

**Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie**

Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki

KATEDRA INFORMATYKI



PRACA MAGISTERSKA

JACEK SPÓLNIK

**ZASTOSOWANIE METOD UCZENIA
MASZYNOWEGO W SYSTEMIE WSPOMAGAJĄCYM
ZARZĄDZANIE ZADANIAM W
PRZEDSIĘBIORSTWIE**

PROMOTOR:

dr inż. Rafał Dreżewski

Kraków 2011

OŚWIADCZENIE AUTORA PRACY

OŚWIADCZAM, ŚWIADOMY ODPOWIEDZIALNOŚCI KARNEJ ZA
POŚWIADCZENIE NIEPRAWDY, ŻE NINIEJSZĄ PRACĘ DYPLOMOWĄ
WYKONAŁEM OSOBIŚCIE I SAMODZIELNIE, I NIE KORZYSTAŁEM ZE
ŹRÓDEŁ INNYCH NIŻ WYMIENIONE W PRACY.

.....

PODPIS

AGH
University of Science and Technology in Krakow

Faculty of Electrical Engineering, Automatics, Computer Science and
Electronics

DEPARTMENT OF COMPUTER SCIENCE



**THE APPLICATION OF MACHINE LEARNING
METHODS IN THE SYSTEM SUPPORTING
TASKS MANAGEMENT WITHIN AN
ENTERPRISE**

JACEK SPÓLNIK

**THE APPLICATION OF MACHINE LEARNING
METHODS IN THE SYSTEM SUPPORTING TASKS
MANAGEMENT WITHIN AN ENTERPRISE**

SUPERVISOR:

Rafał Dreżewski Ph.D

Krakow 2011

Dla Madzi, mojej ukochanej
Żony

Spis treści

1. Wprowadzenie	7
1.1. Cele pracy	7
1.2. Zawartość pracy	8
2. Uczenie maszynowe	10
2.1. Zagadnienia Podstawowe	10
2.1.1. Definicja i znaczenie uczenia się	11
2.1.2. Programy uczące się	11
2.1.3. Koncepcja uczenia się	11
2.2. Indukcja drzew decyzyjnych	12
2.3. Indukcja reguł	12
2.4. Metody probabilistyczne	13
2.5. Podsumowanie	13
3. System zarządzania zadaniami	14
3.1. Opis systemu	14
3.2. Przykłady działania	14
3.3. Przygotowanie danych	14
3.4. Podsumowanie	14
4. Uczenie maszynowe w systemie zarządzania zadaniami	15
4.1. Zastosowanie praktyczne	15
4.2. Realizacja	15
5. Wyniki badań eksperymentalnych	16
5.1. Opis badań	16
5.2. Wyniki	16
6. Zakończenie	17
6.1. Wnioski	17
6.2. Kierunki badań	17

6.3. Podsumowanie.....	17
------------------------	----

1. Wprowadzenie

Zarządzanie zadaniami w każdej organizacji niesie z sobą sporo trudności i wyzwań. Jednym z nich jest na pewno efektywne zarządzanie i monitorowanie realizacji zadań. Trudnym problemem, który staje na naszej drodze, jest analiza ogromnych ilości informacji powiązanych z aktualnym stanem w przedsiębiorstwie. Niestety, jest ona niezbędna aby można było myśleć o efektywnym zarządzaniu na poziomie całej jednostki. Bezpośrednią konsekwencją tego jest niebagatelne znaczenie każdej podejmowanej decyzji, która musi być odpowiednio przemyślana oraz musi być na tyle elastyczna, aby łatwo można było wpływać na dalsze wyniki. Dodatkowo, kluczowa jest znajomość każdej nietrafnej decyzji, co pomaga unikać podobnych błędów w przyszłości.

Uczenie maszynowe jest konsekwencją idei sztucznej inteligencji oraz praktycznego wykorzystania tych idei. Wykorzystuje głównie takie dziedziny jak statystyka, informatyka, psychologia oraz wiele innych [4]. Głównym jej celem jest tworzenie automatycznego, samodoskonalącego się systemu, który wykorzystuje zgromadzone dane jako własne doświadczenie i na tej podstawie potrafi nabywać nową wiedzę tworząc nowe dane. Aktualnie jest stosowane szeroko w przemyśle wspierając tzw. inteligencję biznesową (ang. "Business Intelligence"). Istotą działania algorytmów uczenia maszynowego jest odpowiedni dobór parametrów, nazywanych wiedzą z wykorzystaniem metod pozyskiwania wiedzy oraz sposobów reprezentacji wiedzy [2][1][3].

Praca ta wprowadza koncepcję wykorzystania metod uczenia maszynowego we wspieraniu procesu zarządzania zadaniami w przedsiębiorstwie. Kluczowym elementem jest możliwość analizy i podejmowania inteligentnych decyzji przez algorytmy na podstawie zgromadzonego doświadczenia (dane) i ciągła nauka w celu efektywniejszego działania (nowe dane). Algorytmy z tej dziedziny stanowią nieocenione wsparcie dla menadżerów i innych osób podejmujących decyzje na poziomie całej organizacji.

1.1. Cele pracy

Problemem jaki przed nami staje jest problem efektywnego działania przedsiębiorstwa. (Dylematem, z którym musimy się zmierzyć, jest zagadnienie efektywnego działania przedsiębior-

stwa.) Organizacja musi realizować zadania, a jej wydajność zależy od efektywnego ich wykonywania. O ile pojedyncze zadanie wymaga specjalistycznych umiejętności od jednostki, to optymalne zarządzanie wszystkimi zadaniami na poziomie całego przedsiębiorstwa wymaga wiedzy o aktualnym stanie oraz doświadczenia. Menadżer, zanim zacznie efektywnie wykonywać swoją pracę, często przez jakiś czas musi się wdrażać w środowisko. Dodatkowo potrzeba czasu, aby zdobył potrzebne doświadczenie, a zastąpienie go jest niezwykle kosztowne.

Jeśli chodzi o same zadania, to można je opisać jako obiekty posiadające pewne cechy warunkujące przypisanie do konkretnego wykonawcy. W tej pracy chcemy stworzyć algorytm, który na podstawie istniejącej wiedzy (np. baza danych) będzie potrafił podejmować optymalne decyzje. Dzięki temu menadżer dostaje nieocenione wsparcie, a nawet istnieje możliwość zastąpienia menadżera w niektórych aspektach jego pracy. Jak wiemy, czas menadżera jest bardzo cenny dla każdej organizacji, a pomyłki popełnione przez niego bardzo kosztowne. Dzięki naszemu algorytmowi będziemy w stanie zmniejszyć to ryzyko, sprawiając że sama praca menadżera będzie bardziej komfortowa.

1.2. Zawartość pracy

Praca składa się ze wstępu oraz pięciu rozdziałów, z których każdy dotyczy osobnej części pracy, często wykorzystując rezultaty poprzedniego rozdziału.

Rozdział 2 wprowadza opis teoretyczny uczenia maszynowego. Poznamy historię uczenia maszynowego oraz samą koncepcję uczenia się w algorytmach. Następnie skupimy się na trzech rodzajach algorytmów klasyfikacji, które możemy użyć do rozwiązania naszego problemu. Na końcu rozdziału dokonamy podsumowania, wynikiem którego będzie wybór najlepszego algorytmu wraz z uzasadnieniem naszego wyboru.

Rozdział 3 przedstawia system zarządzania zadaniami, stworzony na potrzeby samej pracy. W systemie wzorujemy się na przedsiębiorstwie informatycznym. System pozwala użytkownikom na pracę z zadaniami, ich tworzenie, przypisywanie do odpowiedniej osoby oraz raportowanie czasu pracy i stanu zadań. Dodatkowo w tym rozdziale znajdziemy specyfikację danych przygotowanych za pomocą tego systemu, które pozwolą nam przeprowadzić praktycznie badania.

Rozdział 4 opisuje praktycznie wykorzystanie algorytmu uczenia maszynowego we wspomaganiu systemu zarządzania zadaniami. Wspomnianym systemem będzie oprogramowanie opisywane w rozdziale 3, oraz będzie wykorzystywać podstawy teoretyczne i wnioski z rozdziału 2.

Rozdział 5 przedstawia wyniki badań eksperymentalnych naszego algorytmu. W rozdziale tym opisujemy przeprowadzone badania oraz ich wyniki.

Rozdział 6 zamyka pracę, jest próbą podsumowania zawartych w niej treści oraz wskazania dalszej możliwości rozwoju idei wprowadzonej w pracy.

2. Uczenie maszynowe

W niniejszym rozdziale przedstawione zostaną zagadnienia teoretyczne bezpośrednio związane z tematyką pracy. Szczegółowo wyjaśnione zostanie pojęcie uczenia się oraz różne podejścia wykorzystujące koncepcję uczenia się. Omówione zostaną między innymi indukcja drzew decyzyjnych, indukcja reguł oraz metody probabilistyczne. Na koniec dokonamy porównania między poszczególnymi koncepcjami wraz z wnioskami, które pozwolą zastosować najlepszy wariant w praktycznej części pracy

2.1. Zagadnienia Podstawowe

Człowiek odkąd stworzył komputery zastanawiał się, czy komputery będą kiedyś dorównywać ludziom. Od samego początku podejmowano próby uczenia komputerów, w jaki sposób powinny postępować aby sprostać wyzwaniom, które stawiamy przed nimi. Sprowadzało się to głównie do tworzenia pewnego algorytmu, który w ścisły sposób radził sobie z pewnym określonym zadaniem. Jednakże aby komputer mógł działać bez pomocy z zewnątrz, należało by go wyposażać w mechanizm pozwalający na uczenie się na podstawie doświadczeń co skutkowało by ciągłą poprawą w działaniu samej maszyny. Wyobraźmy sobie komputery, które szybko i skutecznie diagnozują choroby, wykrywają anomalie pogodowe, inteligentne domy dostosowujące się do modelu życia każdej rodziny. Badania nad możliwością uczenia maszyn prowadzą również do lepszego zrozumienia natury ludzkiego uczenia się. Dziedzina ta ciągle się rozwija, wiele jeszcze brakuje aby uczynić maszynę zdolną do pojmowania wiedzy na poziomie człowieka, aczkolwiek efekty aktualnego stanu wiedzy już są wykorzystywane. Dziedziny takie jak rozpoznawanie mowy, medycyna, rozpoznawanie obrazów czerpią całymi garściami z tego co oferują aktualne algorytmy uczenia maszynowego. Rozdział ten skupia się na podstawowej teorii powiązanej z tematyką pracy. Opisuje znaczenie uczenia się, taksonomię maszynowego uczenia się oraz powiązanie ze sztuczną inteligencją.

2.1.1. Definicja i znaczenie uczenia się

Człowiek ucząc się zdobywa pewne umiejętności jak jazda na rowerze, bieganie, czytanie czy chodzenie. Podczas tego procesu zdarza się nam popełniać błędy, dokonywać kolejnych prób. Wykorzystujemy własne porażki i doświadczenia w celu podejmowania lepszych decyzji w przyszłości. Uczymy się czytając książki, korzystając z pomocy nauczycieli, uogólniamy nasze obserwacje i odkrywamy powiązania między nimi. Prowadzi to do prostego wniosku, chcąc mówić o uczących się programach komputerowych musimy mieć na uwadze programy komputerowe posiadające pewne szczególne właściwości, a dokładniej rzecz ujmując posiadające zdolność do uczenia się [1].

Postaramy się sprecyzować właściwości składające się na zdolność do uczenia się. Zaczniemy od prostego elementu jakim jest *zmiana*. Mówiąc o uczeniu się mamy na myśli pewną zachodzącą w czasie zmianę w aktualnym stanie wiedzy, umiejętności. Oczywiście nie każda zmiana wiąże się z samym uczeniem. Dlatego kolejnym kluczowym elementem jest *poprawa*, czyli zmiana przynosząca korzyść. Precyzując, chodzi o zmiany pozwalające usprawnić system w jego działaniu. O tym co stanowi takie usprawnienie systemu decyduje konstruktor takiego systemu, który wyposaża system w odpowiedni mechanizm oceny takiej poprawy. Ostatnim, równie ważnym elementem jest *autonomiczność* korzystnych zmian zachodzących w systemie. System sam jest w stanie zmieniać się na lepsze, tak jak człowiek sam jest w stanie nauczyć się wielu umiejętności. O takich zmianach będziemy mówić jako o *doświadczeniu* zdobywanemu przez system. Nasze rozważania prowadzą do definicji[1]:

Definicja 1 *Uczeniem się systemu jest każda autonomiczna zmiana w systemie zachodząca na podstawie doświadczeń, która prowadzi do poprawy jakości jej działania.*

Sama definicja dosyć jasno precyzuje kiedy sam proces możemy nazwać uczeniem. Jednakże ocena autonomiczności oraz faktycznej poprawy często nie jest łatwym zadaniem i zależy od samego projektanta systemu. Trafne decyzje prowadzą do systemu, który faktycznie w myśl naszej definicji 1 potrafi się uczyć i wykorzystywać nabyte doświadczenie.

2.1.2. Programy uczące się

2.1.3. Koncepcja uczenia się

Uczenie się w kontekście sztucznej inteligencji oraz automatyki rozumiane jest inaczej niż tradycyjnie. Proces uczenia się systemu ma za zadanie osiągnięcie rezultatów opartych na wiedzy fragmentarycznej, umożliwiać doskonalenie się, tworzyć nowe pojęcia oraz wnioskować indukcyjnie.

Herbert Simon (1983)

Uczenie się oznacza zmiany w systemie, które mają charakter adaptacyjny w tym sensie, że pozwalają systemowi wykonać za następnym razem takie same zadanie lub zadania podobne bardziej efektywnie.

Ryszard Michalski (1986)

Uczenie się to konstruowanie i zmiana reprezentacji doświadczanych faktów. W ocenie konstruowania reprezentacji bierze się pod uwagę: wiarygodność – określa stopień w jakim reprezentacji odpowiada rzeczywistości, efektywność – charakteryzuje przydatność reprezentacji do osiągania danego celu, poziom abstrakcji – odpowiada zakresowi szczegółowości i precyzji pojęć używanych w reprezentacji; określa on tzw. moc opisową reprezentacji. Reprezentacja jest rozumiana jako np. opisy symboliczne, algorytmy, modele symulacyjne, plany, obrazy.

Donald Michie (1991)

System uczący się wykorzystuje zewnętrzne dane empiryczne w celu tworzenia i aktualizacji podstaw dla udoskonalonego działania na podobnych danych w przyszłości oraz wyrażania tych podstaw w zrozumiałej i symbolicznej postaci.

2.2. Indukcja drzew decyzyjnych

Przez drzewo decyzyjne rozumiemy strukturę, która ma zwykłe właściwości drzewa w znaczeniu, jaki temu słowu nadaje się w informatyce, jest więc strukturą złożoną z węzłów, z których wychodzą gałęzie prowadzące do innych węzłów lub liści, oraz z liści, z których nie wychodzą żadne gałęzie. Węzły odpowiadają testom przeprowadzanym na wartościach atrybutów przykładów, gałęzie odpowiadają możliwym wynikom tych testów, liście zaś etykietom kategorii.

2.3. Indukcja reguł

W wielu systemach uczących się hipotezy reprezentowane są za pomocą reguł. Podobnie jak drzewa decyzyjne (które zresztą mogą być przekształcone do zbioru reguł), reprezentacja taka dopuszcza pojęcia wielokrotne, dla pewnego zbioru kategorii. Zazwyczaj dla każdej kategorii mamy regułę postaci:

IF pokrycie THEN , gdzie pokrycie reprezentuje warunki, jakie muszą spełniać wartości atrybutów przykładu , aby był on pokrywany przez tę regułę. (Dla pojedynczych pojęć można poprzestać na tylko jednej regule dla przykładów pozytywnych.) W celu wyznaczenia , gdzie jest reprezentowana za pomocą zbioru reguł tej postaci, znajduje się regułę pokrywającą i

zwraca jej kategorię. Jeśli jest takich reguł wiele, to bądź wybiera się pierwszą z nich (zakładając hierarchiczne uporządkowanie zbioru reguł w ciąg IF-THEN-ELSE-IF) bądź odpowiednio (np. probabilistycznie) łączy się kategorie wskazywane przez te reguły. Szczegóły zależą od konkretnego systemu.

2.4. Metody probabilistyczne

Metody oparte na twierdzeniu sformułowanym przez XVIII-wiecznego matematyka Thomasa Bayesa odgrywają znaczną i ostatnio rosnącą rolę w sztucznej inteligencji i zwłaszcza w uczeniu się maszyn. Można ogólnie powiedzieć, że dobrze znany wzór Bayesa stał się podstawą do rozwoju teorii i algorytmów różnych form wnioskowania probabilistycznego. My ograniczymy się do bliższego omówienia najprostszego, lecz bynajmniej nie pozbawionego znacznej praktycznej użyteczności bayesowskiego algorytmu uczenia się. Pokażemy też, jak z twierdzeniem Bayesa wiąże się jeden z bardziej popularnych tematów ostatnich lat, tzw. zasada minimalnej długości kodu.

2.5. Podsumowanie

Podsumowanie algorytmów uczenia maszynowego w kontekście naszej pracy.

3. System zarządzania zadaniami

Jest to nauka interdyscyplinarna ze szczególnym uwzględnieniem takich dziedzin jak informatyka, robotyka i statystyka. Głównym celem jest praktyczne zastosowanie dokonań w dziedzinie sztucznej inteligencji do stworzenia automatycznego systemu potrafiącego doskonalić się przy pomocy zgromadzonego doświadczenia (czyli danych) i nabywania na tej podstawie nowej wiedzy. Uczenie maszynowe jest konsekwencją rozwoju idei sztucznej inteligencji i metod jej wdrażania praktycznego. Dotyczy rozwoju oprogramowania stosowanego zwłaszcza w innowacyjnych technologiach i przemyśle. Odpowiednie algorytmy mają pozwolić oprogramowaniu na zautomatyzowanie procesu pozyskiwania i analizy danych do ulepszania i rozwoju własnego systemu. Uczenie się może być rozpatrywane jako konkretyzacja algorytmu czyli dobór parametrów, nazywanych wiedzą lub umiejętnością. Służy do tego wiele typów metod pozyskiwania wiedzy oraz sposobów reprezentowania wiedzy.

3.1. Opis systemu

OpisSystemu

3.2. Przykłady działania

Przykłady działania

3.3. Przygotowanie danych

Przygotowanie danych na potrzeby algorytmu uczenia maszynowego

3.4. Podsumowanie

Podsumowanie opisu danych i systemu zarządzania zadaniami.

4. Uczenie maszynowe w systemie zarządzania zadaniami

Uczenie maszynowe w systemie zarządzania zadaniami

4.1. Zastosowanie praktyczne

Zastosowanie praktyczne - opis

4.2. Realizacja

Realizacja - opis

5. Wyniki badań eksperymentalnych

Uczenie maszynowe w systemie zarządzania zadaniami - wyniki badań eksperymentalnych

5.1. Opis badań

Opis badań

5.2. Wyniki

Wyniki badań

6. Zakończenie

Zakończenie

6.1. Wnioski

Wnioski

6.2. Kierunki badań

Możliwe kierunki badań

6.3. Podsumowanie

Podsumowanie

Bibliografia

- [1] P. Cichosz. *Systemy uczące się*. Wydawnictwo Naukowo-Techniczne, Warszawa, 2007.
- [2] P. Z. i L. Bolc. *Wprowadzenie do uczenia się maszyn*. Akademicka Oficyna Wydawnicza, 1993.
- [3] T. M. Mitchell. *Machine Learning*. Wydawnictwo McGraw-Hill, Portland, 1997.
- [4] N. J. Nilsson. *Introduction To Machine Learning*. Robotics Laboratory, Department of Computer Science, Stanford, CA 94305, 1998.