

## Cel projektu

Pierwszym celem projektu jest stworzenie modeli uczenia maszynowego z niepewną wiedzą, rozszerzonych o matematyczne struktury i połączonych z modelami teoretycznymi w formie równań różniczkowych. Drugim celem jest stworzenie metod uczenia dodatkowej wiedzy do maszyn wektorów wspierających (SVM) na podstawie zaproponowanych modeli. W szczególności, rozważamy problemy klasyfikacji i regresji.

## Badania podstawowe

Planowana praca jest podzielona na dwie części, projektowanie modeli i metod uczenia maszynowego z połączonymi strukturami matematycznymi, i tworzenie metod uczenia równań różniczkowych z metodami uczenia maszynowego. Proponowane podejście dla pierwszej części uczenia matematycznych struktur składa się z trzech zadań sekwencyjnych:

1. Aksjomatyzacja problemów klasyfikacji i regresji dla modelu ze zbiorami wiedzy z połączonymi podstawowymi strukturami matematycznymi.
2. Wyprowadzenie teoretycznego uzasadnienia dla modeli zdefiniowanych w poprzednim kroku.
3. Opracowanie nowych metod i rozszerzenie SVM bazujących na zdefiniowanych modelach i teorii.

Proponowane podejście dla drugiej części uczenia równań różniczkowych z metodami uczenia maszynowego składa się z dwóch sekwencyjnych zadań:

1. Uczenie warunków początkowych i brzegowych dla równań różniczkowych do SVM.
2. Opracowanie metod uczenia wiedzy o rozwiązaniach układów równań różniczkowych do SVM.

## Powody podjęcia danej tematyki badawczej

Istnieją dwa główne podejścia do modelowania świata. Pierwsze z nich to stworzenie czysto teoretycznych modeli bazujących przykładowo na równaniach różniczkowych. To podejście jest z sukcesem stosowane w wielu dziedzinach takich jak fizyka, ekonomia. Drugie podejście to użycie ogólnych metod takich jak na przykład metody uczenia maszynowego do wyszukiwania modeli bazujących na danych empirycznych. Drugie podejście stało się ostatnio popularne z powodu coraz większej ilości dostępnych danych jednakże bez dodatkowej wiedzy o strukturze problemu. Szczęśliwie, nie wszystkie zadania mogłyby być lepiej modelowane za pomocą połączenia obu podejść po prostu. Ideą tego projektu jest skorzystanie z metod ogólnych jak SVM i użycie dodatkowej wiedzy pozyskanej z danych o matematycznej strukturze problemu i o modelu teoretycznym w formie równań różniczkowych. Możemy rozważać takie modelowanie w kategoriach inteligencji człowieka. Używamy dwóch typów systemów myślenia: Systemu 1 i Systemu 2 według laureata Nagrody Nobla Daniela Kahnemana [Kahneman2011]. Pierwszy z nich to intuicja, drugi z nich to myślenie racjonalne. Ogólne metody na ładuj System 1, podczas gdy dodatkowa wiedza i modele teoretyczne na ładuj System 2. Myślenie intuicyjne jest mniej dokładne, ale szybsze. Połączenie go ze ścisłymi regułami z Systemu 2 mogłoby poprawić wydajność rozwiązywania pewnych zadań. Metody ogólne uczą się z ograniczonej liczby przykładów i często bazują na założeniach statystycznych o niewidocznych przykładach, które są wygenerowane niezależnie z tego samego nieznanego rozkładu prawdopodobieństwa. Z tego powodu, statystycznie umotywowane metody ogólne pracują lepiej z większą liczbą przykładów. Ideą tego projektu jest poprawienie wydajności metod ogólnych poprzez---zamiast zwiększania liczby przykładów, co może być niemożliwe, dodawanie pewnej wiedzy a priori o matematycznej strukturze problemu i o modelach teoretycznych w formie równań różniczkowych. Długoterminowy cel dziedziny sztucznej inteligencji to stworzenie systemu sztucznej inteligencji. Nie jest możliwe stworzenie systemu sztucznej inteligencji tylko za pomocą metod statystycznych. Potrzebujemy jeszcze ścisłych reguł [Kisielewicz2011]. To jest powód uczenia dodatkowej wiedzy do modeli statystycznych. Kolejnym ideałem jest połączenie modeli statystycznych z modelami teoretycznymi, którymi są zazwyczaj równania różniczkowe. Taka kombinacja modeli mogłaby potencjalnie poprawić wydajność każdego z systemów z osobna. Możemy rozważać takie kombinacje z dwóch stron: 1. modele statystyczne mają dodatkową wiedzę o strukturze w formie równań różniczkowych 2. poprawiamy dokładność modelu bazującego na równaniach różniczkowych za pomocą danych empirycznych reprezentowanych przez model statystyczny.

Rozważmy porównanie modelu ze zbiorami wiedzy do modeli statystycznych. Modele statystyczne bazują na teorii prawdopodobieństwa są bardzo skutecznym podejściem do modelowania świata. Ale powstaje pytanie, czy metody takie jak SVM mogłyby wyprowadzone z bardziej ogólnych modeli nie zakładających, że dane są niezależne i o jednakowym rozkładzie (i.i.d.). Badane były możliwości osłabienia założeń statystycznych modeli uczenia maszynowego jak np. o danych i.i.d.. Podczas gdy w uczeniu maszynowym koncepcja użycia pewnego rodzaju uogólnionej teorii prawdopodobieństwa jest przeważnie nieobecna, w innych dziedzinach są pewne oznaki, że teoria prawdopodobieństwa z aksjomatami zdefiniowanymi przez Kołmogorowa nie jest wystarczająca do opisu świata i potrzebujemy bardziej ogólnej teorii. Z tego powodu w kosmologii w celu modelowania czarnych dziur stosowane jest wolne prawdopodobieństwo bazujące na nieprzemiennej algebrze von-Neumanna [Heller2014]. Idea wolnego prawdopodobieństwa staje się ostatnio bardziej popularna. Teoria ta jest stosowana także do opisu losowych macierzy. Jest ona bardziej nastawiona na matematyczną strukturę, którą jest algebra z mniejszym nastawieniem na losowość. Wszystkie ogólne twierdzenia teorii prawdopodobieństwa, jak prawo wielkich liczb, zostały przemodelowane. Jest to wczesna oznaka tego, że potrzebujemy uogólnionej teorii prawdopodobieństwa do opisu świata. Inny powód jest czysto spekulacyjny. Teoria prawdopodobieństwa została aksjomatyzowana przez Kołmogorowa w 1933, prawie 100 lat temu i jest raczej do góry boko wyeksploatowana. Bardziej ogólne pytanie to jak zastąpić niepewność w formie losowo wygenerowanych przykładów z pewnego nieznanego rozkładu innym typem niepewności w modelach uczenia maszynowego. Uczyliśmy pierwszy krok, definiując model ze zbiorami wiedzy z różnymi typami niepewności, takimi jak ograniczenia predykcyjne w [Orchel2015].

Kluczowym pojęciem w teorii uczenia statystycznego jest wymiar Vapnika-Chervonenkisa (VC). Wymiar VC klasy  $C$  hipotez to największa liczba  $N$ , dla której pewien zbiór  $N$  punktów jest rozdzielony przez reguły z  $C$  [Kulkarni2011]. Przykładowo, klasa liniowych hipotez w przestrzeni  $d$ -wymiarowej ma wymiar VC równy  $d+1$ . Podobne pojęcie to wymiar Poppera opisujący  $C$

trudno falsyfikacji klasy i jest zdefiniowany jako największa liczba  $N$  dla której każdy zbiór  $N$  punktów jest rozdzielony przez reguły z  $C$ . Dla klasy liniowych hipotez w przestrzeni  $d$ -wymiarowej wymiar Poppera jest równy 2, ponieważ 3 współliniowe punkty nie mogą być rozdzielone. W tym projekcie koncentrujemy się na ograniczeniu możliwych konfiguracji punktów, stosując wiedzę a priori o punktach. Przykładowo, jeżeli dodamy wiedzę o konfiguracjach z  $p > d$  punktami tak, że odcinki między punktami z tą samą klasyfikacją nie przecinają się, wtedy wymiar VC jest równy  $p$ . Następnym pytaniem dotyczy tego jak wybiera najlepszą hipotezę dla danego zbioru punktów. Różne ogólne reguły zostały zaproponowane takie jak brzytwa Ockhama, zasada minimalnego opisu, minimalizacja ryzyka strukturalnego (SRM). W ogólności bierzemy pod uwagę najlepsze dopasowanie do danych oraz faworyzujemy hipotezy z mniejszą złożonością.

#### Literatura

- [1] D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
- [2] A. Kisielewicz, *Sztuczna inteligencja i logika*. WNT, 2011.
- [3] M. Heller, *Granice nauki*. Copernicus Center Press, 2014.
- [4] M. Orchel, "Solving classification problems by knowledge sets," *Neurocomputing*, vol. 149, pp. 1109–1124, 2015.
- [5] S. Kulkarni and G. Harman, *An Elementary Introduction to Statistical Learning Theory*, ser. Wiley Series in Probability and Statistics. Wiley, 2011.