



SPRAWOZDANIE Z DZIAŁALNOŚCI
KOŁA NAUKOWEGO UCZENIA MASZYNOWEGO GHOST
W ROKU AKADEMICKIM 2017/2018
ORAZ W SEMESTRZE ZIMOWYM ROKU 2018/2019

Wstęp

Zgodnie z Regulaminem Koła Naukowego *Group Of Horribly Optimistic Statisticians*, celem działania organizacji jest rozwój zainteresowań oraz popularyzowanie zagadnień powiązanych z uczeniem maszynowym i statystyką wśród studentów Politechniki Poznańskiej.

Zarówno w ubiegłym roku akademickim 2017/2018, jak i bieżącym roku 2018/2019 Koło Naukowe GHOST podejmowało wielorakie działania w celu realizacji powyższego zadania. Po pierwsze, grupa GHOST organizowała regularne spotkania mające na celu rozwój wiedzy członków koła na temat algorytmów analizy danych i uczenia maszynowego. Członkowie grupy zdobywali także umiejętności praktyczne poprzez implementowanie poznawanych technik oraz algorytmów, udział w konkursach analizy danych oraz wykonywanie miniprojektów studenckich. Dalej, koło naukowe GHOST organizowało spotkania z zewnętrznymi ekspertami uczenia maszynowego, którzy prezentowali projekty dot. analizy danych realizowane w poznańskich firmach. Stworzyło to także możliwość nawiązania cennych kontaktów pomiędzy kołem naukowym, studentami Politechniki Poznańskiej a lokalnymi firmami zajmującymi się tematyką uczenia maszynowego. Rezultatem działalności koła naukowego są także sukcesy jego członków: wysokie lokaty w ogólnopolskich konkursach, udział w ogólnopolskich konferencjach, przyznanie prestiżowych stypendiów czy udział w międzynarodowych programach stażowych. Ponadto grupa realizowała również zadania popularyzatorskie poprzez organizowanie konkursów i wykładów otwartych wprowadzających do zagadnień uczenia maszynowego i analizy danych.

Koło Naukowe GHOST

Dnia 16 października 2017 roku na spotkaniu nieformalnej Grupy Studenckiej Uczenia Maszynowego i Statystyki GHOST wyłoniono na drodze głosowania skład władz Koła Naukowego GHOST na kadencję 2017/2018 w przypadku jego zatwierdzenia przez Jego Magnificencję Rektora Politechniki Poznańskiej. W pierwszej turze głosowania na stanowisko przewodniczącego koła wybrano Piotra Miarę, studenta kierunku Informatyka na Wydziale Informatyki, a na stanowisko wiceprzewodniczącego wybrano Jakuba Berezowskiego, również studenta kierunku Informatyka na Wydziale Informatyki.

W grudniu 2017 roku Jego Magnificencja Rektor Politechniki Poznańskiej zatwierdził regulamin koła naukowego GHOST, plan działań na rok akademicki 2017/2018 oraz wybór powyższych osób na wspomniane wyżej stanowiska. Reprezentanci grupy zostali także zaliczeni w poczet Rady Kół Naukowych Politechniki Poznańskiej. Na opiekuna koła naukowego został wyznaczony mgr inż. Mateusz Lango, asystent w Zakładzie Inteligentnych Systemów Wspomagania Decyzji Instytutu Informatyki Politechniki Poznańskiej. Wspomniany zakład jest również regulaminową siedzibą koła naukowego.

W marcu 2018 roku koło naukowe zostało zaproszone do Ogólnopolskiego Forum Kół Naukowych Metod Ilościowych zrzeszającym dziewięć kół naukowych z kilku polskich

uniwersytetów. Natomiast w czerwcu 2018 roku koło naukowe GHOST zostało uwzględnione przez inicjatywę „PL in ML” na liście aktywnych społeczności promujących uczenie maszynowe w Polsce. Koło GHOST było jedynym reprezentantem miasta Poznania na tej liście. W tym miejscu pragniemy także odnotować, że w spotkaniach koła cieszą się zainteresowaniem studentów innych poznańskich uczelni, np. Uniwersytetu im. Adama Mickiewicza czy Uniwersytetu Ekonomicznego, a także studentów innych wydziałów Politechniki Poznańskiej.

W październiku 2018, podczas walnego zgromadzenia członków koła GHOST, została przedłużona kadencja przewodniczącego Piotra Miary na kolejny rok akademicki. Jednocześnie na stanowisko wiceprzewodniczącego została mianowana Faustyna Krawiec, aktualnie studentka Informatyki na specjalności Inteligentne Systemy Wspomagania Decyzji.

Spotkania GHOST

Regularne spotkania członków koła naukowego GHOST mają na celu pogłębianie wiedzy w zakresie statystyki oraz uczenia maszynowego. Spotkania odbywają się w formie cotygodniowych wykładów lub prezentacji organizowanych przez członków koła i zwykle są skoncentrowane na jednej, konkretnej technice bądź zagadnieniu. Wykłady mają charakter zamknięty, trwają 90 minut, a każdemu z nich towarzyszy literatura obowiązkowa (zwykle jeden rozdział z podręczników do uczenia maszynowego bądź statystyki). Szczegółowy wykaz spotkań wraz z literaturą można znaleźć na oficjalnej stronie koła naukowego www.ghost.put.poznan.pl.

W ubiegłym roku akademickim tematem przewodnim spotkań były „Głębokie sieci neuronowe do rozpoznawania obrazów” (semestr 1) oraz „Drzewa decyzyjne i zespoły klasyfikatorów” (semestr 2). Poniżej prezentujemy konkretne tematy spotkań.

Cykl spotkań pt. „Głębokie sieci neuronowe do rozpoznawania obrazów” obejmował realizację następujących spotkań:

- Spotkanie organizacyjne
- Powtórzenie zagadnień poznanych w roku akademickim 2016/2017 oraz Podstawy teorii maszynowego uczenia się
- Regresja logistyczna
- Wstęp do sieci neuronowych
- Półautomatyczne tworzenie sieci słowotwórczych
- Wielomianowa regresja logistyczna
- Sieci neuronowe
- Algorytm wstecznej propagacji błędów
- Splotowe sieci neuronowe
- Wstęp do rekurencyjnych sieci neuronowych
- Sieci rekurencyjne

- Nowoczesne architektury rekurencyjne (LSTM)

Cykl spotkań pt. „Drzewa decyzyjne i zespoły klasyfikatorów” obejmował realizację następujących tematów:

- Wprowadzenie do drzew decyzyjnych
- Drzewa decyzyjne
- Lasy losowe
- Lasy losowe – techniki zaawansowane
- Modele drzewiaste dla klasyfikacji wieloetykietowej
- Analiza głównych składowych (jako wstęp do lasów rotowanych)
- Lasy rotowane
- Techniki budowy zespołów klasyfikatorów: Boosting
- Wprowadzenie do uczenia ze wzmocnieniem
- Uczenie ze wzmocnieniem

W ramach regularnych spotkań GHOST zorganizowano również trzy spotkania warsztatowe o głębokich sieciach konwolucyjnych, podczas których członkowie koła zaprezentowali artykuły naukowe dot. tego tematu. Listę studentów wraz z przypisanymi im artykułami przedstawiono poniżej:

- Piotr Miara: *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* Ioffe S. & Szegedy C.
- Magda Lipowska: *Going Deeper with Convolutions* by Szegedy C. et al.
- Marcin Konieczny: *Gradient-Based Learning Applied to Document Recognition: LeNet-5* LeCun Y. et al.
- Jakub Berezowski: *ImageNet Classification with Deep Convolutional Neural Networks* Krizhevsky A. et al.
- Mateusz Skiba: *Visualizing and Understanding Convolutional Networks* Zeiler M. & Fergus R.

Ponadto na kilku spotkaniach członkowie koła prezentowali wybrane przez nich artykuły naukowe, które dotyczyły aktualnie interesujących ich zagadnień np. regresji porządkowej, klasyfikacji danych sekwencyjnych czy głębokiego transferu stylów na zdjęciach.

W semestrze zimowym roku akademickiego 2018/2019 koło naukowe GHOST koło zrealizowało dwa cykle tematycznych spotkań dotyczące uczenia z danych niezbalansowanych oraz uczenia się rankingów. Na pierwszy cykl spotkań składały się następujące spotkania:

- Wstęp do uczenia się rankingów
- Uczenie się rankingów: podejścia punktowe i parowe
- Uczenie się rankingów: podejścia listowe
- Uczenie się rankingów: podejścia probabilistyczne

Podsumowaniem pracy nad zagadnieniem uczenia się rankingów były dwa spotkania poświęcone prezentacjom artykułów naukowych:

- PRanking with Ranking
- Combined regression and ranking
- Click Models for Web Search
- Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank
- Learning to Order Things
- Learning to Rank at Query-Time using Association Rules
- Adapting Ranking SVM to Document Retrieval

Drugi cykl spotkań dot. uczenia się z niezbalansowanych rozkładów obejmował następujące spotkania:

- Wprowadzenie do uczenia z danych niezbalansowanych
- Uczenie z danych niezbalansowanych: podejścia modyfikujące algorytmy i metody modyfikujące rozkład danych
- Źródła trudności w uczeniu z danych niezbalansowanych: analiza teoretyczna oraz jej wykorzystanie w projektowaniu metod modyfikujących rozkład danych
- Uczenie z danych niezbalansowanych: podejścia kosztowe

Analogicznie jak w przypadku pierwszego cyklu, wygłoszono prezentacje dot. artykułów naukowych:

- Adacost: Misclassification Cost-sensitive Boosting
- A systematic study of the class imbalance problem in convolutional neural networks
- Learning Deep Features for One-Class Classification

Dodatkowo na spotkaniach koła zaprezentowano dwa projekty wykonywane przez członków koła: "Training Neural Networks with Adaptive and Non-Adaptive Method" (Norbert Schlaffke, Oliwia Masian, Mateusz Słysz i Mateusz Skiba) oraz "Neural models for semantic code search" (Faustyna Krawiec).

Sesje przemysłowe GHOST

W ubiegłym roku akademickim 2017/2018 grupa zorganizowała trzy spotkania z ekspertami uczenia maszynowego z poznańskich firm informatycznych.

Andrzej Prałat, który pracuje na stanowisku Machine Learning Engineer w firmie Grey Wizard zajmującej się cyberbezpieczeństwem serwerów oraz aplikacji internetowych, wygłosił referat na temat „*Machine Learning for Web Applications Security*”. Podczas prezentacji słuchacze zapoznali się z nowoczesnymi technikami uczenia maszynowego wykorzystywanymi w cyberbezpieczeństwie aplikacji internetowych. Przeprowadzana została także dyskusja ograniczeń obecnych metod analizy danych w tym obszarze

zastosowań oraz omówione zostały możliwe ścieżki rozwoju i nowe zastosowania metod uczenia maszynowego w zabezpieczaniu aplikacji internetowych.

Mateusz Jukiewicz i Jakub Wąsikowski, pracownicy działu Data Science w firmie Roq.ad zajmującej się automatyczną identyfikacją urządzeń tego samego użytkownika, wygłosili prezentację na temat „*Entity Resolution for Cross-Device Identification*”. W czasie spotkania zaprezentowany został problem ujednoznaczniania encji oraz metod uczenia maszynowego do jego rozwiązywania. Prezentację podsumowała dyskusja zastosowania powyższych metod do problemu automatycznej identyfikacji urządzeń tego samego użytkownika.

Marek Wydmuch, Data Scientist w firmie OLX wygłosił prezentacją zatytułowaną „*Embed All The Things! Neural embedding models for wide variety of problems*”. Prezentacja obejmowała wiele interesujących metod, zaczynając od powtórki podstawowych pojęć związanych z głębokim uczeniem maszynowym, a kończąc na najpopularniejszej metodzie neuronowych zagnieżdżeń (word2vec). Druga część prezentacji obejmowała bardziej nowoczesne i zaawansowane techniki zagnieżdżeń jak np. StarSpace czy doc2vec. Pokazano także wykorzystanie powyższych metod do praktycznych problemów rozwiązywanych w firmie OLX. Każdy uczestnik wydarzenia otrzymał gadżety z firmy, informację o stażach, a także miał szansę zdobyć cenne nagrody poprzez odpowiadanie na wnikliwe pytania dot. prezentacji.

W semestrze zimowym bieżącego roku akademickiego zorganizowano spotkanie z Michałem Giełdą - członkiem zarządu firmy AntMicro, który wygłosił prelekcję na temat „*Edge AI - new capabilities, tools and platforms for a new wave of connected devices*”. Podczas prezentacji były prezentowane nowe architektury sprzętowe które wspierają głębokie sieci neuronowe takie jak Jetson Xavier, Movidius Myriad czy TPU. Została także podkreślona potrzeba otwartych architektur sprzętowych takich jak Risc-V i pokazano interesujące projekty nad którymi aktualnie pracuje firma np. SkyWall (automatyczne przechwytywanie dronów) czy kompresja sieci neuronowych.

Więcej informacji dot. powyższych wydarzeń można znaleźć na stronie internetowej koła naukowego GHOST.

Działania popularyzatorskie

W ramach działalności popularyzatorskiej koło zorganizowało dwa spotkania (6 listopada 2017 oraz 8 marca 2018) będące łagodnym wprowadzeniem w tematykę uczenia maszynowego i analizy danych. Spotkania te miały charakter otwarty, a ich celem, oprócz popularyzacji tematyki data science na Politechnice Poznańskiej, było zdobywanie nowych członków koła oraz umożliwienie im wdrożenia się w tematykę regularnych spotkań, które często dotyczą technik bardziej zaawansowanych.

Dnia 1 marca 2018 roku koło naukowe zorganizowało również poznańskiego hub'a konkursu Google Hashcode. Wydarzenie to, w którym wzięło ponad 120 osób, wsparł poznański oddział Polskiego Towarzystwa Informatycznego. Podczas konkursu uczestnicy starali rozwiązać się problem inteligentnego planowania przejazdów samochodów, uwzględniając odpowiednie ograniczenia biznesowe.

Członkowie koła naukowego uczestniczyli także w organizacji spotkania „Polskiego Porozumienia na rzecz Rozwoju Sztucznej Inteligencji” na Politechnice Poznańskiej w październiku 2018 r. podczas którego podpisano porozumienie o współpracy pomiędzy kluczowymi organizacjami zajmującymi się sztuczną inteligencją w Polsce m.in. Polskie Stowarzyszenie Sztucznej Inteligencji (PSSI), Polskie Towarzystwo Sieci Neuronowych (PTSN), Polska Grupa Systemów Uczących się PL SIGML, Polski Oddział IEEE SMC (Polish Chapter of the IEEE Systems, Man, and Cybernetics Society) oraz Polski Oddział IEEE Computational Intelligence Society.

Projekty implementacyjne

Zwieńczeniem wiedzy zdobytej przez członków koła naukowego na regularnych spotkaniach oraz sesjach przemysłowych była implementacja poznanych technik oraz algorytmów, a także udział w konkursach z uczenia maszynowego.

W ubiegłym roku akademickim celem technologicznym koła była nauka języka Python oraz biblioteki do głębokiego uczenia maszynowego TensorFlow. Członkowie koła mieli okazję do implementacji:

- Systemu rekomendacyjnego opartego na idei filtrowania kolaboratywnego
- Algorytmu stochastycznego spadku gradientu
- Wielowarstwowej sieci neuronowej wraz z algorytmem wstecznej propagacji błędów
- Regresji logistycznej (TensorFlow)
- Głębokiej sieci neuronowej (TensorFlow)

Powyższe projekty studenci wykonywali samodzielnie w domu, a na spotkaniach GHOST wyjaśniano różne wątpliwości, czy rozwiązywano niektóre trudności implementacyjne.

Zorganizowano również wewnętrzny konkurs analizy danych na internetowej platformie Kaggle. Członkowie koła mieli do zaprojektowania własny algorytm drzewiasty do zadania regresji. Warto nadmienić, że zwycięskie rozwiązanie uzyskało lepszy wynik na 10 zbiorach testowych niż profesjonalny algorytm drzewiasty dostępny w języku R (rpart).

Sukcesy i udział w konkursach

Poniżej prezentujemy listę sukcesów członków koła naukowego GHOST w poprzednim roku akademickim:

- stypendium Ministerstwa Nauki i Szkolnictwa Wyższego 2017/2018 (1 osoba)
- staż zagraniczny finansowany przez Unię Europejską w ramach programu *Summer of HPC* organizowanym przez PRACE (Partnership for Advanced Computing in Europe), projekt: „Big Data clustering with RHadoop” (1 osoba)

- staż zagraniczny w Amazon Development Center (zespół uczenia maszynowego) w Wielkiej Brytanii (2 osoby)
- staż zagraniczny w Microsoft Research (Machine Learning and Perception Group) w Wielkiej Brytanii (1 osoba)
- 1 miejsce w regionalnej edycji konkursu Microsoft Imagine Hack (rozwiązanie oparte na technikach uczenia maszynowego)
- 2 miejsce w konkursie Brainathon (rozwiązanie oparte na technikach uczenia maszynowego)
- 3 miejsce w konkursie CERHack (rozwiązanie oparte na technikach uczenia maszynowego)
- 5 miejsce (finał) w ogólnopolskim konkursie analizy danych DataNinja organizowanym przez OLX
- Publikacja na międzynarodowej konferencji "Discovery Science" na temat "Discovering Minority Sub-clusters and Local Difficulty Factors from Imbalanced Data" (WoS, 1 osoba)

Członkowie Koła Naukowego GHOST reprezentowali również Politechnikę Poznańską w różnych wydarzeniach i konkursach:

- udział w Inspirational Boot Camp for Girls in ICT 2016 organizowany przez Intel oraz „Perspektywy” (1 osoba)
- udział w programie Google Inside Look 2017 w Londynie (1 osoba)
- udział w Amazon Alexa Hackathon organizowany przez dział Amazon Video w Londynie (2 osoby)
- udział w konkursie HackYeah w Krakowie (4 osoby)
- udział w konkursie BrainCode 2017 organizowany przez Allegro (2 osoby)
- udział w konkursie analizy danych DataNinja organizowanym przez OLX (2 osoby)
- udział w konkursie Google HashCode poprzez organizowany przez koło hub w Poznaniu (7 osób)

W bieżącym roku akademickim członkowie koła osiągnęli następujące sukcesy:

- stypendium Ministerstwa Nauki i Szkolnictwa Wyższego 2018/2019 (1 osoba)
- stypendium "Nowe technologie dla dziewczyn", organizowane przez firmę Intel i fundację Perspektywy (2 osoby)
- zdobycie stażu zagranicznego w Google w Szwajcarii (2 osoby)
- prezentacja wygłoszona na konferencji PLinML: Polish View on Machine Learning w Warszawie przez wiceprzewodniczącą koła
- plakat naukowy na konferencji PLinML: Polish View on Machine Learning w Warszawie pt. "Semantic Code Search"
- plakat naukowy na konferencji PLinML: Polish View on Machine Learning w Warszawie pt. "Radiation, invisible power around you"

- 2 miejsce w konkursie Switch to Smart 2018 organizowanego przez Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW (rozwiązanie oparte na technikach uczenia maszynowego)
- 1 miejsce w konkursie analizy danych "ML STX NEXT"
- zdobycie certyfikatu " Practical Machine Learning" (1 osoba)
- prezentacja na spotkaniu organizowanym przez stowarzyszenie "Women in Machine Learning & Data Science" (1 osoba)

Dodatkowo, członkowie Koła Naukowego GHOST reprezentowali Politechnikę Poznańską w różnych wydarzeniach i konkursach:

- udział w konkursie HackYeah w Warszawie (3 osoby)
- udział w konkursie analizy danych DataNinja 2019 organizowanym przez OLX (6 osób)
- udział zdalny w konkursie Google HashCode (10 osób)
- uczestnictwo w konferencji PLinML: Polish View on Machine Learning (9 osób)
- uczestnictwo w konferencji PP-RAI 2018 (6 osób, w tym 5 osób jako wolontariusze)
- uczestnictwo w konferencji Women in Tech Summit 2018 (3 osoby, w tym dwie osoby jako wolontariusze)

Podsumowanie

Niniejszy dokument prezentuje sprawozdanie z działania Koła Naukowego *Group Of Horribly Optimistic Statisticians* w roku akademickim 2017/2018 oraz w semestrze zimowym 2018/2019. Koło podjęło różne aktywności: organizację sesji przemysłowych, wykładów otwartych, zamkniętych prezentacji dot. zaawansowanych zagadnień uczenia maszynowego czy też implementacji różnorodnych projektów studenckich. O wysokim zaangażowaniu członków koła świadczą liczne osiągnięcia, których lista jest szczególnie imponująca biorąc pod uwagę fakt, że w kole uczestniczy tylko kilkanaście studentów. Uważamy, że poprzez realizację powyższych działań osiągnięty został regulaminowy cel działania grupy, którym jest rozwój zainteresowań oraz popularyzowanie zagadnień związanych z uczeniem maszynowym i statystyką wśród studentów Politechniki Poznańskiej.