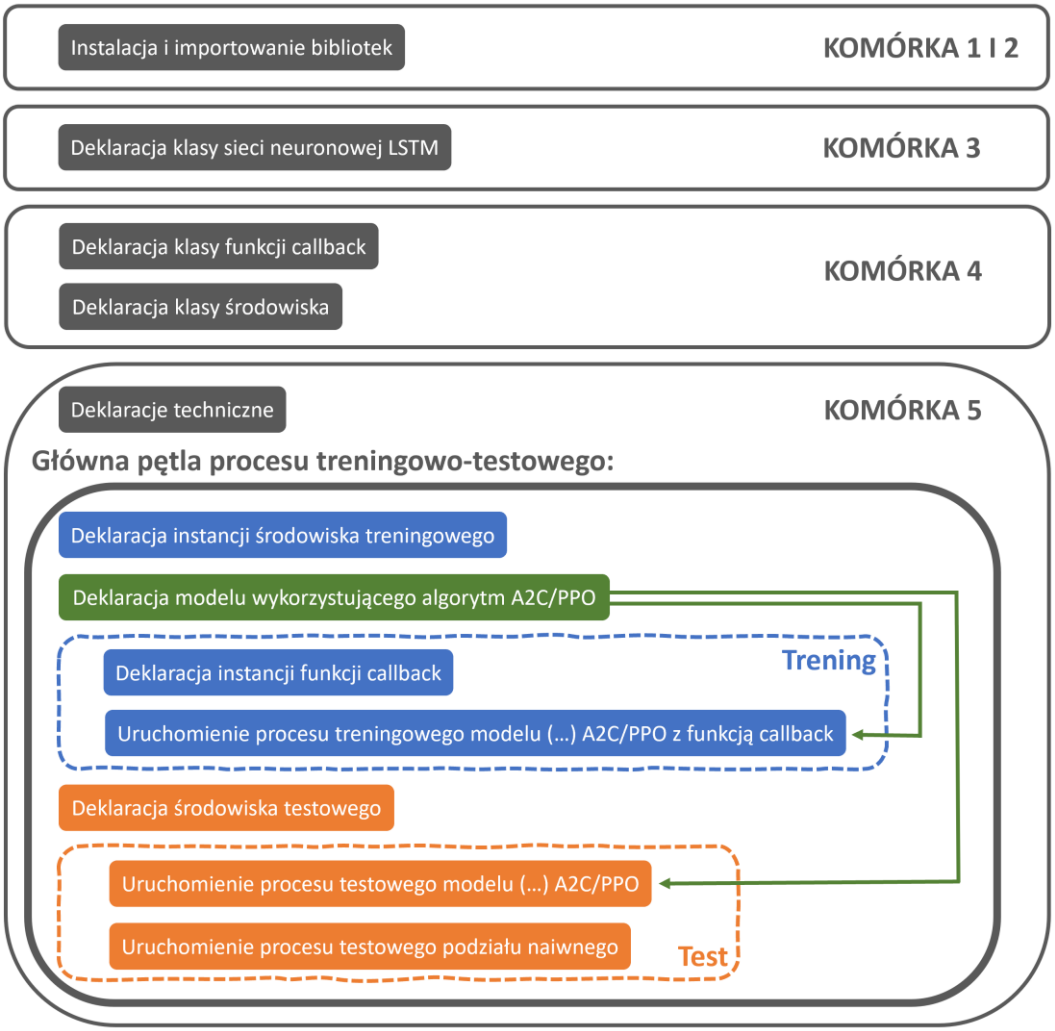
Rysunek 4.3. przedstawia schemat kompilowanego na przestrzeni 1500 dni procesu treningowo-testowego. Realizowano go względem 36 rozpatrywanych szeregów notowań cen akcji. Każdy z nich podzielono na 6 okien czasowych liczących po 500 rekordów danych. Każde dwa sąsiadujące ze sobą okna były przesunięte względem siebie o 200 rekordów. Ich przecięcie było więc równe 300 rekordom. Każde z okien podzielono na dni realizacji, ciągłych

względem rekordów, procesów treningu oraz testu w proporcji 4:1. Pierwszy z nich służył uczeniu programów – powodował więc zgodnie z planem modyfikacje sieci neuronowych. Dane zgromadzone w drugim wykorzystywano jedynie przy ewaluacji skuteczności badanych algorytmów. Praca na nich nie powodowała zmian struktury programu. W ten sposób, spośród dostępnej w zbiorze danych przestrzeni 1719 dni, zagospodarowano pierwszych 1500. Przechodzenie przez cały zbiór treningowo-testowy (a więc i każde z 6 okien czasowych) było powtarzane 30 razy w przypadku każdego z algorytmów.

Przed odniesieniem do samej budowy programów warto wspomnieć charakterystykę wykorzystanych wersji algorytmów i sieci neuronowych. A2C i PPO zostały zaimplementowane w wersjach dla ciągłych przestrzeni akcji. Tego typu warianty różnią się od bardziej tradycyjnych schematów dyskretnych jedynie w dwóch kwestiach. Pierwsza to sposób aktualizacji wartości zwracanych przez Aktora. Druga dotyczy samej ich istoty. Algorytmy w wersjach dla przestrzeni ciągłych nie zwracają prawdopodobieństw akcji, tylko ich konkretne wybory. W opisanym cyklu treningowym, wraz z otrzymywaniem kolejnych informacji ze Środowiska, aktualizowane były sieci neuronowe, reprezentujące zarówno Aktora jak i Krytyka. Bazowały one na architekturze LSTM – popularnym schemacie rekurencyjnym służącym do analizy szeregów czasowych. Jej aktualizacja następowała po każdym dniu treningu.

Kod programów implementujących algorytmy A2C i PPO podzielono na 5 komórek. 2 pierwsze dotyczyły instalowania oraz importowania odpowiednich bibliotek. 2 kolejne poświęcono deklaracjom pomocniczym. Budowa ostatniej z części odzwierciedlała działanie właściwe programów. Można było przedstawić ją w uproszczeniu schematem podzielonym na elementy związane z procesami treningu i testu.

Programy implementujące algorytmy A2C i PPO.



Źródło: opracowanie własne.

Rysunek 4.4. *Na szaro zaznaczono te elementy, których bezpośrednie znaczenie względem opisu działania kodu było niewielkie. Kolorem niebieskim zdefiniowano elementy związane z treningiem, a pomarańczowym te służące testowi. Na zielono zaznaczono jedynie deklarację modelu wykorzystującego algorytm A2C lub PPO. Model ten był podczas działania programu wykorzystywany w obu procesach, co zilustrowano zielonymi strzałkami.*

Rysunek 4.4. przedstawia w uproszczeniu budowę obu programów stosujących algorytmy A2C i PPO do rozwiązywania problemu badawczego. Pierwsze cztery komórki kodu zawierały elementy, których znaczenie dla procesu realizowane było za pośrednictwem tej ostatniej. Z tego powodu opis działania programów najlepiej uchwycić z jej perspektywy. Rozpoczynające komórkę deklaracje techniczne były niezbędne do importowania danych wejściowych opisywanych w Podrozdziale 4.1. oraz eksportowania agregowanych wyników – analizowanych później w Podrozdziale 4.3. Pętlę główną programu realizowano 6 razy, dla każdego ze wcześniej zdefiniowanych okien czasowych. Deklaracja środowiska treningowego

równoznaczna była ze stworzeniem instancji odpowiedniej klasy – zadeklarowanej w komórce 4 przy użyciu dziedziczenia po klasie z importowanej biblioteki. Zbudowany na podstawie klasy importowanej z innej biblioteki model, wykorzystujący algorytm A2C lub PPO, definiowany był na poziomie deklaracji 4 głównymi elementami, które wymieniono poniżej.

1. Środowisko.

- Pozostałe 3 argumenty funkcji pozwalające Aktorowi działać w Środowisku.

2. Sieć neuronowa typu LSTM. Był to obiekt klasy zadeklarowanej w komórce 3 przy użyciu dziedziczenia po klasie z jednej z importowanych bibliotek.
3. Stopa Learning rate. Była to stała wartość liczbowa. Definiowała relację zachodzącą między działaniami eksploracyjnymi, a eksploatacyjnymi. Została wybrana na podstawie metody prób i błędów.
4. Stopa dyskonta przyszłych nagród. Była to stała wartość liczbowa. Sposób rozwiązywania problemu badawczego zdecydował o ustawieniu jej na 0. Posługując się prostym tłumaczeniem terminów anglojęzycznych, można by określić tak wyspecyfikowany algorytm jako maksymalnie „chciwy”.

Deklarację instancji funkcji callback wykonano analogicznie do środowiska i sieci neuronowej. Wykorzystywana w tym celu klasa została zadeklarowana w komórce 4 przy użyciu dziedziczenia po klasie z jednej z importowanych bibliotek. Funkcja callback służy zwracaniu różnych wartości z procesu treningowego, a więc mapowaniu zmian zachodzących w strukturze sieci neuronowej. Z tego względu oznaczone na rynku, uruchomienie procesu treningowego modelu wykorzystującego algorytm A2C, odbywało się przy użyciu tej funkcji. Środowisko testowe zadeklarowano analogicznie do Środowiska treningowego z wykorzystaniem innych danych. Wykorzystywano je do uruchomienia procesu testowego modelu wykorzystującego algorytm A2C oraz procesu testowego podziału naiwnego.

Wszystkie skrypty napisano w Google Colab. Jest to darmowe narzędzie pozwalające na uruchamianie skryptów języka Python. Zapewnia ono dostęp do niezbędnych przy trenowaniu programów, zasobów mocy obliczeniowych procesora (GPU). W badaniach wykorzystano trzy otwarte biblioteki języka Python, które wymieniono poniżej.

- Implementacje algorytmów A2C i PPO pochodziły z biblioteki Stable Baselines 3.
- Sieć neuronowa oraz proces treningu zostały stworzone przy użyciu biblioteki PyTorch.
- Środowisko programów napisano korzystając z biblioteki OpenAI Gym.

Dokładany kod programów implementujących algorytmy A2C i PPO, wykorzystanych na potrzeby badania, zawarto w Załączniku do pracy, por. Rozdział 9.

4.3 Wyniki optymalizacji

Rezultaty osiągane przez programy wykorzystujące algorytmy A2C i PPO zostały zestawione z tymi będącymi następstwem naiwnego podziału portfela – równego względem akcji wszystkich spółek. W tak określonym porównaniu za benchmark przyjęto efekty zwracane przez program używający algorytmu Critical Line Algorithm (CLA). Critical Line Algorithm (CLA) to deterministyczna metoda znajdująca podział portfela, zachowujący balans między wartością oczekiwaną zysku z inwestycji, a jego wariancją. Stanowi ona aktualny i popularny w pracach naukowych benchmark reprezentujący możliwości tradycyjnych narzędzi statystycznych w przypadku ewaluacji skuteczności rozwiązań z dziedziny uczenia maszynowego (ang. *state-of-the-art*). Dane wejściowe dla algorytmu to wartości oczekiwane zysku wynikającego z zakupu poszczególnych instrumentów finansowych oraz macierz kowariancji z ich cen historycznych. W badaniach wykorzystano implementację CLA z biblioteki PyPortfolioOpt, por. Hoogenband (2017). Dokładany kod programu implementującego algorytm CLA, wykorzystanego na potrzeby badania, zawarto w Załączniku do pracy, por. Rozdział 9.

Wyniki programów wykorzystujących poszczególne algorytmy.

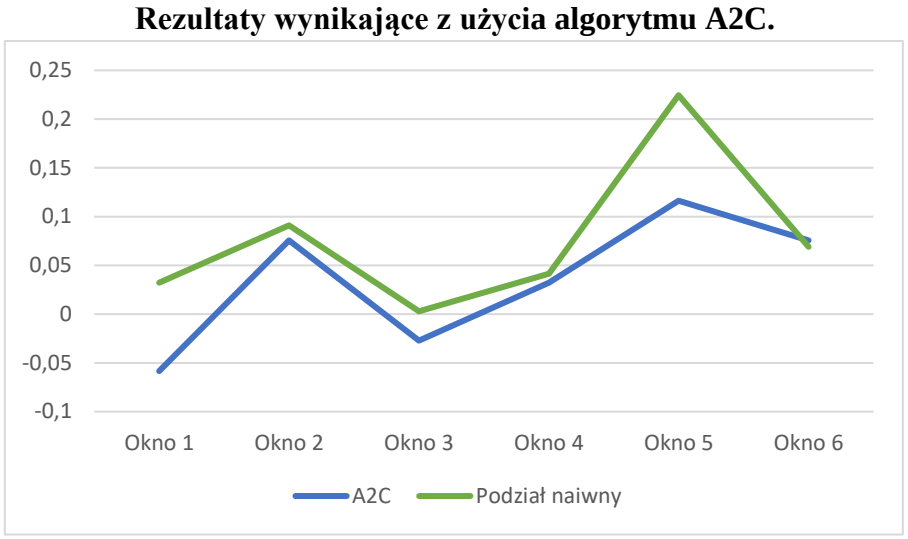
	Okno 1	Okno 2	Okno 3	Okno 4	Okno 5	Okno 6
Podział naiwny	$3{,}21 * 10^{-2}$	$9{,}09 * 10^{-2}$	$2{,}9 * 10^{-3}$	$4{,}16 * 10^{-2}$	$2{,}245 * 10^{-1}$	$6{,}91 * 10^{-2}$
A2C	$-5{,}84 * 10^{-2}$	$7{,}56 * 10^{-2}$	$-2{,}73 * 10^{-2}$	$3{,}24 * 10^{-2}$	$1{,}163 * 10^{-1}$	$7{,}56 * 10^{-2}$
PPO	$-4{,}18 * 10^{-2}$	$1{,}109 * 10^{-1}$	$-2{,}99 * 10^{-2}$	$1{,}82 * 10^{-2}$	$1{,}517 * 10^{-1}$	$6{,}36 * 10^{-2}$
CLA	$8{,}13 * 10^{-2}$	$1{,}414 * 10^{-1}$	$2{,}14 * 10^{-2}$	$4{,}45 * 10^{-2}$	$1{,}604 * 10^{-1}$	$4{,}80 * 10^{-2}$

Źródło: opracowanie własne.

Tabela 4.4. *Wszystkie liczby zostały zaokrąglone do części dziesięciotysięcznych.*

Tabela 4.4. przedstawia wyniki badania – osiągane przez programy implementujące algorytmy i mierzone zwrotami z portfeli inwestycyjnych. Ułamki w komórkach tabeli reprezentują wartości o jakie zainwestowana pierwotnie suma zostałaby zwiększona (dla wartości dodatnich) lub zmniejszona (dla wartości ujemnych) w przypadku realizowania strategii wyznaczanej przez każde z podejść (wymienionych w kolumnie pierwszej),

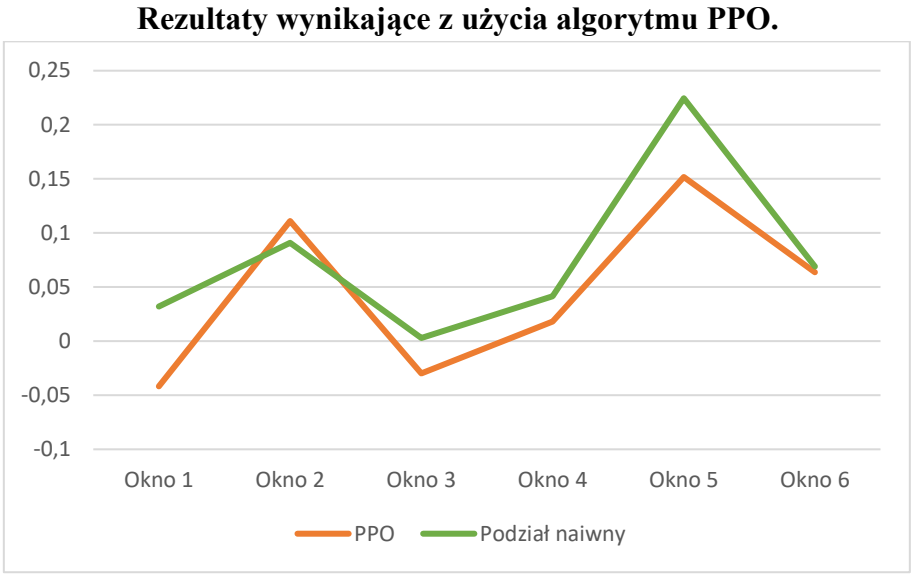
na przestrzeni każdego z 6 okien czasowych (wymienionych w wierszu pierwszym). Portfele związane z oboma algorytmami okazały się być w ogólności gorsze od naiwnego podziału po równo.



Źródło: opracowanie własne.

Rysunek 4.5. *Na osi pionowej odłożono wyrażone ułamkami stopy zwrotu z portfeli inwestycyjnych. Na osi poziomej zaznaczono kolejnych 6 okien czasowych.*

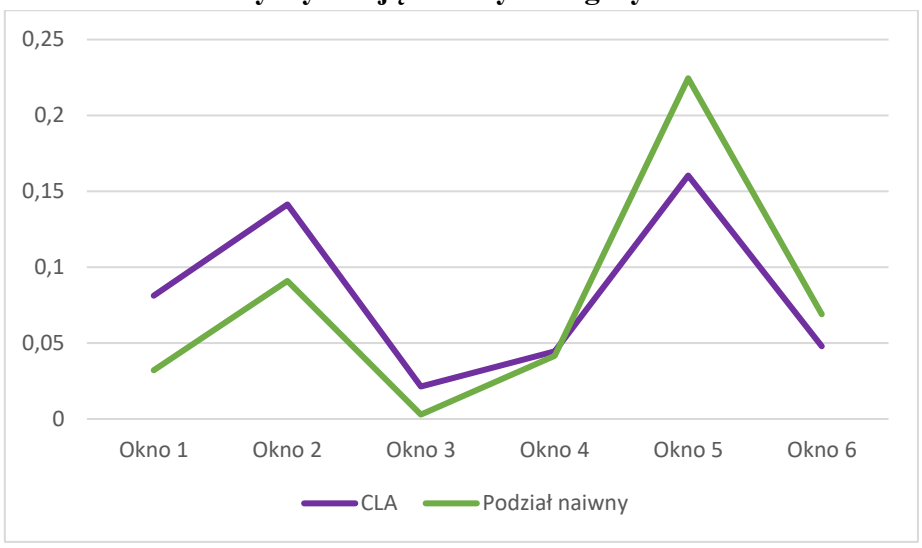
Rysunek 4.5. przedstawia wyniki działania programu wykorzystującego algorytm A2C w zestawieniu z efektami wprowadzenia równego podziału portfela. Jak widać na wykresie, zwrócił on rezultaty lepsze od tych otrzymywanych przy zastosowaniu strategii naiwnej jedynie w przypadku ostatniego okna czasowego. W każdym innym wypadał gorzej. Wykorzystanie PPO nie wypadło lepiej.



Źródło: opracowanie własne.

Rysunek 4.6. *Na osi pionowej odłożono stopy zwrotu z portfeli akcyjnych, a na poziomej okna czasowe.*

Wyżej przedstawiony Rysunek 4.6. pokazuje osiągnięcia programu stosującego algorytm PPO. Jak w przypadku A2C, je też porównano z efektami stosowania równego podziału. Wykorzystanie PPO przełożyło się na wyniki lepsze od tych będących następstwem realizacji strategii naiwnej tylko na przestrzeni drugiego okna czasowego. W pozostałych 5 porównaniach okazywało się być gorszym rozwiązaniem. Nieco lepiej na wykorzystywanym zbiorze danych zadziałał natomiast program realizujący założenia CLA.

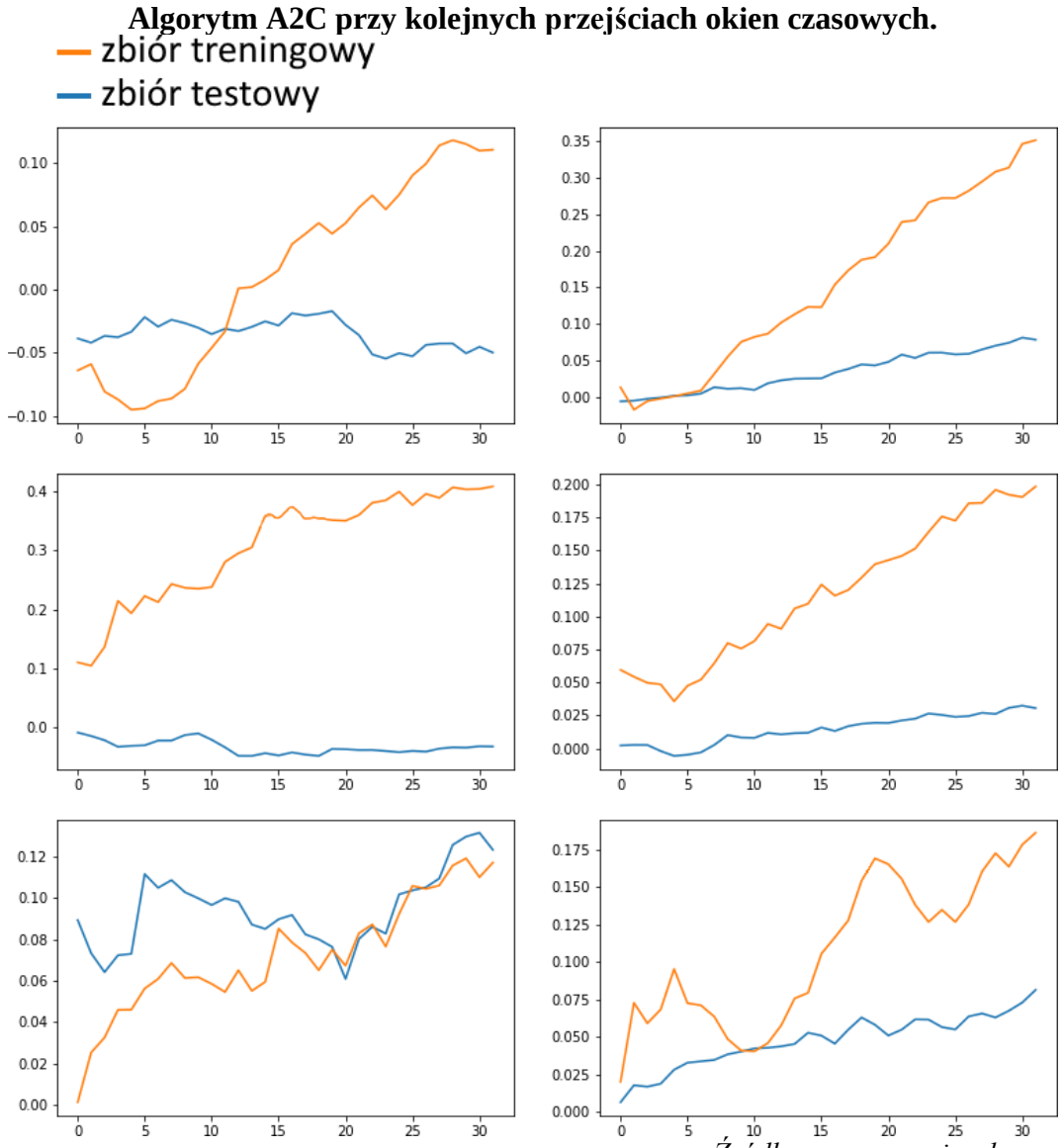


Źródło: opracowanie własne.

Rysunek 4.7. *Ponownie, na osi pionowej odłożono stopy zwrotu z portfeli. Na poziomej przedstawiono 6 następujących po sobie okien czasowych.*

Na Rysunku 4.7. zilustrowano wyniki programu wykorzystującego deterministyczny algorytm CLA oceniane względem efektów stosowania podziału naiwnego. Okazuje się, że wybrany dla potrzeb badania benchmark wypadł lepiej niż oba rozwiązania z dziedziny uczenia maszynowego. Rezultaty będące efektem działania CLA były lepsze od tych wynikających z przyjęcia strategii równego podziału w 4 spośród 6 przypadków.

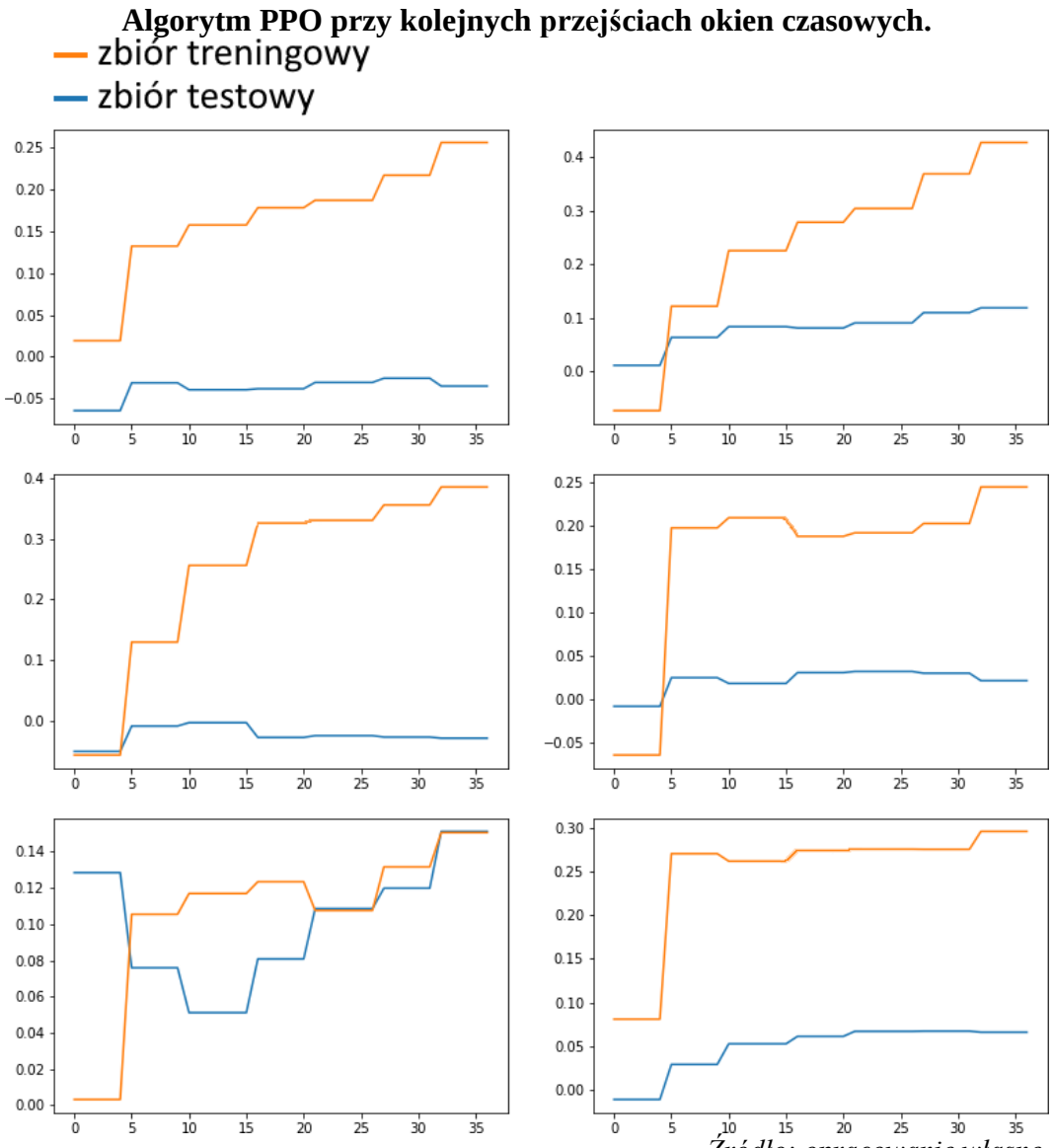
W przypadku zarówno A2C jaki i PPO, dla większości okien czasowych, na przestrzeni kolejnych iteracji uczenia algorytmicznego, zarejestrowano wyraźne postępy. Było tak we wszystkich rozpatrywanych procesach treningowych oraz w połowie procesów testowych.



Źródło: opracowanie własne.

Rysunek 4.8. *Wyniki zaprezentowano na 6 wykresach, z których każdy dotyczy osobnego okna czasowego. Na wszystkich osiach pionowych odłożono stopy zwrotu z portfeli inwestycyjnych. Na osiach poziomych zaznaczono rosnące liczby iteracji procesu treningowo-testowego.*

Rysunek 4.8. przedstawia wyniki osiągane przez program wykorzystujący algorytm A2C w trakcie treningu i testu, na przestrzeni 6 okien czasowych. Jak widać, dla rosnącej liczby iteracji procesu treningowo-testowego, osiągał on co raz lepsze rezultaty. W przypadku procesów treningowych było tak dla wszystkich rozpatrywanych okien czasowych. Procesy testowe wykazywały wzrost efektywności podejmowanych decyzji w 4 z 6 przypadków. Poprawa rezultatów zwracanych przez program wykorzystujący drugi algorytm uczenia ze wzmacnianiem przebiegała jeszcze korzystniej.



Źródło: opracowanie własne.

Rysunek 4.9. *Każdy z 6 wykresów prezentuje wyniki zwracane przez program na przestrzeni jednego z 6 okien czasowych. Osi pionowe odmierzają stopy zwrotu z portfeli akcyjnych. Osi poziome służą odliczaniu iteracji procesu treningowo-testowego.*

Przedstawione na rysunku 4.9. wykresy pokazują, że algorytm PPO determinował osiąganie co raz lepszych wyników przy rosnącej liczbie iteracji procesu treningowo-testowego. Zarówno w przypadku procesów treningowych jak i testowych było tak dla wszystkich 6 okien czasowych.

Oceniając zaprezentowane wyniki badania można stwierdzić na rozpatrywanym zbiorze danych względną nieskuteczność rozpatrywanych metod uczenia ze wzmacnianiem w rozwiązywaniu problemu optymalizacji struktury złożonego z akcji portfela giełdowego. Algorytmy A2C i PPO zwróciły w ogólności gorsze wyniki niż naiwny, równy podział. Każdy z nich wypadł lepiej niż określony w ten sposób punkt odniesienia jedynie na przestrzeni 1 z 6 definiowanych dla potrzeb badania okien czasowych. Narzędzia uczenia ze wzmacnianiem poradziły sobie z postawionym w pracy wyzwaniem gorzej niż reprezentujący tradycyjne metody statystyczne, deterministyczny algorytm CLA. Zwracane przez wykorzystujący go program rezultaty przewyższały te będące następstwem naiwnego podziału portfela inwestycyjnego w 4 spośród 6 okien czasowych. Przebieg treningów wskazywał na poprawne zaimplementowanie algorytmów A2C i PPO w środowisku problemu badawczego. W przypadku obu kolejne iteracje cyklu uczenia powodowały na przestrzeni znacznej większości rozpatrywanych okien czasowych co raz skuteczniejsze działania programów. Było to możliwe do obserwacji zarówno na treningowym, jak i testowym zbiorze danych. Rezultat przeprowadzonego badania określić można ostatecznie jako mniej satysfakcjonujący względem sukcesów osiąganych przez algorytmy, których efekty działania przedstawiano w przytaczanych wcześniej publikacjach naukowych. Taki stan rzeczy może wynikać z trzech powodów.

Po pierwsze, należy zwrócić uwagę na ograniczenia zaangażowanych do potrzeb tej pracy zasobów. Korzystano ze względnie niewielkich mocy obliczeniowych procesora (GPU) działających na stosunkowo niewielkim zbiorze danych. Skutkowało to małą liczbą iteracji procesów treningowych, odbywanych względem krótkich szeregów czasowych dotyczących wąskiego zbioru rozpatrywanych spółek. W połączeniu z losowością związaną z doborem danych probabilistyczny charakter obu algorytmów decyduje o tym, że jednoznaczna ewaluacja ich ogólnych możliwości w tak zorganizowanym badaniu może nie być całkowicie miarodajna. Pewne zestawy danych uwypukliłyby zalety sieci neuronowych uczonych przez A2C lub PPO. Inne je zacierają.

Drugim problemem może być to, że na etapie konstrukcji badania wszystkie wykorzystane w pracy narzędzia były już od przeszło 2 lat opublikowane na warunkach

otwartego dostępu. Oczywistym jest więc, że ich działanie musiało być już dawno zdyskontowane przez rynek. Narzędzia z prac naukowych, których wyniki przytaczano wcześniej były często wykorzystywane na rozpatrywanych przez autorów zbiorach danych po raz pierwszy, ponieważ stanowiły unikalne modyfikacje najnowszych architektur sieci neuronowych, czy algorytmów.

Ostatnią i najważniejszą wątpliwością pozostaje to w jakim stopniu obrany na potrzeby pracy paradygmat analizy technicznej rozumiany wnioskowaniem jedynie na podstawie przeszłych zdarzeń, widocznych na wykresach notowań cen historycznych, pozwala na trwałe osiąganie ponadprzeciętnego zysku. Przytaczana w pierwszym rozdziale historia sukcesu związana z zastosowaniem analizy technicznej wymyka się jej pełnej, tradycyjnej definicji. To, że decydenci funduszu Medalion wykorzystywali zaawansowane matematycznie metody probabilistyczne przy ciągłych badaniach krótkoterminowych, nie stosowali analizy fundamentalnej i polegali w znacznej mierze na wykresach zdarzeń przeszłych nie znaczy, że ostatecznie ograniczali się jedynie do nich.

5 Uwagi końcowe

Programy zbudowane na potrzeby przeprowadzonego badania funkcjonowały poprawnie w zadanym środowisku. Ulepszały wyraźnie wyniki osiągane na przestrzeni kolejnych treningów. Nie działały natomiast na tyle dobrze, aby można było stwierdzić ich praktyczną użyteczność, definiowaną porównaniem względem wybranej, deterministycznej, metody statystycznej. Przedstawione w pracy rezultaty wykazały niską skuteczność wybranych na podstawie przeglądu literatury metod uczenia ze wzmacnianiem przy rozwiązywaniu problemu optymalizacji struktury portfela akcyjnego. W świetle wymienionych estymacji głównych problemów badawczych wynik ten nie jest natomiast zaskakujący. Stworzenie "programu do zarabiania pieniędzy na giełdzie" to niewątpliwie bardzo skomplikowane wyzwanie.

Przy określaniu szacowanych wad badania odniesiono się do tego, że realizowano je przy użyciu ograniczonych zasobów. Przekrój ogółu światowej giełdy akcyjnej mapowano walorami dotyczącymi 36 spółek. Wykorzystywano szeregi czasowe ograniczone w ramach przedziału blisko 8 lat. Wielkość zbioru notowań cen historycznych zdeterminowano w oparciu o dwa czynniki. Pierwszym były ograniczenia dostępnych mocy obliczeniowych. Drugi stanowiły przywoływane przykłady badań finalizowanych satysfakcjonującymi wynikami. Bazowały one na węższych zestawach danych. Należy podkreślić, że zaangażowane do potrzeb tej pracy zasoby nie pozwalały na efektywne zastosowanie algorytmów na znacząco większych zbiorach notowań cen akcji. Przeprowadzona w Rozdziale 3. analiza literatury nie stwarzała jednak podstaw dla przypuszczeń o ich niedostatecznej objętości. Wątpliwości dotyczące badania związane z ograniczeniami zasobów danych i mocy obliczeniowych można więc ograniczyć do liczby iteracji procesu treningowego. Dla każdego algorytmu realizowano go 30 razy na przestrzeni każdego z 6 okien czasowych. W skali eksperymentów dokumentowanych przytaczanymi w Rozdziale 3. artykułami naukowymi są to niewielkie liczby. Wydaje się natomiast, że zwiększanie ich przy użyciu tych samych metod i na tym samym zbiorze danych, może mieć wątpliwy sens praktyczny. Takie podejście ocenić można jako wątpliwe pod względem wartości naukowej. Zestawione z rynkiem oznaczałoby podjęcie konkurencji na polu wykorzystywanych mocy obliczeniowych. Przez możliwości największych podmiotów przetwarzających informacje, można je więc uznawać za nieosiągalne z perspektywy pojedynczego inwestora.

Wśród estymowanych wad badania wymieniono odtwórczość wszystkich użytych w nim metod. Konstrukcja obu stworzonych programów opierała się na wykorzystaniu trzech

otwartych bibliotek (open source) języka Python: Stable Baselines3, PyTorch oraz OpenAI Gym. Z założenia podstawowej efektywności rynku finansowego wynika, że musi on w jakimś stopniu dyskontować działanie tych narzędzi. Eliminuje w ten sposób stwarzane przez nie możliwości arbitrażowe. Przytaczane w Rozdziale 3. przykłady prac badawczych, rejestrujących korzystniejsze wyniki, wykorzystywały często rozwiązania autorskie. Stworzenie nowego algorytmu uczenia maszynowego o zastosowaniu giełdowym wiązałaby się z większą wartością naukową niż z podnoszenie liczby iteracji obecnie wykorzystywanych metod na tym samym zbiorze danych. Należy jednak pamiętać, że byłoby to przy obecnym poziomie zaawansowania dziedziny duże wyzwanie. Zastosowanie takiego rozwiązania na rynku finansowym oznaczałoby podjęcie konkurencji na płaszczyźnie kreatywności matematycznej z licznymi zespołami naukowców z całego świata.

Można zauważyć, że wspólną cechą uczenia maszynowego oraz analizy technicznej jest duża wariancja efektów zastosowań względem konkretnych zestawów danych. W przypadku obu tych dziedzin często nie sposób mówić o rozwiązaniach definitywnie działających. Rozważanie problemu leżącego w ich przecięciu może potęgować ten efekt. Z przytaczanych przykładów literackich wiemy natomiast jedynie, że działałby w konkretnych próbach na konkretnym zbiorze danych. Problem ten wyróżniono szacując wady badania. Określono go wątpliwą zasadnością przyjęcia podejścia zgodnego z założeniami analizy technicznej – rozumianej wnioskowaniem na podstawie zdarzeń przeszłych, widocznych na wykresach notowań cen akcji. Odchodząc od takiego paradygmatu, można by starać się rozwiązywać problem optymalizacji portfela akcyjnego przy użyciu niewielkich mocy obliczeniowych i istniejących metod. Modyfikacja dotyczyłaby rodzaju danych. Ciekawym kierunkiem eksploracji tej idei jest szacowanie trendów krótkoterminowych na podstawie przepływu informacji ze źródeł pozagiełdowych. Decyzje o zmianach pozycji handlowych można by podejmować opierając się o wcześniej określone dla całości gospodarki lub wybranego rynku barometry koniunktury. Różne algorytmy uczenia maszynowego pozwalałyby nadawać odpowiednie wagi nie tylko danym ilościowym dotyczącym produkcji lub zużycia materiałów, ale też informacjom pochodzącym ze sprawozdań finansowych spółek i raportów makroekonomicznych. Pomysł mogłoby okazać się wartościowy naukowo oraz perspektywiczny w ujęciu skuteczności mierzonej zwracanymi stopami zwrotu z inwestycji. Wydaje się być natomiast znacznie bardziej złożony niż opisana w tej pracy implementacja algorytmów.

6 Bibliografia

Książki:

Azzopardi P., “Behavioural Technical Analysis: An introduction to behavioural finance and its role intechnical analysis”, Harriman House, (2010).

Berlinski D., “The Advent of the Algorithm”, Harcourt Books, (2000).

Castro J., “Biblioteca española” (1781).

Crevier D., “AI: The Tumultuous Search for Artificial Intelligence”, BasicBooks, (1993).

Crevier D., “AI: The Tumultuous Search for Artificial Intelligence”, BasicBooks, (1993).

Dębski W., “Rynek finansowy i jego mechanizmy. Podstwy teorii i praktyki”, Wydawnictwo Naukowe PWN (2014).

Hodges A., “Alan Turing: The Enigma”, Princeton University Press, (2012).

Kaźmierczak A., “Pieniądz i bank w kapitalizmie”, (1993).

Kirkpatrick C., Dahlquist J., “Technical Analysis: The Complete Resource for Financial Market Technicians”, Financial Times Press, (2006).

Luger G., Stubblefield W., “Artificial Intelligence: Structures and Strategies for Complex Problem Solving”, Benjamin-Cummings Pub Co, (1997).

Marshall A., “Principles of Economics”, (1890).

Marsland S., “Machine Learning. An algorithmic perspective”, Chapman and Hall, (2014).

McCorduck P., “Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence”, A K Peters/CRC Press, (2004).

Milewski R., Kwiatkowski E., “Podstawy Ekonomii”, Wydawnictwo Naukowe PWN, (2000).

Minsky M., Papert S., “Perceptrons: An Introduction to Computational Geometry”, MIT Press, (1969).

Mitchell T., “Machine Learning”, McGraw-Hill Education, (1997).

Mohri M., Rostamizadeh A., Talwalkar A. “Foundations of Machine Learning”, MIT Press, (2012).

Nilsson N., “Artificial Intelligence: A New Synthesis”, Morgan Kaufmann Publishers, (1998).

Roseveare H., “The Financial Revolution, 1660–1760”, Longman, (1991).

Russell S., Norvig P., “Artificial Intelligence: A Modern Approach”, Pearson Education, (2003).

Samuelson P., Nordhaus W., “Economics”, McGraw-Hill, (1980).

Sawka K., “Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow” wyd. 2, Helion, (2020).

Shelly M., “Frankenstein”, (1818).

Singh A., Bhadani R., “Mobile Deep Learning with TensorFlow Lite, ML Kit and Flutter: Build scalable real-world projects to implement end-to-end neural networks on Android and iOS”, Packt Publishing, (2020).

Sipser M., “Introduction to the Theory of Computation”, Cengage Learning, (2013).

Weatherall J., “The Physics of Wall Street: A Brief History of Predicting the Unpredictable”, (2013).

Zuckerman G., “The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution”, Portfolio, (2019).

Artykuły naukowe:

Atje R., Jovanovic B., “Stock markets and development”, European Economic Review, (1993).

Corzo T., Prat M., Vaquero E., “Behavioral Finance in Joseph de la Vega's Confusion de Confusiones”, Journal of Behavioral Finance, (2014).

Crowe C., Meade E., “The Evolution of Central Bank Governance around the World”, Journal of Economic Perspectives, (2007).

Cybenko G., “Approximations by superpositions of sigmoidal functions”, Mathematics of Control, Signals, and Systems, (1989).

Di Persio L., Honchar O., “Recurrent Neural Networks Approach to the Financial Forecast of Google Assets”, International Journal of Mathematics and Computers in Simulation, (2017).

Freund Y., Schapire R., “Large margin classification using the perceptron algorithm”, (1999).

Garnelo M., Shanahan M., “Reconciling deep learning with symbolic artificial intelligence: representing objects and relations”, (2019).

Gelderblom O., Jong A., Jonker J., “The Formative Years of the Modern Corporation: The Dutch East India Company VOC, 1602–1623”, The Journal of Economic History, (2013).

Greenwood J., Jovanovic B. “Financial Development, Growth, and the Distribution of Income”, Journal of Political Economy, (1990).

Hammond P., “The Efficiency Theorems and Market Failure”, Department of Economics, Stanford University, (1997).

Hirchoua B., Ouhbi B., Frikh B., “Deep Reinforcement Learning Based Trading Agents: Risk Curiosity Driven Learning for Financial Rules-Based Policy”, Expert Systems with Applications, (2021).

Hoogenband M., “Markowtiz’ Critical Line Algorithm”, (2017). [online] (https://studenttheses.uu.nl/bitstream/handle/20.500.12932/25398/Final_version_scriptie_Michael.pdf?sequence=2h)

Huang G., Zhou X., Song Q., “Deep Reinforcement Learning for Portfolio Management Based on the Empirical Study of Chinese Stock Market”, (2021).

Irwin S., Park C., “What Do We Know About the Profitability of Technical Analysis?”, Journal of Economic Surveys, (2007).

Kaelbling L., Littman M., Moore A., “Reinforcement Learning: A Survey”. Journal of Artificial Intelligence Research, (1996).

King R., Levine R., “Finance and Growth: Schumpeter Might be Right”, The Quarterly Journal of Economic, (1993).

Korajczyk R., “A Measure of Stock Market Integration for Developed and Emerging Markets”, World Bank Economic Review, (1996).

Legg S., Hutter M., “A Collection of Definitions of Intelligence”, (2007).

Levine R., Zervos S., “Stock Markets, Banks, and Economic Growth”, American Economic Review, (1998).

Li Y., Zheng W., Zheng Z., “Deep Robust Reinforcement Learning for Practical Algorithmic Trading”, (2019).

Lighthill J., “Artificial Intelligence: A General Survey”, Science Research Council, (1973).

Liu X., Yang H., Chen Q., Zhang R., Yang L., Xiao B., Dan Wang C., “FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance”, (2020).

Lo A., Mamaysky H., Wang J., “Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation”, Journal of Finance, (2000).

Lucas R., “Macroeconomic Priorities”, The American Economic Review, (2003).

Manyika J., “Getting AI Right: Introductory Notes on AI & Society”, (2022).

Mckinnon R., “Money and Capital in Economic Development”, The Brookings Institution, (1973).

Mizrach B., Weerts S., “Highs and Lows: A Behavioral and Technical Analysis”, (2007).

Olser K., “Support for Resistance: Technical Analysis and Intraday Exchange Rates”, FRBNY Economic Policy Review, (2000).

Pagano M., “Financial markets and growth: An overview”, European Economic Review, (1993).

Pandian J., Noel M., “Control of a bioreactor using a new partially supervised reinforcement learning algorithm”, Journal of Process Control, (2018).

Ponomarev E., Oseledets I., Cichocki A., “Using Reinforcement Learning in the Algorithmic Trading Problem. Journal of Communications Technology and Electronics”, (2019).

Poole D., Mackworth A., Goebel R., “Computational Intelligence: A Logical Approach”, Oxford University Press, (1998).

Quinn S., Roberds W., “An Economic Explanation of the Early Bank of Amsterdam, Debasement, Bills of Exchange, and the Emergence of the First Central Bank”, FRB Atlanta Working Paper, (2006).

Robbins L., “An Essay on the nature and significance of Economic Science”, (1932).

Rumelhart D., Hinton G., Williams R., “Learning representations by back-propagating errors”, Nature, (1986).

Sadighian J., “Deep Reinforcement Learning in Cryptocurrency Market Making”, (2019).

Schumpeter J., “Theorie der wirtschaftlichen Entwicklung”, (1911).

Shannon C., “Programming a Computer for Playing Chess”, Philosophical Magazine, (1950).

Shaw E., “Financial Deepening in Economic Development”, Oxford University Press, (1973).

Silver D., Hubert T., Schrittwieser J., Antonoglou I., Lai M., Guez A., Lanctot M., Sifre L., Kumaran D., Graepel T., Lillicrap T., Simonyan K., Hassabis D., “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm”, (2017).

Stiglitz J., “Financial Markets and Development”, Oxford Review of Economic Policy, (1989).

Théate T., Ernst D., “An Application of Deep Reinforcement Learning to Algorithmic Trading”, (2020).

Turing A., “Computing Machinery and Intelligence”, (1950).

Walras L., “Éléments d’économie politique pure”, (1874).

Zhang T., Xiao M., Zou Y., Xiao J., Chen S., “Robotic Curved Surface Tracking with a Neural Network for Angle Identification and Constant Force Control based on Reinforcement Learning”, (2020).

Artykuły publicystyczne:

Akston H., “Beating the Quants at Their Own Game”, Seeking Alpha, (2009). [online] (https://seekingalpha.com/article/114523-beating-the-quants-at-their-own-game)\

Chandra A., “McCulloch-Pitts Neuron — Mankind’s First Mathematical Model Of A Biological Neuron”, Towards Data Science, (2018). [online] (https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1)

Cummans J., “A Brief History of Bond Investing”, Bondfunds, (2014). [online] (http://bondfunds.com/education/a-brief-history-of-bond-investing/)

Dickson B., “What is symbolic artificial intelligence?”, TechTalks, (2019). [online] (https://bdtechtalks.com/2019/11/18/what-is-symbolic-artificial-intelligence/)

Floyd D., “Buffett's Bet with the Hedge Funds: And the Winner Is …”, Investopedia, (2019). [online] (https://www.investopedia.com/articles/investing/030916/buffetts-bet-hedge-funds-year-eight-brka-brkb.asp)

Harrington C., “Fundamental vs. Technical Analysis”, CFA Institute, (2003). [online] (https://www.cfainstitute.org/en/research/cfa-magazine/2003/fundamental-vs-technical-analysis)

Hayes A., Scott G., Kvilhaug S., “Black-Scholes Model Definition”, Investopedia, (2022). [online] (https://www.investopedia.com/terms/b/blackscholes.asp)

Kelion L., “DeepMind AI achieves Grandmaster status at Starcraft 2”, BBC, (2019). [online] (https://www.bbc.com/news/technology-50212841)

Knight W., “AlphaGo Zero Shows Machines Can Become Superhuman Without Any Help”, MIT Technology Review, (2017). [online] (https://www.technologyreview.com/2017/10/18/148511/alphago-zero-shows-machines-can-become-superhuman-without-any-help/)

Koch C., “How the Computer Beat the Go Master”, Scientific American, (2016). [online] (https://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/)

Loiseau J., “Rosenblatt’s perceptron, the first modern neural network”, Towards Data Science, (2019). [online] (https://towardsdatascience.com/rosenblatts-perceptron-the-very-first-neural-network-37a3ec09038a)

Maggiulli N., “Why the Medallion Fund is the Greatest Money-Making Machine of All Time”, Of Dollars And Data, (2019). [online] (https://ofdollarsanddata.com/medallion-fund/)

Rosenbaum E., “What Warren Buffett’s losing battle against S&P 500 says about this market”, CNBC, (2021). [online] (https://www.cnbc.com/2021/01/08/how-warren-buffetts-uphill-battle-against-the-sp-500-is-changing.html)

Saletan W., “Chess Bump: The triumphant teamwork of humans and computers”, Slate, (2007). [online] (https://slate.com/technology/2007/05/the-triumphant-teamwork-of-humans-and-computers.html)

Statt N., “DeepMind’s StarCraft 2 AI is now better than 99.8 percent of all human players”, The Verge, (2019) [online] (https://www.theverge.com/2019/10/30/20939147/deepmind-google-alphastar-starcraft-2-research-grandmaster-level)

Yi Peng N., “How Renaissance beat the markets with Machine Learning”, Towards Data Science, (2020). [online] (https://towardsdatascience.com/how-renaissance-beat-the-markets-with-machine-learning-606b17577797)

Raporty okresowe:

McKinsey & Company, “Ask the AI experts: What’s driving today’s progress in AI?”, (2017). [online] (https://www.mckinsey.com/capabilities/quantumblack/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai)

UNESCO, “The race against time for smarter development” (2021). [online] (https://www.unesco.org/reports/science/2021/en)

Bazy danych:

Companies Market Cap (companiesmarketcap.com; dostęp: 13.07.22r.)

7 Spis rysunków

8 Spis tabel

9 Spis kodów

Kody programów zawarto w Załączniku do pracy.

10 Streszczenie

Praca dotyczy problemu optymalizacji portfela inwestycyjnego. Przy konstrukcji rozwiązań zastosowane zostaje podejście zgodne z analizą techniczną. Zauważa się rosnący potencjał zastosowań uczenia ze wzmacnianiem przy rozwiązywaniu problemów optymalizacyjnych. Przedstawiony zostaje mechanizm matematyczny decydujący o funkcjonowaniu metod tej dziedziny w modelach probabilistycznych. Przeprowadzona analiza prac naukowych determinuje wybór probabilistycznych algorytmów A2C i PPO. Zbudowany zostaje zbiór danych odpowiedni do symulowania optymalizacji portfela względem przekroju światowej giełdy. Tworzy się programy implementujące algorytmy na zbiorze danych. Przy założeniu ewaluacji względem równego podziału i w porównaniu do rezultatów z zastosowania deterministycznego algorytmu CLA, ich działanie ocenia się jako niesatysfakcjonujące. Szacuje się 3 powody takiego stanu rzeczy: niedostateczne zasoby danych i mocy obliczeniowych, odtwórczy charakter wykorzystanych metod oraz ograniczenie paradygmatem analizy technicznej.