Onderwerpen Masterproef Opleiding Wiskunde

Academiejaar 2024-2025

Vakgroep Toegepaste Wiskunde, Informatica en Statistiek

Titel: Wiskundige modellen voor besmettelijke ziekten
Promotor: Prof. Marnix Van Daele
Begeleider: Prof. Willy Govaerts

Korte beschrijving:
In Deel III (Epidemische modellen) van de cursus Wiskundige Modellering van Prof. M. Van Daele (3de bachelor wiskunde) wordt de basistheorie over epidemische modellen ontwikkeld. Het gaat hierbij om compartimentele modellen waarbij de populatie opgesplitst wordt in een aantal besmette en een aantal niet-besmette compartimenten. Het model bestaat dan uit een stelsel gewone differentiaalvergelijkingen, voor ieder compartiment een vergelijking. Als er slechts een besmet compartiment is, dan wordt het begrip van het basisreproductiegetal ingevoerd door het volgen van de secundaire infecties die veroorzaakt worden door een enkel besmet individu dat in de ziektevrije vatbare populatie ingebracht wordt. De waarde van dit basisreproductiegetal in een ziektevrij evenwichtspunt bepaalt of de ziektevrije toestand stabiel is tegen het uitbreken van een infectie. Als er verschillende besmette compartimenten zijn, dan is er een meer algemene wiskundige benadering nodig, waarbij het begrip van de volgende generatie (next generation) matrix een rol speelt. Dat is een matrix waarvan het element op de plaats (i, j) het aantal secundaire infecties is dat in het i-de compartment veroorzaakt wordt door een infectie in het j– de compartiment. Het is dus een (n, n) matrix waarbij n het aantal besmette compartimenten is. In de literatuur zijn er hiervoor verschillende benaderingen die nauw verwant zijn, maar niet dezelfde basisveronderstellingen over de gebruikte modellen maken. We vermelden [2] en [3] (verschillend, alhoewel van dezelfde auteurs), [4, Ch. 9.10] en [6, Ch. 5.2]. De verschillende benaderingen gebruiken gelijkaardige geavanceerde stellingen over specifieke types van matrices (Z-sign pattern, niet-negatieve matrix, M-matrix) en verwijzen naar [1] voor specifieke resultaten. In de praktijk leiden ze meestal tot dezelfde resultaten, als de basisveronderstellingen van de respectieve benaderingen vervuld zijn In deze masterproef willen we de benaderingen kritisch vergelijken en de resultaten testen voor diverse epidemische modellen zoals die in de literatuur overvloedig voorhanden zijn. Het artikel [5] is eerder een overzichtsartikel waarin minder ingegaan wordt op de wiskundige achtergrond en meer op de epidemische toepassingen en op uitbreidingen. Het gaat onder meer over de afhankelijkheid van het basisreproductiegetal van specifieke parameters, hetgeen verwant is met controle van het basisreproductiegetal. Onder de voorbeelden vermelden we het West Nile virus, cholera, Anthrax, en Zika. Het boek [6] is een rijke bron van nog meer toepassingen en uitbreidingen.

Referentie
1. A. Berman and R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1970.
2. P. van den Driessche and J. Watmough, Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences, 180:29-48,2002. DOI:10.1016/S0025-5564(02)00108-6. 3.
3. P. van den Driessche and J. Watmough, Further notes on the basic reproduction number. In F. Brauer, P. van den Driessche, and J. Wu (Eds.), Mathematical epidemiology (pp. 159-178), 2008. Springer.
4. Fred Brauer and Carlos Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology (2nd edition, 2012). Springer Texts in Applied Mathematics 40.
5. Pauline van den Driessche, Reproduction numbers of infectious disease models. Infectious Disease Modelling, 2 (2017) 288-303.

6. Fred Brauer, Carlos Castello-Chavez and Zhilan Feng, Mathematical Models in Epidemiology, 2019 Springer Verlag, New York, Texts in Applied Mathematics 69.

Titel: Ontwikkeling van een kwaliteitscontrole systeem voor kwantum gegenereerde randomgetallen.

Promotor: Prof. Marnix Van Daele
Begeleider: Filip Pynckels (Auditeur-generaal ICT, FOD Binnenlandse Zaken)

Korte beschrijving:
Op basis van de studie Cryptosystems in an ever-changing world kan besloten worden dat een goede encryptie steunt op een goede hardware randomgenerator die niet te beïnvloeden is door externe invloeden. Vaak vindt men dergelijke hardware ingebed in de CPU van een computersysteem, maar deze kan door bijvoorbeeld elektromagnetische " foutinjectie" bijgestuurd worden in de richting van voorspelbaarheid. Een "proof of concept" randomgetallen generator op basis van "kwantum tunnel effect" werd ontwikkeld (zie onderstaande foto). Dit is evenwel nog maar de eerste stap in een volledig ontwikkeltraject (hardware aanpassingen, verkleinen, softwareontwikkeling en optimalisatie, kwaliteitscontrole van de gegenereerde randomgetallen, EMP-resistent maken, design van een cover, schrijven van documentatie, …).

De volgende stappen zijn:

• Een correcte methode ontwikkelen (theoretisch en programmatie van een controller met slechts 8Kb geheugen waarvan 2Kb voor variabelen) om de gegenereerde entropie om te zetten in randomgetallen die via de USB-poort naar een computersysteem worden gestuurd.
• Device drivers ontwikkelen (Windows, macOS en Linux) die gebruik maken van een correcte te ontwikkelen methode (theoretisch en programmatie) om de gegenereerde randomgetallen te ontvangen, snel fulltime te controleren op kwaliteit en door te geven aan andere programma's op het computersysteem. Bovenstaande stappen kunnen echter maar worden voltooid zodra er een standalone kwaliteitscontrole systeem voor randomgetallen beschikbaar is dat toelaat om de uitvoer van de boven beschreven stappen te controleren. Tijdens de ontwikkelfase van het project kan dit controle systeem dienen om de ontwikkeling te toetsen op correctheid, tijdens het gebruik van het eindproduct zal het systeem toelaten aan de gebruiker om eender wanneer na te gaan dat de randomgetallen generator correct werkt.

Voor deze scriptie wordt van de student verwacht dat hij/zij:
• Bestaande kwaliteitscontrole systemen en/of methodes voor gegenereerde randomgetallen vindt en bestudeert.
• Een goed kwaliteitscontrole systeem voor gegenereerde randomgetallen beschrijft en ontwikkelt of verder uitbouwt.
Meer specifiek wordt van de student verwacht dat hij/zij: Kennis opdoet van kwantum entropie generatie, elektromagnetische interferentie, elektromagnetische fout injectie, methodes om het hardware systeem te compromitteren, … (Opleiding van maximaal 2 uur).
• Kennis opdoet van de hardware aspecten van de hardware entropie generator in zijn huidige vorm (Opleiding van maximaal 1 uur).
• Bestudeert aan welke vereisten kwalitatieve random sequenties moeten voldoen (zelfstudie met overlegmomenten).
• Bestudeert welke de sterktes en zwaktes van bestaande kwaliteitssystemen zijn (zelfstandig werk met overlegmomenten).
• Een bestaand kwaliteitssysteem kiest en aanpast of er één ontwikkelt (zelfstandig werk met overlegmomenten).
• Het systeem dat hij/zij gekozen of gemaakt heeft verdedigt op basis van de zwaktes/sterktes ten opzichte van niet gekozen systemen (zelfstandig werk met overlegmomenten).
• Een korte beschrijving maakt hoe het door de student gekozen of ontwikkelde systeem moet gebruikt worden (zelfstandig werk met overlegmomenten).
• Het bovenstaande verwerkt in een samenvattende scriptie.

Referentie
• “Cryptosystems in an ever-changing world”, Freya Verbeke (begeleid door Prof. Marnix Van Daele, UGent, Faculteit Wetenschappen, Vakgroep Toegepaste Wiskunde, Informatica en Statistiek), 8 augustus 2019
• “Random number generator ver 2.0”, Robin Pynckels & Filip Pynckels, 20 februari 2021 (niet voltooid document), https://pynckels.github.io/2020-4-Random_number_generator_ver_2-0.pdf

---

Titel: Een ander onderwerp in de numerieke wiskunde
Promotor: prof. Marnix Van Daele

Korte beschrijving:
Bespreekbaar

Titel: Statistische besluitvorming na gebruik van machine learning technieken
Promotor: Prof. Stijn Vansteelandt

Korte beschrijving:
Meer en meer statistische analyses maken gebruik van machine learning technieken om bijvoorbeeld het effect van een behandeling of interventie te schatten. Zoals alle statistische analyses, genereren ook machine learning technieken onzekerheid op de resultaten (ten gevolge van steekproefvariatie); omwille van de complexiteit van de technieken wordt dit in de verdere analyse echter meestal genegeerd. Als gevolg hiervan zijn de betrouwbaarheidsintervallen die men bekomt, typisch zwaar vertekend.
In deze masterproef zult u betrokken worden in recente ontwikkelingen, alsook onderzoek binnen de onderzoeksgroep van de promotor, om na te gaan hoe men hiermee kan omgaan. De bestudeerde methodes zullen theoretische en/of door middel van simulatiestudies en concrete data-analyses worden geëvalueerd. Er kan gekozen worden voor hetzij een meer theoretisch onderwerp (bvb., een studie van auto-dml) of betrokkenheid in lopend onderzoek (ter bespreking), of een meer toegepast onderwerp (bvb., gebruik van debiased machine learning voor de analyse van intensieve zorgen data of geneesmiddelenstudies; evaluatie door middel van simulatie van de impact van de keuze van algoritmes in stacked learners, van het aantal folds in cross-fitting, ...; ontwikkelen van een R-pakket voor assumption-lean regression; ...).

Titel: Wanneer moeten standaard errors rekening houden met clustering?
Promotor: Prof. Stijn Vansteelandt

Korte beschrijving:
Het niet erkennen van clustering of afhankelijkheid tussen metingen in een regressie-analyse kan aanleiding geven tot sterk vertekende standaard errors. Statistici hebben manieren uitgewerkt om hiermee rekening te houden, namelijk door de cluster als categorische variabele in het model op te nemen als hetzij ‘fixed’ of ‘random’ effect. Soms blijkt dit echter overbodig te zijn, of zelfs tot een inflatie in standaard fouten te leiden. Via deze masterproef wensen we hier meer inzicht in te krijgen, enerzijds door een studie van recent werk van Nobelprijswinnaar Guido Imbens, en anderzijds door het probleem zelf onder de loep te nemen door middel van analytisch en/of simulatieonderzoek.

Titel: Afhankelijke kansvariabelen en hun som
Promotor: Prof. David Vyncke

Korte beschrijving:
Het risico waaraan een financiële of verzekeringsportefeuille blootgesteld is, hangt nauw samen met de verdeling van een som van kansvariabelen. Indien de kansvariabelen onafhankelijk zijn, kan men die som schrijven als een convolutie, maar een dergelijke voorwaarde is in de praktijk zelden voldaan. Rekening houden met de reële afhankelijkheidsstructuur brengt echter heel wat problemen met zich mee. Arbenz et al (2011) ontwierpen een algoritme dat de verdeling snel zou moeten berekenen, maar dat algoritme is beperkt tot positieve kansvariabelen en vereist bovendien dat de volledige copula gekend is. In de praktijk is echter vaak slechts gedeeltelijke informatie over de afhankelijkheid bekend. Door gebruik te maken van dergelijke informatie is het mogelijk om de (verdeling van de) som te begrenzen, zoals geïllustreerd in Bernard et al (2017) en Lux & Papapantoleon (2019). In deze masterproef bestudeert de student verscheidene technieken om de verdeling van een som van afhankelijke kansvariabelen te berekenen en/of te begrenzen.

Referentie:

- Arbenz P., Embrechts P. & Puccetti G. (2011). The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. Bernoulli 17(2), 562–591.
- Bernard C., Rüschendorf L. & Vanduffel S. (2017). Value-at-risk bounds with variance constraints. Journal of Risk and Insurance 84, 923–959.
- Lux T. & Papapantoleon A. (2019). Model-free bounds on Value-at-Risk using extreme value information and statistical distances. Insurance: Mathematics and Economics 86,

---

Titel: Een ander onderwerp in de financiële of actuariële wiskunde
Promotor: prof. David Vyncke

Korte beschrijving:
Bespreekbaar

Titel: Topological data analysis with the Mapper algorithm
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Topological data analysis (TDA) attempts to extract geometric properties present in numerical datasets [1]. Concretely, algebraic topology is used to identify certain geometric regularities in the data. From the point of view of topology, a cloud of points in an n-dimensional space is a totally disconnected set. If we use a notion of distance or similarity, it is possible to connect points, or clusters of points, to construct simplicial complexes. The Mapper algorithm, introduced by Singh, Mémoli and Carlsson [2], is a computational method for extracting descriptions of high-dimensional datasets from simplicial complexes. It has been used to reduce and capture topological and geometric information present in data. Recently, in [3] it was proposed to use Mapper to generate different coverings from a numerical dataset. In the context of rough set theory [4], such coverings serve to construct approximations of data, which in turn can be used inside various machine learning algorithms, including classification, feature selection, instance selection, etc.

The goal of this thesis is to explore the various possibilities of Mapper for generating data coverings. Depending on the interests of the student, this can be done in a purely theoretical way (investigating the mathematical underpinnings and properties of the method), and/or by applying Mapper's output to specific benchmark machine

learning problems. A possible research direction could also be to extend the approach to fuzzy set theory, although this is definitely not mandatory (nor is any previous knowledge of fuzzy sets required for this thesis).

Referentie
[1] F. Chazal, B. Michel, An introduction to topological data analysis: fundamental and practical aspects for data scientists, Frontiers in Artificial Intelligence 4, 2021.
[2] G. Singh, F. Mémoli, G. Carlsson, Topological methods for the analysis of high dimensional data sets and 3D object recognition, Eurographics Symposium on Point-Based Graphics, 2007.[3] M. Restrepo, C. Cornelis, Mapper-based rough sets, Proceedings of International Joint Conference on Rough Sets (IJCRS), May 2024, in press.
[4] Z. Pawlak, Rough Sets, International Journal of Computer and Information Sciences 11(5), pp. 341-356, 1982.

Titel: Similarity learning for fuzzy rough rule induction
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression [1].FRS make use of a similarity relation, which expresses the degree to which two elements in a dataset are related. Instead of using a normal, fixed similarity relation, it is possible to apply similarity learning or distance metric learning (DML)[2] to learn the optimal relation from the data. This has been successfully applied to neighborhoud-based machine learning methods, such as k-nearest neighbours. Rule induction is a machine learning model which learns rules from the data with the aim of e.g. predicting the class of new samples. One such technique which makes use of FRS theory, is QuickRules [3], which creates rules in a greedy way. The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the predictive and descriptive potential of similarity learning in the context of rule induction.

The specific research questions that the thesis should address are:
• How can we apply DML to rule induction?
• Which DML methods perform best in this context, what are their strengths and weaknesses w.r.t. predictive and descriptive performance?
• Can we improve these methods or create our own technique which outperforms them?
• Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, ...) on which the approach performs better or worse? Prior knowledge of fuzzy rough sets is not necessary, and a Python implementations of QuickRules will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie
[1] S. Vluymans, et al. “Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey.” Fundamentae Informaticae 142.1-4 (2015): 53-86.
[2] J. L. Suárez, S. García, F. Herrera, A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges, Neurocomputing 425 (2021) 300–322.
[3] Richard Jensen, Chris Cornelis, and Qiang Shen. Hybrid fuzzy-rough rule induction and feature selection. pages 1151–1156, 09 2009.

Titel: Experimental evaluation of new fuzzy rough set models for machine learning tasks
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression. [1]
A prominent example is fuzzy rough nearest neighbour classification (FRNN), which is a lazy learner that associates an upper and a lower approximation with each decision class and classifies test instances according to their membership in these. This has a transparent interpretation: upper approximation membership encodes to what extent a test instance is similar to the training instances of a class, and so possibly belongs to this class, whereas lower approximation membership encodes to what extent a test instance is not similar to the training instances of other classes and so necessarily belongs to this class.

One of the most promising FRS models uses Ordered Weighted Averaging (OWA) operators in the calculation of the upper and lower approximations. Recently, a new class of fuzzy rough set models have been developed at Ghent University, based on fuzzy quantification models. [2]The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the machine learning potential of the newly proposed FRS models.

The specific research questions that the thesis should address are:
• Are the new fuzzy rough set models based on fuzzy quantification an improvement on the existing OWA FRS model for feature selection?
• What are the strengths and weaknesses of each FRS model?
• Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, ...) on which the approach performs better or worse? Prior knowledge of fuzzy rough sets is not necessary, and Python implementations of the different FRS models will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie
[1] Vluymans, Sarah, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundam. Informaticae 142.1-4 (2015): 53-86.[2] Theerens, Adnan, and Chris Cornelis. "Fuzzy Rough Sets Based on Fuzzy Quantification." arXiv preprint arXiv:2212.04327 (2022).

Titel: Normalised outlier scores and their application in robust fuzzy rough sets
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression [1]. However, outliers in the dataset can have a significant impact on the performance of the model. To overcome this issue, outlier detection techniques have been introduced, where outlier scores are assigned to each data point. Outlier scores are often normalised to improve the interpretability of results. Since these normalised outlier scores can be viewed as fuzzy sets, they are well fitted to be combined with fuzzy rough sets. However, the effectiveness of normalised outlier scores in robust fuzzy rough set models is yet to be explored. The objective of this thesis is to investigate the application of normalised outlier scores in robust fuzzy rough set models.

Specifically, the goal is to:
• Conduct a comprehensive review of the literature on normalised outlier scores.

• Develop a framework that integrates normalised outlier scores into robust fuzzy rough sets.
• Implement the proposed framework on benchmark datasets and evaluate its effectiveness in improving the accuracy of classification based on fuzzy rough sets. Two potential frameworks for integrating normalised outlier scores are readily available: Choquet-based fuzzy rough sets [2] and confidence-based fuzzy quantifier FRS [3]. Prior knowledge of fuzzy rough sets is not necessary, and Python implementations of the different FRS models will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie
[1] Vluymans, Sarah, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundam. Informaticae 142.1-4 (2015): 53-86.
[2] Theerens, Adnan, Oliver Urs Lenz, and Chris Cornelis. "Choquet-based fuzzy rough sets." Int. Journal of Approximate Reasoning 146 (2022): 62-78.
[3] Theerens, Adnan, and Chris Cornelis. "Fuzzy Quantifier-Based Fuzzy Rough Sets." 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2022.
[4] Theerens, Adnan, and Chris Cornelis. "Fuzzy Rough Sets Based on Fuzzy Quantification." arXiv preprint arXiv:2212.04327 (2022).

_____T

Titel: Statistical inference for optimal treatment regimes
Promotor: Prof. Oliver Dukes

Korte beschrijving:
Identifying the optimal treatment strategy or decision for a patient lies at the heart of personalized medicine. However, providing valid tests and confidence intervals in this setting turns out to be challenging when the optimal regime is non-unique. A simple way of obtaining valid inference is by sample-splitting; the optimal regime is learn on a training sample, and confidence intervals are conducted on a test sample. however, this wastes information, and results may depend heavily on the choice of split. The aim of this project is to construct confidence intervals for the mean outcome under the optimal strategy that improves upon naive sample-splitting.

_____

Titel: Debiased machine learning with instrumental variables and binary outcomes
Promotor: Prof. Oliver Dukes

Korte beschrijving:
This project will consider inference for the causal effect of a binary exposure in a setting with a valid instrumental variable. This is a variable that has no direct effect on the outcome, is itself not subject to unmeasured confounding, and is predictive of the exposure. It can be used to infer the ATE even when the exposure-outcome relationship is distorted by unmeasured confounding. Existing results for statistical inference have relied on an additive homogeneity condition which is not usually plausible with a binary outcome. In this project, we will explore inference under a different assumption. We will begin by obtaining efficiency bounds under differing homogeneity conditions, and then understanding the asymptotic properties of corresponding 'debased machine learning' estimators. The considered methods will be evaluated theoretically and/or by means of simulation studies and concrete data analyses

Titel: Information-theoretic lower bounds for causal effects
Promotor: Prof. Oliver Dukes

Korte beschrijving:
Semiparametric causal inference often focuses on the estimation of statistical functionals, such as the average treatment effect. If we are unwilling to impose parametric assumptions of the data generating process, a natural question is how well we can hope to estimate our parameter. This turns out to be relatively straight forward when nuisance parameters are smooth or sparse enough; one can then relate to semiparametric efficiency theory to obtain a lower bound on the asymptotic variance. Matters become considerably more challenging when we do not have sufficient smoothness or sparsity however. This theory will review how concepts and techniques from information theory can help determine notions of optimality, as well as point towards optimal estimators, in the general case.

Titel: Gezamenlijke modellen voor tijd tot sterfte en kwaliteit van leven (terwijl men leeft)
Promotor: Prof. Els Goetghebeur

Korte beschrijving:
Geneesmiddelen voor kankerpatiënten (in een laat stadium) hebben typisch niet alleen een impact op de overlevingstijd maar komen soms ook met een kost of voordeel wat de kwaliteit van leven betreft. `Mixed Models' zijn een vorm van lineaire modellen voor herhaalde metingen van de kwaliteit van leven, die aan elke patiënt een random intercept (en soms ook andere random regressiecoefficiënten) geven. Conditioneel op de random intercept kunnen ze nog steeds onafhankelijke uitkomsten vooronderstellen. De toepassing van die modellen in onze setting gebeurt vaak op zo'n manier dat men de facto`missing data na de dood' imputeert alsof ze`missing at random' waren om de gemiddelde kwaliteit van leven over de tijd te berekenen. Dat levert sterk imaginaire (onrealistische) resultaten op als men een groep met en een zonder experimentele behandeling met elkaar vergelijkt. Alternatieve (cross sectionele') modellen zijn zeer geschikt om de verwachte kwaliteit van leven te schatten over de tijd, maar doen geen poging om de correlaties tussen de uitkomsten over de tijd mee te modelleren. In deze thesis zullen we vertrekken van mixed models en hun verwachte uitkomsten berekenen *onder de levenden* na verloop van tijd. We zullen deze modellen fitten op een dataset van kankerpatiënten in een laat stadium en onderzoeken hoe goed de modellen er in slagen om diverse aspecten van de resultaten onder de twee behandelingen correct te laten schatten. We zullen daartoe hun statistische eigenschappen in die context vergelijken en/of simulaties uitvoeren.

Efficiënte Multi-Probabilistische Predictie als Verdediging tegen Adversarial Attacks op Grote Classificatieproblemen

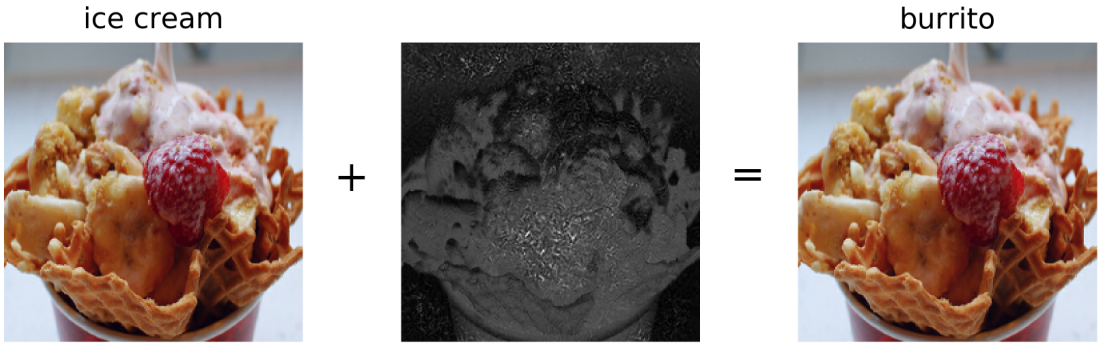
Begeleiders

- Jonathan Peck (WE02)
- Bart Goossens (TW07)
- Yvan Saeys (WE02)

Contact: `Jonathan.Peck@UGent.be`

Beschrijving

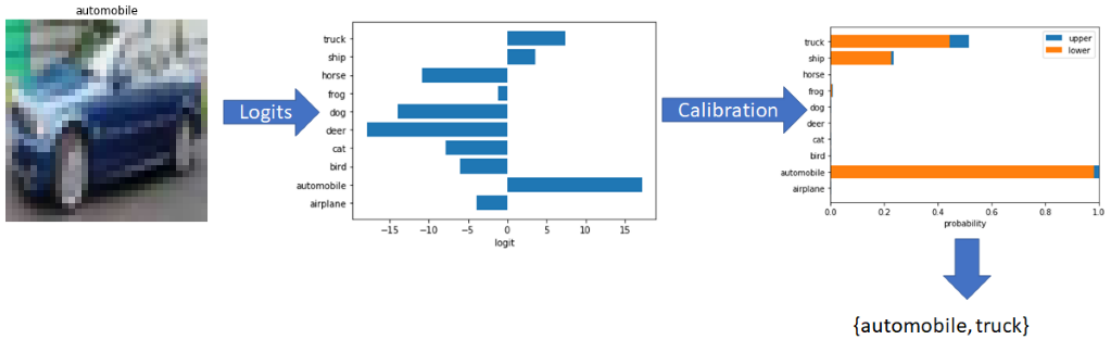
Neurale netwerken worden tegenwoordig overal gebruikt in domeinen zoals objectherkenning, natuurlijke taalverwerking en spraakherkenning. Deze modellen zijn echter *black boxes*: ze zijn moeilijk te interpreteren en hun precieze werking is hierdoor moeilijk te verklaren, wat problematische gevolgen kan hebben. Een gekend probleem is het fenomeen van *adversarial perturbations*: kleine, vaak onzichtbare wijzigingen in de invoer die ervoor zorgen dat het netwerk een totaal andere uitvoer geeft [1]. We tonen een voorbeeld hiervan in figuur 1. Het neuraal netwerk herkent de linkerafbeelding inderdaad als een ijsje, maar als de (onzichtbare) perturbatie in het midden wordt toegevoegd, krijgen we de rechterafbeelding. Deze afbeelding wordt plotseling verward met een burrito.



Figuur 1. Een adversarial perturbation in beeldherkenning.

Het doel van deze thesis is een bestaande methode te verbeteren die kan verdedigen tegen dit soort aanvallen. De aanpak die we hier hanteren, vertrekt van het neuraal netwerk als *black box*

en probeert de onzekerheid van de predictie te kwantificeren. Figuur 2 toont een schematisch overzicht van de werkwijze op een afbeelding uit de CIFAR-10 dataset. Het neuraal netwerk verwerkt eerst de afbeelding en produceert een scorevector met één score per klasse. Deze scores worden dan *gecalibreerd* zodat we voor elke klasse een bovengrens en ondergrens krijgen op de probabiliteit. Op basis van een gevoeligheidsparameter $\varepsilon \in [0, 1]$ wordt dan een finale selectie gemaakt van de mogelijke klassen die bij dit beeld horen. In het voorbeeld komen we uit op twee mogelijkheden: *automobile* of *truck*. Algemeen kan dit echter eender welke deelverzameling van alle mogelijke klassen zijn. In het bijzonder kan de predictieset dus ook leeg zijn of alle klassen bevatten. Indien de verzameling leeg is of teveel klassen bevat, kunnen we besluiten dat het model onbetrouwbaar is op de gegeven invoer.



Figuur 2. Schematisch overzicht van de verdedigingsmethode.

De focus van de thesis ligt op de calibratiestap, die als volgt gaat:

1. Bepaal de scorevector $(s_1, \ldots, s_K)$. Deze vector krijg je door de invoer aan het neuraal netwerk te geven. Er is één component voor elke klasse, met in totaal K klassen.
2. Bepaal een vector van ondergrenzen $(l_1, \ldots, l_K)$ en een vector van bovengrenzen $(u_1, \ldots, u_K)$ zodat de kans dat de invoer behoort tot klasse i tussen l_i en u_i ligt. Deze stap wordt afgehandeld door een bestaand algoritme: de IVAP van Vovk et al. (2015) [3].
3. Gebruik de onder- en bovengrenzen om een predictieset V te bepalen die de correcte klasse bevat met kans $1 - \varepsilon$, waar $\varepsilon \in [0, 1]$ een parameter van het algoritme is. Dit wordt gedaan via volgend optimalisatieprobleem:

$$\max_{\alpha_1, \ldots, \alpha_K, q} \alpha_1 + \ldots + \alpha_K + q$$
$$\text{zodat} \begin{cases} \alpha_1, \ldots, \alpha_K \in \{0, 1\} \\ q \in [0, 1] \\ q + \alpha_i(1 - u_i) \leq 1 \\ \sum_{i=1}^{K} \alpha_i(l_i - 1) \geq \varepsilon - 1 \end{cases}$$

Hier zijn $\alpha_1, \ldots, \alpha_K$ binaire indicatorvariabelen die de predictieset V bepalen: $i \in V \iff \alpha_i = 1$. Dit algoritme wordt verder toegelicht in [2]. Merk op dat dit algoritme perfect kan werken met

black boxes, aangezien er geen assumpties worden gemaakt over het onderliggende model. Deze methode is dus toepasbaar op alle classificatie-algoritmes die scorevectoren produceren.

Hoewel dit algoritme een effectieve verdediging geeft tegen adversarial attacks, schaalt het niet gunstig met het aantal klassen: er is namelijk één variabele nodig per klasse in het optimalisatieprobleem, wat in het ergste geval een kwadratische tijdscomplexiteit geeft in het aantal klassen. Voor grote classificatieproblemen met duizenden klassen is dit algoritme niet erg efficiënt.

De probleemstelling van deze thesis is het verbeteren van de efficiëntie van dit algoritme zodat een betere tijdscomplexiteit in het aantal klassen behaald kan worden. Hierbij worden de theoretische garanties liefst zoveel mogelijk gerespecteerd.

Referenties

1. Peck, Jonathan, Bart Goossens, and Yvan Saeys. "An introduction to adversarially robust deep learning." IEEE Transactions on Pattern Analysis and Machine Intelligence (2023). PDF
2. Peck, Jonathan, Bart Goossens, and Yvan Saeys. "Calibrated multi-probabilistic prediction as a defense against adversarial attacks." 28th BENELEARN 2019. Vol. 1196. Springer, 2020. PDF
3. Vovk, Vladimir, Ivan Petej, and Valentina Fedorova. "Large-scale probabilistic predictors with and without guarantees of validity." Advances in Neural Information Processing Systems 28 (2015). PDF
4. Shafer, Glenn, and Vladimir Vovk. "A tutorial on conformal prediction." Journal of Machine Learning Research 9.3 (2008). PDF

Beschermen van afbeeldingen tegen AI-gebaseerde manipulatie

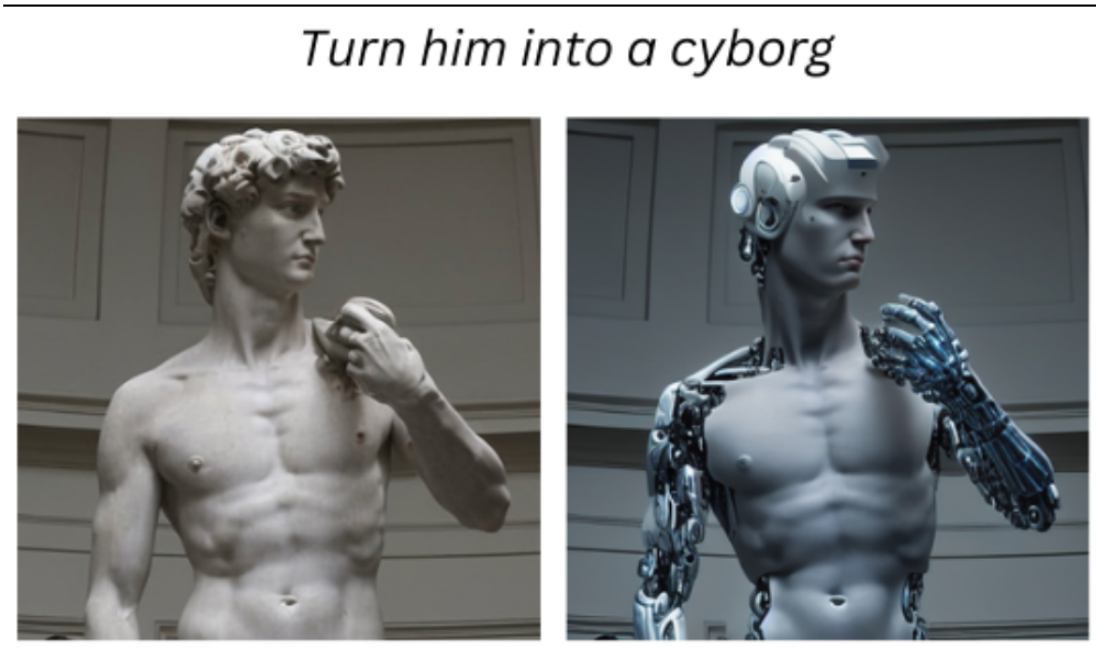
Begeleiders

- Jonathan Peck (WE02)
- Bart Goossens (TW07)
- Yvan Saeys (WE02)

Contact: Jonathan.Peck@UGent.be

Beschrijving

Generatieve AI-modellen hebben een revolutie ontketend in het manipuleren van beelden: gebruikers kunnen tegenwoordig instructies in natuurlijke taal geven en het AI-systeem zal de gevraagde wijzigingen uitvoeren [1]. Hoewel deze systemen uiteraard zeer nuttig kunnen zijn voor allerlei toepassingen, brengen ze ook risico's met zich mee. Zo vereenvoudigen ze bijvoorbeeld de creatie van "revenge porn" of ander schadelijk beeldmateriaal. Ze maken de verspreiding van misinformatie ook eenvoudiger wanneer beelden gemanipuleerd kunnen worden voor politieke doeleinden.

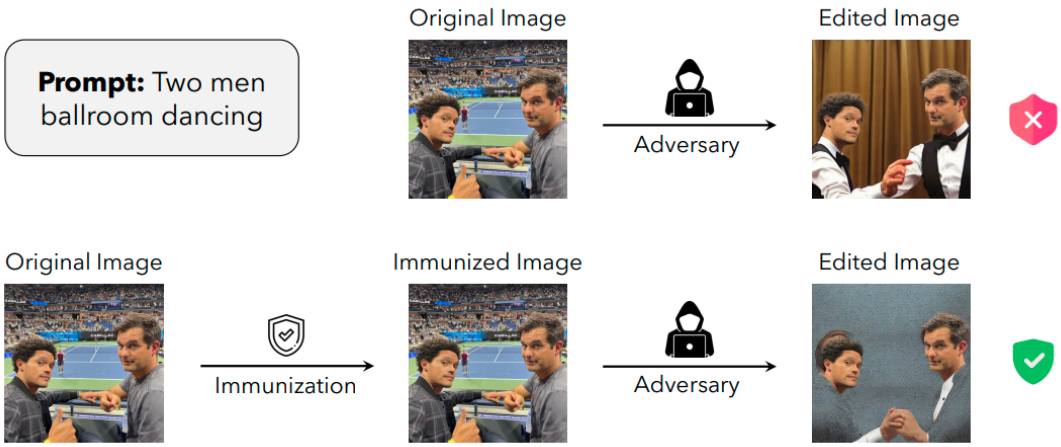


Figuur 1. Manipulatie van een afbeelding met generatieve AI.

Het doel van deze thesis is het ontwikkelen van een methode die AI-gebaseerde manipulatie van afbeeldingen tegengaat. Hiertoe willen we een algoritme ontwikkelen dat, gegeven een afbeelding als invoer, een nieuwe afbeelding produceert die aan twee voorwaarden voldoet:

1. De nieuwe afbeelding is visueel moeilijk te onderscheiden van de originele invoer.
2. Bestaande generatieve AI-systemen kunnen de nieuwe afbeelding niet betrouwbaar manipuleren.

De werkwijze wordt samengevat in figuur 2. Hier starten we met een originele afbeelding van twee personen bij een tenniswedstrijd. De afbeelding ondergaat dan een *immunisatie*, wat ons een nieuwe afbeelding oplevert die visueel identiek lijkt aan het origineel. Als we met een generatieve AI echter de setting van de afbeelding proberen te wijzigen naar een dansvloer, dan levert dit nonsens op voor de beschermde afbeelding terwijl het probleemloos lukt voor het origineel.



Figuur 2. Beschermen van afbeeldingen tegen AI-gebaseerde manipulatie.

Een *proof-of-concept* van dit idee is al gepubliceerd [2]. De methode die daar werd ontwikkeld, is echter niet heel robuust: als de originele afbeelding lichtjes gewijzigd wordt (bijvoorbeeld door toevoeging van ruis), dan is de bescherming niet meer doeltreffend. Bovendien werkt de bescherming ook enkel tegen één specifiek AI-systeem; er zijn geen garanties voor andere generatieve modellen.

Het doel van deze thesis is het verbeteren van deze methode zodat ze effectief blijft wanneer de afbeeldingen onderhevig zijn aan ruis en tegen een brede waaier aan generatieve systemen.

Referenties

1. Brooks, Tim, Aleksander Holynski, and Alexei A. Efros. "InstructPix2Pix: Learning to follow image editing instructions." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. PDF
2. Salman, Hadi, et al. "Raising the cost of malicious AI-powered image editing." arXiv preprint arXiv:2302.06588 (2023). PDF

Robuustheid van diffusiemodellen tegen adversarial attacks

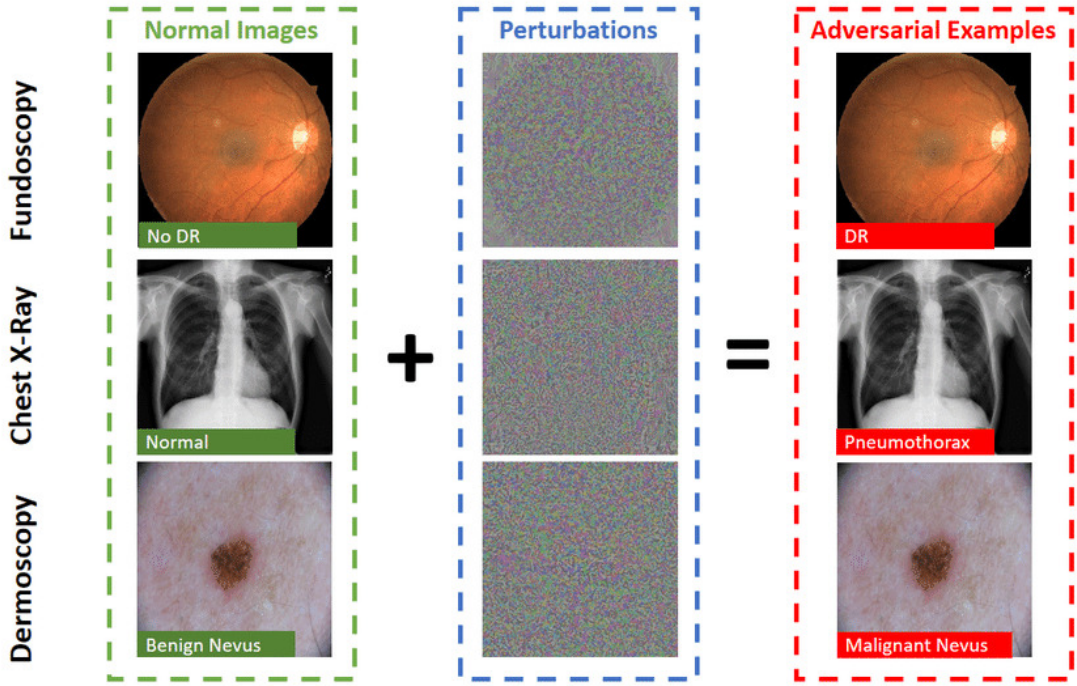
Begeleiders

- Jonathan Peck (WE02)
- Bart Goossens (TW07)
- Yvan Saeys (WE02)

Contact: `Jonathan.Peck@UGent.be`

Beschrijving

Zogenaamde *diffusiemodellen* zijn de huidige *state-of-the-art* in beeldverwerking [1]. Tegelijkertijd blijven neurale netwerken kwetsbaar voor *adversarial attacks*: onzichtbare wijzigingen van de originele beelden die de uitvoer van het model volledig kunnen veranderen [2]. Adversarial attacks vormen een gevaar voor neurale netwerken wanneer ze gebruikt worden in gevoelige contexten, zoals medische diagnoses. We tonen voorbeelden hiervan in figuur 1 waar de diagnose voorspeld door het netwerk helemaal verandert ondanks het feit dat de afbeeldingen visueel identiek zijn.



Figuur 1. Voorbeelden van adversarial attacks.

Men heeft aangetoond dat diffusiemodellen de robuustheid van neurale netwerken drastisch kunnen verbeteren [3]. Deze methode is echter computationeel enorm intensief: de gebruikte modellen zijn zeer groot en vereisen heel veel data om getraind te worden. Dergelijke hoeveelheden data zijn doorgaans niet voorhanden in medische contexten, waardoor deze methode niet toegepast kan worden.

De doelstelling van deze thesis is verkennen of de methode van [3] ook gebruikt kan worden in situaties waar er geen gigantische datasets beschikbaar zijn en grote modellen niet praktisch zijn. Hiertoe kunnen we volgend stappenplan nemen:

1. Ontwerp en train diffusiemodellen met niet meer dan 5 miljoen parameters op kleine datasets zoals Fashion-MNIST of CIFAR-10. Hiervoor kunnen we bestaande implementaties in PyTorch gebruiken [4].
2. Ga na in welke mate deze diffusiemodellen de robuustheid van kleine netwerken zoals MobileNet verbeteren wanneer de methode van [3] hierop wordt toegepast. Hiervoor kunnen we bestaande libraries zoals ART gebruiken [5].

De resultaten van deze thesis kunnen we gebruiken om een trade-off te maken tussen het aantal parameters van de modellen en de gewenste robuustheid. Dit kan een belangrijke gids zijn in praktische toepassingen van deep learning.

Referenties

1. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in Neural Information Processing Systems 33 (2020): 6840-6851. PDF

2. Peck, Jonathan, Bart Goossens, and Yvan Saeys. “An introduction to adversarially robust deep learning.” IEEE Transactions on Pattern Analysis and Machine Intelligence (2023). PDF

3. Carlini, Nicholas, et al. “(Certified!!) Adversarial robustness for free!.” arXiv preprint arXiv:2206.10550 (2022). PDF

4. Diffusers: state-of-the-art diffusion models. https://github.com/huggingface/diffusers.

5. Adversarial Robustness Toolbox. https://adversarial-robustness-toolbox.readthedocs.io/en/latest/